WHAT IS PERCEPTUAL ORGANIZATION FOR?

Andrew P. Wilkin
Jay M. Tenenbaum

Fairchild Laboratory for Artificial Intelligence Research

ABSTRACT—Perceptual organization has typically been viewed as a description of the image in compact or convenient form. We interpret it as a primitive *explanation* of the processes that generated the image, based on constraints derived from the discovery of regularities that are unlikely to be accidental. Deeper explanations are constructed by labeling, elaborating, and refining the primitive one. This interpretation of primitive perception has significant implications for perception at all levels, and for the relation between perception and cognition.

1. Introduction

Perceptual organization is easier to experience than to define (see Fig. 1.) The exquisite, curving parallel striations, the blobs, swirls, patches, edges, etc. leap out at us even when we have no idea what in particular we're seeing. These rich and vivid phenomena have long been the object of study by psychologists, most notably in the Gestalt school (e.g., [I], and more recently, [2,3].) The concern, however, has been almost entirely with the *what* of organization, hardly at all with the *why.* Where the issue of purpose has been addressed at all, it has generally been assumed that the goal of organization is to *describe* the image in a compact or convenient form, re-arranging it but adding no empirical content.

We propose that perceptual organization is not a description of the image at all, but a primitive, skeletal causal *explanation.* When we perceive the striations of Fig. 1 as a fixed profile, swept along a smooth trajectory and smoothly deforming along the way, the meaning of our percept is not just that the image *can* be generated that way, but that it *really was.* The basis for these primitive inferences is the discovery of similarities—literal spatiotemporal ones—that are extremely unlikely to arise by accident.

The way we discover similarities and use them to understand what we see, even at the primitive level of perceiving visible surfaces, strikingly resembles processes of discovery and explanation at the highest levels of abstraction and sophistication, exemplified in domains such as linguistics, biology, and geology. Within the constraints that non-accidental regularities provide, deeper interpretation proceeds by labeling, refining, and elaborating the initial model, discovering new regularities along the way as additional knowledge can be brought to bear.

Primitive perception is intelligence in microcosm: specialized, limited to the immediate spatiotemporal arena of the visual world, sacrificing flexibility for astounding power and speed; in a sense, an idiot savant. Yet all the essential principles and methods of scientific discovery and explanation appear to be in place.

Figure 1. Perceptual organization is easier to experience than to define.

2. The Laws Of Organization

To investigate the purpose of perceptual organization, we must have some notion of what perceptual organization does: its *language*—the set of elements and relations we perceive and the *laws* or principles that map expressions in that language onto particular images. We need not review the long history of organizational laws in the Gestalt tradition. It suffices to observe that relations of literal resemblance figure prominently in perceptual organization: when we *can* perceive a piece of the image by copying or smoothly sweeping simple elements, we are strongly inclined to do so. When we see parallel lines, we see their common form, and we see the simple relation, e.g. translation, that holds between them. Any list of the relations associated with organization would surely include, along with parallelism: colinearity or smooth continuation; common fate or rigidity; coherent spatial or temporal flow; symmetry; smoothly changing brightness, color, or texture over a region; etc. All of these, like parallelism, can be expressed as the replication or continuation of a base element through space or time, perhaps with smooth deformation along the way. To say that we perceive the image in these terms whenever possible is to characterize perceptual organization as the vigorous rooting out of relations of similarity.

3. Causal Links

The key to the meaning of perceptual organization is that all of these relations frequently arise for a variety of good reasons—ultimately because the world around us is coherent over space and time. They are however extremely unlikely to hold by chance, among elements that aren't directly related. Therefore, when we *can* derive one piece of the image from another by a simple transformation, it almost certainly isn't an accident. The more complex the transform's operands, and the more precise the transformation, the less likely the relation is to be spurious. This argument has been developed in more detail in [4], with a related but distinct account by Lowe and Binford [5]. We won't belabor it here; a simple example will make the point.

If we see two highly irregular but perfectly parallel curves, we can be quite certain that their correspondence is not an accident, even if we have no idea where the curves came from. There might be many reasons for their similarity, but there must be *some* reason. Asserting that the relation can't be accidental imposes a definite constraint on whatever explanation we might adopt for the curves: any explanation that claims the shapes of the curves arose independently, that they just happen to look the same, must be wrong. As we shall see, this is actually a rather strong constraint.

Continuing the example of parallel curves, asserting that their parallelism is not an accident implies that any process that determined the shape of either curve determined the shapes of both curves in common; and that conversely, any process that acted independently on one curve but not the other acted transparently, in the sense that its effect at most was merely to copy the underlying common form. However we try to construct an explanation that violates these conditions, we will find we have implied that the relation *is* accidental, in whole or in part. The bare assertion of non-accidentalness thus irrevocably divides the entire generating process into two parts: a common determinant of shape, and a transparent replication. This broad division provides a primitive skeletal model for the generating process as the replication of a common form. Any subsequent enlightenment we receive can only elaborate and refine the model's internal structure, but never really supplant or negate it.

Again, some simple examples will point out what the skeletal model does and does not specify. There is an endless list of plausible explanations that account for the parallel curves by replication of a common shape. For instance, they might both have been traced from a third curve; either might be a photograph of the other; they might be the borders of a winding road, or tracks left in the ground by a rake, etc. All of these plainly account for the curves by replicating a single shape, although they do so in very different ways, even assigning different directions of causality. All are elaborations of the same basic model, but without further evidence we are free to fantasize about the nature of the common cause and the manner, causal path, and degree of directness of the replication.

Now suppose we try to modify one of these models to make it inconsistent with the skeletal model, but without attributing the parallelism to chance. We might plausibly decide that both curves were hand-traced from a third, but suppose we also asserted that the tracer's hand slipped badly while reproducing one of the curves. We must now introduce a second change to balance the first, because the curves *are* parallel. If we claimed that exactly the same mistake was independently made in tracing the other curve, we would be attributing that part of the parallelism to chance. On the other hand, if we claimed that the mistake was deliberately repeated on the second tracing—a non-accidental relation—we would be adding the original mistake to the underlying form, and its duplication to the replication component, elaborating but not violating the skeletal model. If we claimed that the mistake was erased and corrected, and therefore invisible in the final result, then the combination of mistake-erasure-correction is predictably and lawfully transparent to shape, just a step in the replication process, so the skeletal model is still preserved. As these simple examples suggest, any explanation inconsistent with the skeletal model negates the assertion of non-accidentalness. Thus, as long as that assertion is retained, any explanation we eventually adopt must amount to an elaboration or refinement of the skeletal model.

If perceptual organization provides skeletal causal models, and if these models can be elaborated but never discarded, then our most primitive organizational percepts ought always to survive, embellished but still intact, to the highest levels of interpretation. And of course they nearly always do. Some pictures, e.g. micrographs, are so alien to our ordinary world that the naive observer may have no idea what they portray, beyond the spots, swirls, bands, etc. of pure organization. When the naive observer becomes an informed one, understanding in detail what the picture means and where it comes from, the spots and swirls never vanish. On the contrary, the informed observer is usually able to *explain* them in detail. A spot, for instance, may no longer be seen as a spot but as a cell, yet its form survives. A finely striated structure may become a coherent bundle of nerve fibers, but again its underlying form survives intact.

We have argued that regularities such as parallelism are unlikely to be accidental, and that the assertion that they are not implies a primitive skeletal model of the generating process, as the replication of a common form. Since perceptual organization does discover similarities, and does represent them as replications of common forms, we interpret perceptual organization as expressing exactly these primitive explanations. Perceptual organization is discovery, not description.

4. From Perception to Cognition

Viewing organization as explanation immediately draws a connection between primitive perception and the broader cognitive realms of discovery, induction, and causal reasoning: the discovery of compelling regularities;—relations that "can't be an accident"—is a basic ingredient in discovery and explanation even at the highest levels of abstraction. "Seeing the pattern" in a body of phenomena is the critical step toward deeply understanding the phenomena. Is this seeming relation between primitive perception and cognition a superficial one? Frankly, we feel the resemblance is too strong to be an accident. We want to explore this connection in part because even the hint of unity across so broad a gulf is intrinsically exciting, and also because we feel the connection can shed some immediate light on the manner in which perceptual organization contributes to such basic activities as perceiving surfaces.

Many sophisticated problems of causal explanation actually resemble perception quite closely in overall form: given a picture of the present—not a literal picture, but a more abstract body of observations—construct a model of the processes and events that generated it, in light *of* knowledge of the domain. We will draw some observations from three such problems—inferring ancient languages, species, and geological structures from modern ones.

First, we find repeatedly that the discovery of non-accidental similarities is the point of departure for explanation, as we have argued it is in perceptual organization. Moreover, the discovery of similarities may be clearly distinguished from their *attribution* to particular sorts of events. For instance, a dolphin blatantly resembles a fish. If we conclude that a dolphin is a kind of fish, descended from fishes, we are wrong. Even so, the resemblance is no accident, but a striking instance of systematic convergent evolution. The causal link is there, but is attributable to common selection pressures rather than common ancestors.

In fact, the attribution of observed similarities to specific causes, such as convergent evolution, often amounts to choosing one of a small, fixed set of labels, determined by the nature of the regularity, and by prior knowledge of the domain. In biology, structural similarity may be attributed to common descent, convergent evolution, or imitation of one form by another. In linguistics, words may be similar across languages by common linguistic descent, or later borrowing or intermixing. In geology, the curvature of parallel sedimentary beds may be attributed to folding—deformation of a shape that originally resembled a layer cake or draping—deposition on top of a pre-existing bump, like snow burying a tree stump. Exactly the same discovery/labeling pattern can be seen in surface perception, although of course the labels are different. We will give some examples in the next section.

Relationships such as parallelism and symmetry are ubiquitous. Their significance therefore depends little on details of the domain. On the other hand, because they arise for so many diverse reasons, few specific conclusions about the generating process can be drawn on their evidence alone. However, there are many examples of transforms derived from more specialized domain knowledge. Like parallelism, these are extremely unlikely to arise by chance, and so supply the same kind of causal link. However, since they rest on very particular properties of the domain, their presence can usually be explained in only one way. An example from linguistics is the systematic phonological shifting that occurs as languages evolve. Where two languages can be related by these transformations, there is only one explanation, divergence from a common ancestor; and the details of the observed transformations can give an accurate picture of the drifting process and the common ancestor. Again, we will see in the next section that special transformations (such as projection) play a similar role in vision.

Finally, of the problems we have mentioned, geological interpretation bears especially on perceptual organization: geological structures are contiguous spatial entities, and the data may often be literal pictures, e.g. cross-section views. Although the reasoning process may ultimately become abstract and deliberate, the basic regularities are provided in large measure directly by perceptual organization—the parallel bands of sedimentation, the curvature of folding, the edges of faults and unconformities, etc. In this case and many others, the process of discovering similarities for high-level explanation isn't just *like* perceptual organization; it *is* perceptual organization.

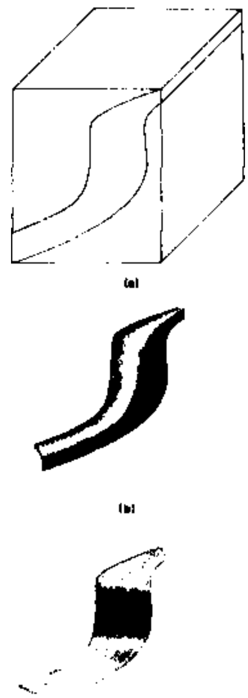

Figure 2. Parallel lines may be labeled in several ways. Here, different labelings have been forced by disambiguating context.

5. Surface Perception

In this section we will draw some parallels between basic surface perception and the high-level problems discussed above.

First, the distinction between discovery of regularity and attribution or labeling is widespread in vision as well. Our perception of a set of parallel lines may flip between wavy lines on a flat surface, lines ruled on a wavy surface, and a curved cylinder, modulated by bits of context [6](see Fig. 2,) just as curved parallel lines may signify folding or draping in geology. Edges, another product of perceptual organization, also admit a small set of labels [7,8]. If we label a bright spot as a highlight, then we label the coherent form on which it lies as a glossy surface, rather than a matte one [9]. Reducing our explanations to such lists of idealized options appears to be a useful simplification in vision, just as it is in other domains. Mistaking a shadow for a surface boundary is just like mistaking a dolphin for a fish.

Second, we saw that specialized transformations—e.g. phonological ones—may reveal new, unobvious regularities. Discovering these regularities, and observing the parameters of the transformations that revealed them, may impart very specific information about the generating process. In vision, of course, the most obvious examples of such transformations are provided by the laws of photometry and projective geometry. Specialized as they are, these laws are always with us.

Much recent work in computational vision has been aimed at recovering quantitative three-dimensional information. In general, models of the imaging process have assumed center-stage. The constraints these models give are underdetermined: all recovery techniques have taken up the slack by imposing a variety of continuity or smoothness constraints and domain restrictions (e.g. of planarity, constant reflectance, etc.) However, we can now provide a very different and more satisfying view of three-dimensional recovery: the basic regularities we have discussed—parallelism, colinearity, etc., are inherently neither two nor three dimensional. We may seek those regularities in the two-dimensional image, and build our skeletal explanations in 2-space. However, concatenating these general transformations with the specialized ones of projection and photometry defines a broader set of potential regularities, just as the imposition of phonological laws does. In effect, by applying these transformations in reverse to the image, we generate the set of three-dimensional configurations consistent with the image, then search for the same basic regularities in that expanded space. When we discover them, the non-accidentalness argument and its consequences apply just as in two dimensions. The three-dimensional structure is recovered, almost as a byproduct, from the particular reverse projection through which the regularity was found. We may thus reverse the usual emphasis, viewing three-dimensional recovery as simply falling out of perceptual organization, when organization is augmented by the special transformations of photometry and projection.

Some existing work illustrates the merit of this view in special cases, where regularities are invisible in the image, being revealed only through reverse projection: the structure of random-dot stereograms exists only through stereopsis [10]. Seemingly chaotic motion in the image may become regular rigid motion, if the regularity is sought in three dimensions rather than two [II]. Regularity of texture [12] or contour [6] may be sought in three dimensions to infer surface shape. Although in these instances depth and shape are recovered by seeking regularity through reverse projection, the unified view of shape recovery as perceptual organization in three-dimensions has not been generally recognized.

6. Conclusions

Elevating perceptual organization from the role of description to that of discovery and explanation radically alters our view of primitive perception: at the very earliest levels, we are creating primitive skeletal explanations based on the discovery of non-accidental similarities. As our interpretations deepen, these primitive models are labeled, elaborated, refined, and disambiguated, but never discarded. The search for regularity in three dimensions is no different in principle than in two, once we are in possession of the special transformations that relate images to scenes. In a reversal of the usual emphasis, we view the recovery of three-dimensional structure as a byproduct of the discovery of regularities in the expanded space these transformations define.

We've also suggested that primitive perception is intelligence in microcosm: the methods and principles of discovery and explanation appear on the face of it to be much the same. Discovering an edge, then deciding that it is a shadow rather than a surface boundary is much like discovering the structural resemblance between dolphins and fishes, then attributing the similarity to convergent evolution rather than common descent. Using projective geometry to discover 3D rigid motion that superficially appears chaotic is much like using phonological laws to discover nonobvious systematic relations among languages: in both cases specific information is derived by observing the parameters of the transformations through which the regularity was found.

It should be clear that this brief paper is just a snapshot, passing over many significant issues entirely, touching others only lightly. All of them expand into deep and rich areas for investigation. In our own research we are focusing on the language and rules of spatiotemporal organization, closely guided by our understanding of the purpose of organization. More generally, we hope that this view of organization will foster interest in the area and provide guidance for the more detailed study of its principles and mechanisms. We hope as well that the hints of deep commonalities between the lowest levels of perception and the highest levels of cognition will contribute to bridging the gulf that has unfortunately separated these fields.

REFERENCES

[I] W. Kohler Gestalt Psychology *Liverright, New York,* 1929.

[2] F. Attneave Pragnanz, and Soap-Bubble Systems: A Theoretical Exploration. *University of Oregon, Psychology Department, Eugene, Oregon,* 1979.

[3] Leeuwenberg, E. L. J. *Structural Information of Visual Patterns: An Efficient Coding System in Perception.* Mouton, The Hague, 1967

[4] A. P. Witkin and J. M. Tenenbaum On the Hole of Structure in Vision, to appear in *Human and Machine Vision, ed. A. Rosenfeld, Academic Press,* 1983.

[5] D. G. Lowe and T. O. Binford Segregation and Aggregation: An Approach To Figure-Ground Phenomena. *Proceedings of the DARPA Image Understanding Workshop Science Applications Incorporated, McLean, Virginia,* Septem 1982, pp. 168-178.

[6] K. A. Stevens Surface Perception From Local Analysis of Texture and Contour. *M.I.T., Cambridge, Massachusetts,* February 1979.

[7] Waltz, D. L. *Generating Semantic Descriptions from Drawings of Scenes With Shadows.* AI-TR-271, M.I.T., Cambridge, Mass., 1972

[8] A. P. Witkin Intensity-Based Edge Classification. *Proceedings of the Second National Conference on Artificial Intelligence, American Association for Artificial Intelligence, Stanford, California,* August 1982, pp. 36-41.

[0] J. Beck Surface Color Perception. *Cornell University, Ithaca, New York,* 1974.

[10] Julesz, B. *Foundations of Cyclopean Perception* University of Chicago Press, Chicago, Illinois, 1971

[II] S. Ullman The Interpretation of Visual Motion. *M.I.T. Press, Cambridge, Massachusetts,* 1979.

[12] A. P. Witkin Recovering Surface Shape and Orientation From Texture. *Journal of Artificial Intelligence, Volume 17, Number* 1-5 August 1981, pp. 17-45.