Perception as Bayesian Inference
DC Knill, WA Richards, eds
New York: Cambridge University Press, 1996

The world, the brain, and the speed of sight

Ronald A. Rensink
Vision Sciences Laboratory
Department of Psychology
Harvard University
Cambridge, MA 02138

(Commentary on Chapter 11:"The perception of shading and reflectance, EH Adelson and AP Pentland)

Adelson & Pentland use an engaging metaphor to illustrate their position on scene analysis: interpretations are produced by a workshop that employs a set of specialists, each concerned with a single aspect of the scene. The authors argue that it is too expensive to have a supervisor co-ordinate the specialists and that it is too expensive to let them operate independently. They then show that a careful sequencing of the specialists leads to solutions of minimum cost, at least for their world of Mondrian panels.

The authors admit that their approach is based on a relatively simple domain, and will have to be developed further if it is to be applied to more realistic situations. But can it really provide the basis for a better understanding of more general-purpose vision? In what follows, it will be shown that this approach is based on several rather strong hypotheses, some of which will have to be seriously modified or even replaced if an extension is to be made to more general domains. Three sets of issues are of particular concern here: the kinds of scenes that can be accurately interpreted this way, the relevance of such models to human vision, and the trade-offs between interpretative power and processing speed.

Scene domain

The operation of the workshop begins with a specialist that uses contour information to recover the three-dimensional (3D) shape of the surface. Since there is no feedback from the other specialists, error here is irreversible. The potential therefore exists for these errors to cascade through the subsequent stages and lead to large inaccuracies in the final interpretation. Shape recovery must be accurate enough to prevent such cascades.

From the description of the shape specialist, it is not clear whether planarity is enforced absolutely, or is only maximized. If planarity is simply maximized, a region could correspond to a nonplanar surface, and so would not have the uniform luminance assumed in the analysis. This in turn could lead to serious inaccuracies in the final interpretation. However, these problems can presumably be avoided by invoking the appropriate planarity constraint.

A more serious concern is that the shape specialist operates on contours alone. This amounts to a claim that (at least for the world of Mondrian panels) there is no need for shape from shading—indeed, shading is to be completely disregarded. If true, this would be a most interesting result, and so must be carefully verified. However, shading cannot be ignored for more general domains. Such information can help to establish which areas in the image belong to the same surface in the scene, and this in turn can have serious consequences for the recovered 3D structure. For example, shading can influence the perceived ordering of depth (Figure 1a) and the perceived completion of edges and surfaces (Figure 1b). To get around this problem it is necessary to somehow take reflectance into account, possibly by having reflectance estimates be directly proportional to intensity (as is done for the lighting direction). But whatever the method used, it will still be necessary to verify that shape recovery remains accurate.

More generally, if Adelson & Pentland's approach is to be extended to larger domains, it will need to determine the kinds of interpretations generated by the fixed sequence of specialists, and to determine how well these capture the (statistical) structure of the domain under consideration. Ultimately, the relevance of this approach will depend on the existence of a world sufficiently "friendly" that its structure can be reasonably approximated by the relatively restricted set of "sequence-generated" interpretations.

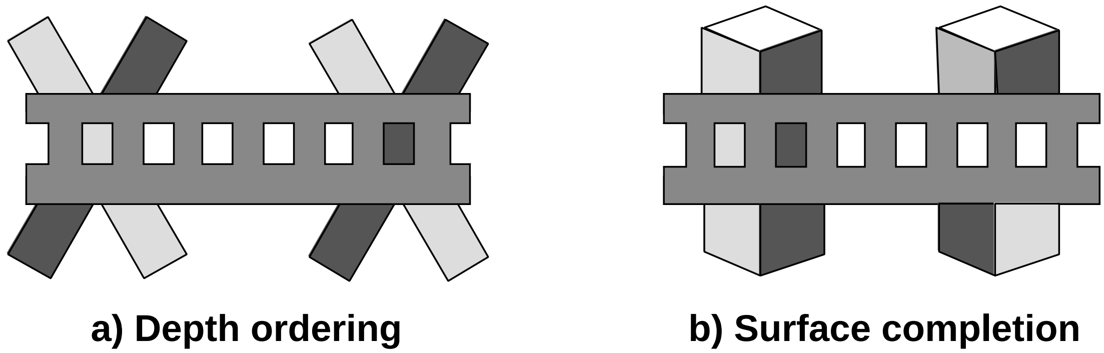

Figure 1. Influence of shading on 3D surface structure. In (a), the pattern of shading determines which of the oblique panels is perceived as being immediately behind the occluder. In (b), the pattern of shading determines whether the occluder is perceived as obscuring one long post (surfaces completed) or two short cubes (surfaces left incomplete).

Relevance to human vision

For Adelson & Pentland, the criterion for a "correct" interpretation is that it corresponds to what humans perceive. Although this is fine for purposes of general discussion, more detail is needed if their approach is to be put on a more rigorous basis.

To begin with, it is important that the solutions be tested against the results of controlled psychophysical experiment. Although Adelson & Pentland have proposed one particular set of constraints, it is possible that others are used instead. For example, the shape specialist could maximize the number of orthogonal corners (Kanatani, 1990), and the lighting specialist could be based on a nonlambertian reflectance function. It is only by a careful quantitative comparison with the results of controlled studies that the actual set of constraints used can be definitively established.

It is also necessary to specify the kind of task to be used in these experiments. What humans "perceive" depends to a large extent on what they are asked to do, since different kinds of tasks can access different kinds of representations. For example, the interpretation of line drawings as 3D objects at preattentive levels differs considerably from that based on "casual" viewing (Enns & Rensink, 1991). The criterion for a "correct" interpretation must therefore be tightened up.

Finally, if the workshop metaphor is taken to heart, it is also important to test not only the results of the specialists acting together, but the results of the individual specialists as well. This could be done, for example, by presenting actual 3D polyhedral objects (with or without colored faces) to the observer, and determining if the interpretation is the same as that based on a picture or drawing of the scene.

Processing speed

Given that each of the specialists can be based (at least potentially) on different constraints, it is apparent that the power of Adelson & Pentland's approach does not depend strongly upon the particular form of these constraints. But what then is the critical factor?

A closer look at the workshop shows that an important element in its operation is the *sequencing* of the specialists: The slants of the planes are first recovered from image measurements, the lighting direction is then recovered from the slants, and the reflectances are then recovered from the slants and lighting directions. This is an instance of *dynamic programming* (see, e.g., Kumar et al., 1994), a technique that breaks a task down into a sequence of nested subtasks, arranged such that the solutions of the earlier subtasks help solve those later in the sequence. The unidirectional flow of information from the earlier subtasks to the later ones allows iteration to be avoided, which keeps computational complexity low.

Thus, the difference in "cost" between fixed sequencing and direct supervision (i.e., optimal interaction among the subtasks) involves processing effort rather than quality of solution. Since the cost of the supervisor differs in kind from the cost of the other specialists, adding these costs together is like adding apples and oranges. What is required instead is a framework that shows how processing effort can be traded off against quality of interpretation (see, e.g., Rensink, 1992). This could then serve as the basis for a true computational analysis, which would explain not only the reasons for the particular structure of the specialists, but also the reasons for the particular pattern of their interactions.

References

Enns, J. T., & Rensink, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review*, **98**, 335-351.

Kanatani, K. (1990). *Group-Theoretical Methods in Image Understanding*. Berlin: Springer.

Kumar, V. Grama, A., Gupta, A., and Karypis, G. (1994). *Introduction to Parallel Computing: Design and Analysis of Algorithms.* Redwood City, CA: Benjamin / Cummings.

Rensink, R. A. (1992). *The Rapid Recovery of Three-Dimensional Orientation from Line Drawings.* Ph.D. Thesis (also Technical Report 92-25), Department of Computer Science, University of British Columbia, Vancouver, BC, Canada.