Perceptual Organization as a Foundation for Graphics Recognition

Eric Saund, James Mahoney, David Fleet, Dan Larner
Xerox Palo Alto Research Center
3333 Coyote Hill Rd.
Palo Alto, CA 94304

1 Summary

This paper motivates an approach to graphics recognition grounded in a framework for human and machine vision known as *Perceptual Organization*. We review some of the characteristics of this approach that distinguish it from traditional engineering of document recognition systems, and we suggest why and how the techniques and philosophy of Perceptual Organization might lead to advances in the very practical matters of interpreting diagrams, drawings, and sketches.

2 What is Perceptual Organization?

Perceptual Organization (PO) refers to a cross-disciplinary tradition in the study of the *intermediate* stages of vision. Perception lies above the level of sensory processes such as image enhancement, contrast normalization, and feature measurement, but below the levels of end tasks such as object recognition, data indexing, and scene modeling. Prior to interpreting semantically meaningful objects, PO builds representations for geometric and visual structure: What are the readily apparent regions of an image? What grabs an observer's attention? What features tend to group together to form larger units? What emergent patterns become apparent? Perceptual Organization is as far as your visual system gets when you view abstract art: you can make sense out of parts of the image locally and perhaps assemble some limited gist of the whole, but you don't necessarily know what it is you're supposed to be looking at. That abstract art is in fact interesting to look at says something important about how the most powerful extant vision systems work.

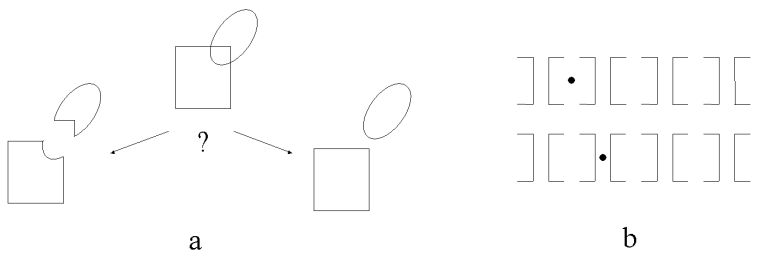

Figure 1: Demonstrations of Perceptual Organization. a. *smooth continuation:* which decomposition into parts seems more likely? b. *closure:* which dot appears to be on an object?

Classically, Perceptual Organization has been concerned with the kinds of visual structure identified by the school of Gestalt Psychology, dating from the 19th century. The "Gestalt Laws" of *smooth continuation*, *common fate*, *symmetry*, *similarity*, *proximity*, *closure*, etc. are taught in some form in almost every introductory Psychology class. The Gestalt phenomena were traditionally studied by psychophysicists who devised wickedly clever visual displays attempting to expose broad principles governing human vision.

More recently, computer vision researchers have attempted to model these processes computationally, while beginning to focus on general issues associated with intermediate stages of visual analysis. As expressed by Witkin and Tenenbaum (1982), at the very heart of visual analysis is our "ability to impose organization on sensory data - to discover regularity, coherence, continuity, etc., on many levels." We believe that the philosophy, methodology, and techniques of Perceptual Organization are now ready to contribute to areas in document image analysis. Graphics images, in particular, present an ideal context in which to carry out this cross-fertilization of disciplines.

3 What Would it Mean to Apply PO to Graphic Documents?

Let us consider three examples of how general processes for detecting intermediate level visual structure might be applied to graphics recognition, in contrast to a straw-man "engineering" approach.

Example 1: text/line art segmentation. Text and graphical material are often separated early in graphics recognition programs, so that text recognition and spatial analysis procedures can be applied respectively to appropriate input data. Traditional methodology is to focus on the properties that specifically distinguish text from line art, then build a classifier to exploit such properties of ideal text as connected component dimension, connected component spacing, and linear alignment of components' edges along parallel toplines and baselines.

A PO approach might approach this problem somewhat differently. Instead of focusing on the text/line art distinction specifically, PO would call for a more general process of identifying compact regions having consistent texture properties. These would call out not only ideally printed text, but also graphic elements that resemble text in appearance, certain kinds of smudges and noise, and scrawled handwritten notes that violate the nominal constraints that most text obeys. These objects would then serve as a starting point for more specialized text classification and recognition.

Example 2: hash region detection. Mechanical drawings depict cutaway views by parallel hash lines. These are typically detected by specialized procedures for finding repeating parallel-lines, parameterized by spacing, line weight, and orientation. These procedures are launched when certain cues are encountered such as a line found inside a contour classified as "object boundary".

A PO approach might instead perform an initial global region segmentation step designed to identify regions distinguished by uniform textural properties such as average

lightness, distribution of characteristic grain size, spatial frequency or image orientation, and distribution of texture primitives like curve ends, corners, and bars. This process would not only detect hash line regions, but also graphic representations for dirt, concrete, wood grain, etc., as well as blocks of text. The intermediate stage texture regions would serve as candidates for more specialized classification. This processing might also help in deciding which contours comprise an object boundary by virtue of enclosing a uniform texture region.

Example 3: detection of enclosing boxes and ovals. Charts and diagrams frequently represent conceptual objects as text enclosed by rectangular boxes, or ovals. A standard approach to detecting these is to exploit their geometric regularities. Boxes contain two sets of parallel lines at right angles; Hough techniques prove useful here. Ovals can be fit by a piecing together the outputs of circular arc finding routines.

A PO approach might instead begin by detecting compact objects formed by closed paths. This would turn up not only rectangles and ovals, but other enclosing graphic shapes like octagon, and triangles, as well as salient graphical structure involving closed paths. As in the first two examples, the intermediate level structure provides building blocks for more specialized recognition routines.

Perceptual Organization, as such, does not "solve" any of these problems; much research remains to be done to flesh out the algorithms and architectures. Rather, our point is that PO approaches the problem in a different way, placing emphasis on finding intermediate level structure, prior to turning to knowledge about a particular task or image domain.

4 Characteristics of a Perceptual Organization Approach

These examples illustrate several characteristics of the Perceptual Organization approach.

- *weak prior models:* It is well understood that knowledge about an image source can be exploited to direct search and to select appropriate representations for inferences to match task requirements. Classically, document image analysis has pursued *strong prior models* incorporating as much prior knowledge as possible. This approach can become brittle as actual data departs from the assumptions of prior models. By contrast, PO suggest the use of *weak models*, to identify salient visual structure common to broad ranges of image domains. As a result, knowledge is kept modular, postponing the use of domain-specific models.

- *rich and redundant representations:* Perceptual Organization seeks to make explicit many types visual structure in ways that may amount to an overcomplete or redundant description of an image. For example, a rectangle can be described economically with five numbers, while its perceptually significant aspects include the presence of a blob with a rough size, orientation, and elongation; straight lines; parallels; corners; a closed region; and more.

- *incremental description:* In PO, typically each assertion is limited in scope. Images are described in terms of large numbers of simple statements instead of all-encompassing models with a multitude of parameters.

- *partial, conflicting, overlapping, and ambiguous assertions:* Perceptual Organization is content to deliver partial interpretations of data, and to have some of these interpretations overlap or conflict with one another. For example, two squares sharing a common edge may both give rise to assertions that interpret the edge as "belonging" to one square or the other, or both.

5 Why Take a PO Approach?

There are a variety of reasons to believe that Perceptual Organization, the intermediate level representation of generic visual structure, offers several benefits for document image analysis.

Documents are human artifacts designed to be viewed and interpreted by human visual systems which employ processes of Perceptual Organization. It is no accident that properties of proximity, symmetry, closure, and smooth continuation are used in abundance in the design of graphic documents. Therefore, it is sensible to believe that modeling some of these processes in machines will support processes mimicing human strategies for interpreting and understanding them.

Perceptual Organization will lead to more natural and intuitive user interfaces for document image analysis since the elements of visual structure identified by PO provide an impedance match to users. If a graphic editor program could let users select graphic elements on the basis of the visual objects they comprise instead of on an element-by-element basis, for example the entire set of markings comprising a texture fill of a region, users could make modifications with much greater facility.

Perceptual Organization is sound engineering. It attempts to introduce knowledge in a modular fashion. Techniques that don't rely on detailed knowledge of any particular image document type hold the promise of generality and stability across applications and image domains. This will include images of structured, formal documents, as well as informal documents such as hand-drawn sketches and annotations.

Perceptual Organization promises a broad foundation for a variety of applications. Intermediate level visual events become building blocks for later processes. In this sense, Perceptual Organization encompasses classical feature detection, where the features can be more complex and more abstractly construed than simple local measurements. Indeed, some work in graphics recognition can be viewed in this framework, including advances in region segmentation, adaptive thresholding and preclassification, and grouping approaches to assembling larger structures from smaller partial collections.

6 Sophistication and Robustness

The benefit of PO will be the ability to support machine conversion and interactive tasks with graphic documents that current systems either cannot cope with or else require spe-

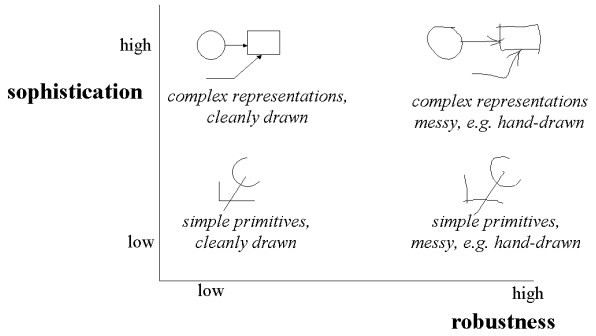

Figure 2: Classical document recognition approaches pursue sophisticated models incorporating greater domain knowledge. Perceptual Organization seeks, first, robust descriptions of intermediate level visual structure that will apply across document genres.

cialized programming and training to handle. For example, hand-drawn sketches are, for people, not much more difficult to read than cleanly drafted drawings, yet the subtle details of whether lines, curves, junctions, regions, and text obey the constraints of drafting technique lead to failure of classical systems.

Accordingly, two key objectives in improving graphics recognition systems are to achieve greater sophistication and to achieve greater robustness (see Fig. 2). *Sophistication* refers to the knowledge complexity of the interpretive models employed. A system that knows only about line, rectangle, and ellipse elements is less sophisticated than one that knows about the syntax of how these combine to form organization charts and process flow diagrams. *Robustness* refers to the range of variability the recognition system can tolerate. A system that can only recognize printed CAD drawings is less robust than one that can also recognize the equivalent rendered as a hand-drawn sketch.

Most attention in classical document image analysis has been devoted to the vertical axis, organizing knowledge so that domain-dependent constraints can interpret graphics in terms of complex models while applying top-down methods to cope with noisy data. Perceptual Organization is concerned with achieving robustness first. This is not simply a matter of "modeling the noise" within a knowledge-intensive system. Instead, according to a PO approach, graphically different figures that are perceived equivalently by human observers are not considered a matter of noise, but of truly equivalent intermediate level visual structure.

References

Witkin, A.P. and Tenenbaum, J.M. (1983) On the role of structure in vision. In Beck, J., Hope, B., and Rosenfeld, A. (eds.), **Human and Machine Vision**, Academic Press, pp. 481–543.