

A Pyramid Model of the Perception of Partially Visible Figures

Michael R. Scheessele (mscheess@iusb.edu)

Department of Computer & Information Sciences, Indiana University - South Bend
1700 Mishawaka Ave., South Bend, IN 46634 USA

Zygmunt Pizlo (pizlo@psych.purdue.edu)

Department of Psychological Sciences, Purdue University
1364 Psychological Sciences Bldg., West Lafayette, IN 47907 USA

Introduction

Frequently, figures in our visual field are only partially visible. One figure may partially **occlude** another, for example, or a particular figure may appear **fragmented** due either to camouflage or to low contrast between it and the background. Despite such challenges, the human visual system routinely perceives figures that may only be partially visible.

One prior theory of the perception of partially occluded figures (Nakayama, Shimojo, & Silverman, 1989) states that contours “intrinsic” to a figure of interest must be distinguished from those “extrinsic” to it and that this classification requires depth cues. Our theory proposes that the human visual system can use a *variety of cues, local or global*, to perform this classification and that this classification serves as the basis for perception of both partially occluded *and* fragmented figures. Further, we propose that an exponential pyramid, from the machine vision literature, provides a good model of how the human visual system implements this classification.

Exponential Pyramid Model Description

The Exponential Pyramid has been proposed as an adequate model of the human visual system (Rosenfeld, 1990; Pizlo, Salach-Golyska, & Rosenfeld, 1997). Our model uses a “non-overlapped quad-pyramid”. Assume that the bottom layer of the pyramid has n processing nodes. The next layer has $n/4$ nodes, the one above that $n/16$ nodes, and so on. The top layer has only one node. Each node in a layer connects with four distinct ‘child’ nodes in the immediately lower layer and one ‘parent’ node in the immediately higher layer. Such a pyramid has $(\log_4 n) + 1$ layers. Each node in the pyramid has limited memory and processing capability. An image is input to the bottom layer (Jolion & Rosenfeld, 1994). The image may also be represented at each higher layer (with increasing spatial scale or ‘receptive field size’).

Our model features a bottom-up processing stage followed by a top-down stage. In the bottom-up stage, local variance of various contour features (e.g., orientation, length) is computed. When the variance of a contour feature abruptly changes between successively higher layers, the presence and position of a figure in the image is indicated (i.e., the figure ‘comes into view’). In the top-down stage, the statistical information computed in the bottom-up stage is used to classify image contours as either intrinsic or

extrinsic to the target figure. The model has only one free parameter: the standard deviation of decisional noise. Model and human performance were compared across 11 experimental conditions.

Method

In each trial of the **human psychophysical experiments**, a polygonal figure was partially occluded by simple shapes – diamonds (Exp. 1, two conditions) and squares (Exp. 2, nine conditions). A subject’s task was to respond whether the figure was presented in its upright or rotated (180°) position. Contours of occluders differed from those of the figure in terms of orientation (Exp. 1) and length (Exp. 2). Depth cues from occluders were minimal. **Model simulations** were run for all 11 conditions using the same sets of stimuli as those used by the human subjects.

Results

Subjects used orientation (Exp. 1) and length (Exp. 2) differences between the contours of a target figure and those of occluders, in detecting the figure. Model simulations accounted well for human performance in the 11 conditions of Experiments 1 and 2.

Conclusions

The human visual system can detect and use a variety of cues, local or global, to classify contours as either intrinsic or extrinsic to a partially visible figure. Our exponential pyramid-based computer model provides a good account of how the human visual system implements this process.

References

Jolion, J. M., & Rosenfeld, A. (1994). A pyramidal framework for early vision. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, 18, 55-68.

Pizlo, Z., Salach-Golyska, M., & Rosenfeld, A. (1997). Curve detection in a noisy image. *Vision Research*, 37, 1217-1241.

Rosenfeld, A. (1990). Pyramid algorithms for efficient vision. In C. Blakemore (Ed.), Vision: Coding and efficiency. Cambridge, Great Britain: Cambridge University Press.