Cognitive Psychology 58 (2009) 1–48

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds ☆

Arash Fazl, Stephen Grossberg *, Ennio Mingolla

Department of Cognitive and Neural Systems, Center for Adaptive Systems and Center of Excellence for Learning in Education, Science, and Technology, Boston University, 677 Beacon Street, Boston, MA 02215, USA

ARTICLE INFO

Article history:
Accepted 6 May 2008
Available online 23 July 2008

Keywords:
Category learning
Object recognition
Spatial attention
Object attention
Parietal cortex
Inferotemporal cortex
Saccadic eye movements
Attentional shroud
Adaptive Resonance Theory
Surface perception

ABSTRACT

How does the brain learn to recognize an object from multiple viewpoints while scanning a scene with eye movements? How does the brain avoid the problem of erroneously classifying parts of different objects together? How are attention and eye movements intelligently coordinated to facilitate object learning? A neural model provides a unified mechanistic explanation of how spatial and object attention work together to search a scene and learn what is in it. The ARTSCAN model predicts how an object's surface representation generates a form-fitting distribution of spatial attention, or "attentional shroud". All surface representations dynamically compete for spatial attention to form a shroud. The winning shroud persists during active scanning of the object. The shroud maintains sustained activity of an emerging view-invariant category representation while multiple view-specific category representations are learned and are linked through associative learning to the view-invariant object category. The shroud also helps to restrict scanning eye movements to salient features on the attended object. Object attention plays a role in controlling and stabilizing the learning of view-specific object categories. Spatial attention hereby coordinates the deployment of object attention during object category learning. Shroud collapse releases a reset signal that inhibits the active view-invariant category in the What

☆ A.F. was supported in part by the Air Force Office of Scientific Research (AFOSR F49620-01-1-0397), the National Science Foundation (NSF SBE-0354378), and the Office of Naval Research (ONR N00014-01-1-0624). S.G. was supported in part by the Air Force Office of Scientific Research (AFOSR F49620-01-1-0397), the National Science Foundation (NSF SBE-0354378), and the Office of Naval Research (ONR N00014-01-1-0624). E.M. was supported in part by the National Science Foundation (NSF SBE-0354378) and the Office of Naval Research (ONR N00014-01-1-0624).

* Corresponding author. Fax: +1 617 353 7755.
E-mail address: steve@cns.bu.edu (S. Grossberg).

0010-0285/$ - see front matter
doi:10.1016/j.cogpsych.2008.05.001

cortical processing stream. Then a new shroud, corresponding to a different object, forms in the Where cortical processing stream, and search using attention shifts and eye movements continues to learn new objects throughout a scene. The model mechanistically clarifies basic properties of attention shifts (engage, move, disengage) and inhibition of return. It simulates human reaction time data about object-based spatial attention shifts, and learns with 98.1% accuracy and a compression of 430 on a letter database whose letters vary in size, position, and orientation. The model provides a powerful framework for unifying many data about spatial and object attention, and their interactions during perception, cognition, and action.

1. Introduction

What is an object? How can we learn what an object is without any external supervision? In particular, how does the brain learn to recognize a complex object from multiple viewpoints? Consider what happens when we first look at an object that is not instantly recognizable. We make scanning eye movements, directing our foveas around to a variety of points of interest, or views, on the object. The object's retinal representations of these views are greatly distorted by cortical magnification in cortical area V1 (Fig. 1). The brain somehow combines several such distorted views into an object recognition category that is invariant to where we happen to be gazing at the moment. Future encounters with the same object can therefore lead to fast recognition no matter what view we happen to see.

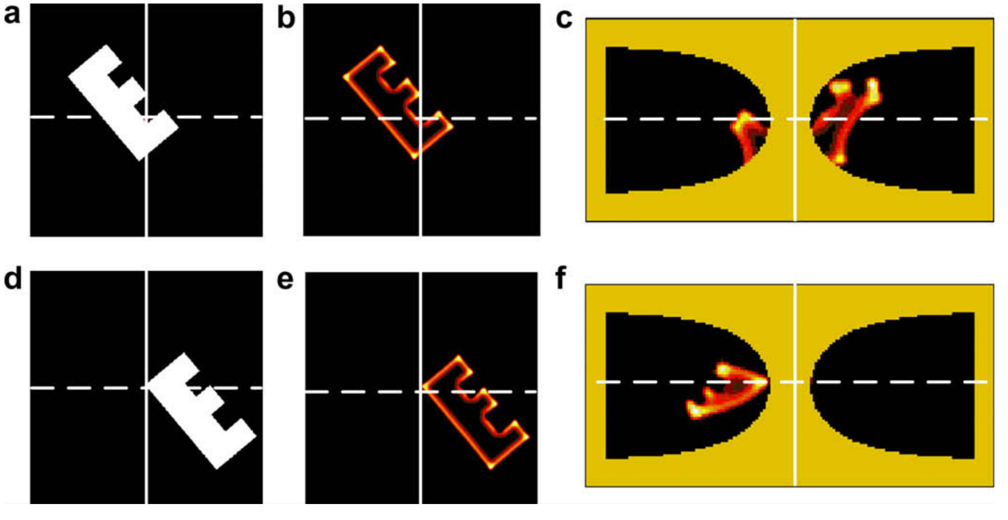

Fig. 1. Visual input distortion due to cortical magnification. The activity generated in the primary visual cortex by a foveation (view) of an object depends on the position of the fixation point on the object. Each saccade greatly distorts this map. (a and d) Images cast on the retina when the eye looks at different positions of the same tilted letter E. The center of the gaze is indicated by the intersection of the solid vertical and dashed horizontal lines. (b and e) Processing of the corresponding images of (a) and (d) by center-surround operators enhances contrast along edges, particularly at corners. (c and f) Simulated cortical magnification using the logarithmic-polar transformation (see text for details). (c) corresponds to the boundary images in (b) whereas (f) corresponds to those in (e). A gaze that is centered at the middle of the letter E, as in (a), activates peri-foveal areas of both hemispheres, whereas gazing at the top left corner of the letter E, as in (d), activates the left hemisphere only. For clarity, unlike human brain topography, the cortical representations are flipped upside down and foveal ends are juxtaposed.