Please cite this article in press as: Pasupathy A. The neural basis of image segmentation in the primate brain. Neuroscience (2014), http://dx.doi.org/10.1016/j.neuroscience.2014.09.051

Neuroscience xxx (2014) xxx–xxx

REVIEW

THE NEURAL BASIS OF IMAGE SEGMENTATION IN THE PRIMATE BRAIN

A. PASUPATHY *

Department of Biological Structure, Washington National Primate Research Center, University of Washington, Seattle, WA 98195, USA

Abstract—Image segmentation is a fundamental aspect of vision and a critical part of scene understanding. Our visual system rapidly and effortlessly segments scenes into component objects but the underlying neural basis is unknown. We studied single neurons in area V4 while monkeys discriminated partially occluded shapes. We found that many neurons tuned to boundary curvature maintained their shape selectivity over a large range of occlusion levels as compared to neurons that are not tuned to boundary curvature. This lends support to the hypothesis that segmentation in the face of occlusion may be solved by contour grouping.

***This article is part of a Special Issue entitled: Choice of model organisms for Neuroscience Research.* © 2014 Published by Elsevier Ltd. on behalf of IBRO.**

Key words: object recognition, shape representation, monkey, ventral pathway.

Contents

INTRODUCTION

The visual world that reaches our eyes is encoded as local contrast values in the activity patterns of retinal ganglion cells. This representation is isomorphic to the visual stimulus and continuous in that there are no demarcations for where one object ends and another begins. We nevertheless perceive the world not as a uniform pixelated representation, but as a meaningful arrangement of objects and regions. This is achieved by a process called image segmentation which takes as its input the continuous retinal representation and parses it into components that ultimately underlie the percept that is the brain's best guess for the current state of the outside world. Image segmentation facilitates scene understanding and makes our interactions with the world around us more effective. It has been shown to improve stimulus discrimination (Croner and Albright, 1999) and provides structure for deploying visual attention (Qiu et al., 2007). While we understand a great deal about how isolated stimuli are encoded in various stages of the visual processing hierarchy, very little is known about how, where, and when images are parsed into components. How scenes are segmented is one of the most important unanswered questions in vision and discovering the underlying principles will constitute a major advance in the field and could lead to better artificial vision systems. Furthermore, while it is universally accepted that feedback and recurrent processes contribute to complex brain function, the underlying mechanisms and circuitry in the visual cortex are largely unknown. In fact, there are essentially no examples of neurophysiological manipulations that can be used to control cortical feedback with the precision with which feedforward signals, driven from sensory input, can be manipulated and used to modulate neuronal responses. Because image segmentation is thought to engage feedback and recurrent processes (Kosai et al., 2014), it provides a relatively untapped opportunity to understand and manipulate cortical feedback, possibly by changing stimulus and task conditions. This could have major implications for a deeper understanding of cortical processing in general.

The approach

Segmentation is computationally challenging—even the most cutting edge machine vision systems are unable to replicate the segmentation abilities of the human visual system. To understand the neural basis of segmentation, it would be tempting to try to decode the visual cortical representations of a wide-variety of stimuli with extensive clutter and occlusions, stimulus characteristics that make segmentation a hard problem. But currently, this turns out to be an impractical strategy because the space of

---

*Address: Department of Biological Structure, University of Washington, 1959 N.E. Pacific Street, Box 357420, HSB G520, Seattle, WA 98195, USA. Tel: +1-206-685-0768.
E-mail address: pasupat@uw.edu

Abbreviations: RF, receptive field; ROC, receiver operating characteristic.

http://dx.doi.org/10.1016/j.neuroscience.2014.09.051

complex images is too large, the time available to record any given neuron in the lab is limited to brief periods due to experimental constraints, and neuronal responses of most visual cortical neurons are nonlinear functions of visual stimuli, and we do not have a good understanding of the underlying nonlinearities or the bases of representation. These constraints make it extremely difficult, if not impossible, to analytically evaluate the neuronal dynamics associated with segmentation on the basis of responses to an arbitrary set of stimuli. A more fruitful approach, in our experience, has been one of targeted hypothesis testing: we identify plausible hypotheses based on shape theory and human psychophysical literature and then focus on designing well-balanced, customized stimuli that can directly address those hypotheses. In this case, the stimulus design targets a localized region of shape space relevant to the hypotheses being tested and facilitates systematic and controlled tests that can reveal the underlying nonlinearities and representational bases. Below, we review our recent experiments (Kosai et al., 2014) to test one longstanding psychophysical hypothesis that image segmentation and subsequent recognition of partially occluded objects are achieved by contour grouping (Wertheimer, 1938).

Contour-based segmentation and primate V4

Gestalt psychologists have hypothesized that visual scenes are perceptually grouped into objects and that the component objects are detected and recognized by first grouping contours based on principles of similarity, proximity, continuity, common fate, symmetry, convexity, etc. (Wertheimer, 1938, see Wagemans et al., 2012, for review). This strategy of applying Gestalt principles to contours has been a popular tool for segmentation in computer vision (Leung and Malik, 1998). This stands in contrast to region-based segmentation, where the image is partitioned into pixel sets with coherent image properties such as brightness, color and texture (Leung and Malik, 1998)—an approach more commonly used in traditional computer vision algorithms. Depending on the specific task design, psychophysical studies lend support to contour-based strategies (Jolicoeur et al., 1986; Ben-Av et al., 1992; Houtkamp et al., 2003), region-based strategies (Fine et al., 2003) or a combination (Mumford et al., 1987).

One possible locus for contour-based segmentation in the primate brain is area V4, an intermediate stage in the ventral (i.e., form processing) pathway, where many neurons encode shape in terms of their boundary characteristics (Pasupathy and Connor, 2001). For example, a V4 neuron may respond strongly to shapes that include a sharp convexity to the lower right and weakly to shapes that do not (Fig. 1). A second neuron may respond preferentially to a set of shapes that include a concavity to the left. We have shown that a population of such neurons can provide a complete and accurate representation of two-dimensional shapes on the basis of their boundary characteristics (Pasupathy and Connor, 2002). These curvature-tuned neurons would be an ideal neural substrate for contour-based segmentation; but, because most shape tuning characterizations are conducted with isolated stimuli, we do not know whether or how these neurons contribute to segmentation. We therefore studied the responses of curvature-tuned V4 neurons as animals discriminated partially occluded shapes to determine how they might contribute to the segmentation of occluded objects.

Non-human primate model

To understand the neural basis of image segmentation, we conducted single unit studies in macaque monkeys as they performed a shape discrimination task. Our choice of animal model is informed by several factors. First, macaque monkeys are highly visual animals. Their lives in their natural habitat suggest high visual acuity and hand–eye coordination. Their visual system is comparable to that of humans in terms of visual acuity (Cavonius and Robbins, 1973) and in the manner in which they explore their environment. Monkeys and humans can easily discriminate complex images and objects that are only 2° in diameter at central fixation (e.g., Asaad et al., 1998). Monkeys are very similar to humans in their exploration of high-interest targets in scenes (Berg et al., 2009). Voluntary eye movements are qualitatively similar in humans and monkeys (Fuchs, 1967); monkeys like humans, have coordinated eye movements important for maintaining stereopsis (Schor and Tyler, 1981). Several behavioral studies in monkeys suggest that they segment visual scenes into objects and regions the way humans do (Munakata et al., 2001). Theories of segmentation, based on human psychophysics are consistent with neurophysiological studies in monkeys. Specifically, shape theory and human psychophysics suggest that T-junctions are highly informative about occlusion and that segmentation of occluded objects may originate at T-junctions

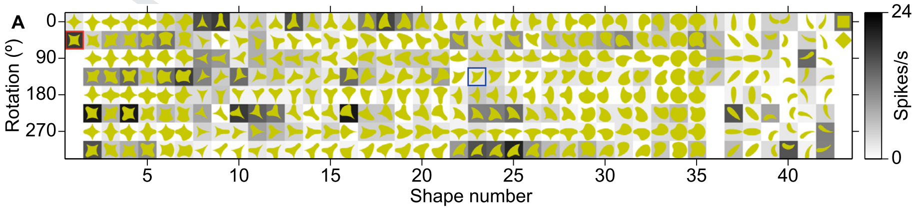

Fig. 1. Responses of a V4 neuron tuned to boundary curvature. Shape preference was characterized using a set of 43 shapes (columns) presented at 8 rotations (rows) in a passive fixation task. Some shapes (1, 36 and 43) were shown at fewer rotations due to rotational symmetry. The background intensity of each icon depicts the average response to that shape. Responses were strongest for shapes containing a sharp convexity to the lower right. Shapes highlighted by red (preferred) and blue (non-preferred) squares were chosen as the discrimination stimuli for the behavioral task (see Fig. 4). Previously published in Kosai et al. (2014).

Please cite this article in press as: Pasupathy A. The neural basis of image segmentation in the primate brain. Neuroscience (2014), http://dx.doi.org/10.1016/j.neuroscience.2014.09.051