available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainresrev

ELSEVIER

BRAIN RESEARCH REVIEWS

Review

Dorsal–ventral integration in object recognition

Reza Farivar

Department of Psychology, McGill University, Montreal, QC, Canada H3A 1B1

ARTICLE INFO

Article history:
Accepted 23 May 2009
Available online 28 May 2009

Keywords:
Parallel visual pathway
Dorsal
Ventral
Visual
Object recognition
Structure-from-motion
Face perception
Recognition
Stereopsis
Perception
View-invariance
Cue-invariance

ABSTRACT

The idea of two parallel hierarchical pathways in vision has fueled a great deal of research and enhanced our understanding of visual processing in the brain. However, after 25 years, it has become clear that the earlier distinctions in terms of neuroanatomy and functional dissociation are less pure than originally considered. Dorsal visual areas may exhibit object-selective responses and many 3-D cues of shape, particularly structure-from-motion, appear to be computed exclusively by dorsal areas. These findings imply a more important role for dorsal visual areas in object recognition than previously considered and also place restrictions on the nature of ventral object representations. These representations will need to include information about the objects in 3-D, making them more viewpoint-invariant. They will also need to be invariant to the 3-D cue used to describe them. Through the discussion of relevant findings in psychophysics, single-unit electrophysiology, neuroanatomy and functional imaging, I suggest that these qualities are indeed present in ventral stream representations. Thus dorsal visual areas that extract 3-D structure of shapes from certain cues, can relate these representations to cue-invariant and view-invariant representations in the ventral stream.

Contents

Fax: +1 514 398 3255.
E-mail address: Reza.farivar@mail.mcgill.ca.

0165-0173/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainresrev.2009.05.006

1. Introduction

In 1982, an idea was presented that dramatically influenced thinking about the primate visual system. Ungerleider and Mishkin (1982), based on the pattern of behaviour following lesions to dorsal (occipito-parietal) and ventral (occipito-temporal) regions of the monkey cortex, suggested that the visual cortex can be decomposed into two pathways—a dorsal pathway concerned with spatial properties of vision (answering the question "where?") and the ventral pathway concerned with identification of the visual objects (answer the question "what?"). However, after 25 years, many challenges have been raised to that original elegant and simple view (Merigan and Maunsell, 1993) (Hegde and Felleman, 2007), and an alternative description of the two pathways exists in terms of vision for perception (ventral stream) and vision for action (dorsal stream; Goodale and Milner, 1992). While the original model and its variant still serve as useful paradigms for interpreting results from psychophysics, neurophysiology, neuroanatomy, neuropsychology, and functional imaging, they are still evolving to incorporate newer findings. The objective of this article is to highlight a number of studies that together suggest the two pathways are functionally integrated in normal object recognition to enhance cue-invariant and viewpoint-invariant recognition by use of 3-D information. This may at first appear to contradict the original ideas of Ungerleider and Mishkin (1982) or those of Goodale and Milner (1992), but at closer inspection it will be evident that normal object recognition and all the variable viewing conditions that may challenge it necessitate the integrative action of these two streams.

First, the discussion will focus on the nature of object recognition in the ventral stream. It will be suggested that ventral object representations are largely viewpoint-invariant, although this invariance may not be represented at the single-cell level. Additionally, it will be suggested that familiarity with objects drives the development of representations that are more viewpoint-invariant. Finally, given that expertise with an object class requires extensive knowledge and familiarity with many members of the class, it is suggested that categories of objects that one has developed expertise in have a greater facility at achieving viewpoint-invariant representations for individual members of that class.

Second, the case is presented for dorsal–ventral integration in object recognition. Considering the primary discussion of viewpoint-invariant representations in the ventral pathway, it is suggested that shapes defined by 3-D cues that are dorsally extracted (particularly structure-from-motion) are ultimately processed by ventral stream mechanisms for recognition.

Through these two syntheses, it is proposed that normal object recognition likely requires the integrative action of the dorsal and ventral streams. This leads to several conjectures as to the properties of ventral stream representations, such as their invariance with respect to 3-D depth cues.

2. Distinctions in object recognition models

Models of visual object recognition can be divided along multiple, orthogonal dichotomies. The grandest dichotomy is between models that assume viewpoint-invariance in the neural representation of objects, and those that assume that viewpoint-invariant effects can be explained by uses of multiple individual viewpoints in an image-based manner. In the latter case, the brain interpolates intermediate views and thus allows us to recognize known objects from novel angles (Riesenhuber and Poggio, 2000). The viewpoint-invariant models suggest that the brain builds a structural representation of objects from available views, and this structural representation, analogous to a 3-D model, may be used to recognize the seen objects from novel views (Biederman, 1987; Marr and Nishihara, 1978). While there is support for both models, the viewpoint-dependent models have the upper hand in explaining the vast majority of data obtained on representations of complex shapes, but in general, many agree that a combination of structural and image-based descriptions is necessary for normal object recognition (see Peissig and Tarr, 2007 for a review).

2.1. *Category-level and subordinate-level recognition*

An important aspect to consider in evaluating models of object recognition is to what extent they explain category-level and subordinate-level recognition performance. The human object recognition system must not only recognize objects as belonging to specific categories, such as "cat", "chair", "car", etc., but must also be able to recognize individuals within that category—my neighbour's cat, my car, etc. An area of object recognition research that informs us best about this aspect of object recognition is that of face recognition, a within-category type of object recognition that is essential to normal human engagement.

It is often argued that faces are processed differently than other objects, but it is unclear whether this difference is due to faces being processed by a specific "module" (Kanwisher et al., 1997), or by a general purpose visual object expertise system (Gauthier et al., 2000). We are experts at recognizing faces, because this is a skill that is essential for our normal social interactions. Some of the effects observed uniquely for faces can also be observed for objects with which one has developed some expertise (Tarr and Cheng, 2003). While a number of reports do suggest that some cortical areas, such as the middle fusiform and inferior occipital gyri, may be uniquely involved in the processing of faces (Grill-Spector et al., 2004; Kanwisher et al., 1997; Op de Beeck et al., 2006), some data exists to also suggest that at least some aspect of the observed face selectivity in these regions may also be observed for objects that one has developed expertise with (Gauthier et al., 2000; Xu, 2005). It is still a matter of considerable debate whether the