Available online at www.sciencedirect.com
ScienceDirect
Vision Research 46 (2006) 3994–4006
Vision Research
www.elsevier.com/locate/visres

Spatial vs temporal continuity in view invariant visual object recognition learning

Gavin Perry, Edmund T. Rolls *, Simon M. Stringer

Oxford University, Centre for Computational Neuroscience, Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, UK

Received 8 February 2006; received in revised form 22 June 2006

Abstract

We show in a 4-layer competitive neuronal network that continuous transformation learning, which uses spatial correlations and a purely associative (Hebbian) synaptic modification rule, can build view invariant representations of complex 3D objects. This occurs even when views of the different objects are interleaved, a condition where temporal trace learning fails. Human psychophysical experiments showed that view invariant object learning can occur when spatial but not temporal continuity applies because of interleaving of stimuli, although sequential presentation, which produces temporal continuity, can facilitate learning. Thus continuous transformation learning is an important principle that may contribute to view invariant object recognition.

Keywords: Object recognition; Continuous transformation; Trace learning; Inferior temporal cortex

1. Introduction

There is now much evidence demonstrating that over successive stages the ventral visual system develops neurons that respond to objects or faces with view, size, and position (translation) invariance (Rolls, 1992, 2000, 2006; Rolls & Deco, 2002; Desimone, 1991; Tanaka, Saito, Fukada, & Moriya, 1991). For example, it has been shown that the macaque inferior temporal visual cortex has neurons that respond to faces and objects with invariance to translation (Tovee, Rolls, & Azzopardi, 1994; Kobotake & Tanaka, 1994; Ito, Tamura, Fujita, & Tanaka, 1995; Op de Beeck & Vogels, 2000; Rolls, Aggelopoulos, & Zheng, 2003), size (Rolls & Baylis, 1986; Ito et al., 1995), contrast (Rolls & Baylis, 1986), lighting (Vogels & Biederman, 2002), spatial frequency (Rolls, Baylis, & Leonard, 1985; Rolls, Baylis, & Hasselmo, 1987), and view (Hasselmo, Rolls, Baylis, & Nalwa, 1989; Booth & Rolls, 1998). It is crucially important that the visual system builds invariant representations, for only then can one-trial learning about an object generalize usefully to other transforms of the same object (Rolls & Deco, 2002). Building invariant representations of objects is a major computational issue, and the means by which the cerebral cortex solves this problem is a topic of great interest (Riesenhuber & Poggio, 1999; Biederman, 1987; Ullman, 1996; Rolls & Deco, 2002).

One proposed method for the learning of invariance in the visual system is to utilize the temporal continuity of objects in the visual environment (over short time periods) to help the learning of invariant representations (Földiák, 1991; Rolls, 1992; Wallis & Rolls, 1997; Rolls & Milward, 2000; Rolls & Stringer, 2001). Temporal continuity can be utilized by, for example, associative learning rules that incorporate a temporal trace of activity in the post-synaptic neuron (Földiák, 1991; Rolls, 1992; Wallis & Rolls, 1997). These rules encourage neurons to respond to input patterns that occur close together in time, which, given the natural statistics of the visual world, are likely to represent different transforms (views) of the same object. Temporal continuity is also a feature of other proposals (Stone, 1996; Bartlett

* Corresponding author. Fax: +44 1865 310447.
E-mail address: Edmund.Rolls@psy.ox.ac.uk (E.T. Rolls).
URL: www.cns.ox.ac.uk (E.T. Rolls).

0042-6989/$ - see front matter
doi:10.1016/j.visres.2006.07.025

& Sejnowski, 1998; Becker, 1999; Einhäuser, Kayser, König, & Körding, 2002; Wiskott & Sejnowski, 2002).

Recently, spatial continuity in the different views of a transforming object has been proposed as another principle of invariance learning (Stringer, Perry, Rolls, & Proske, 2006). In continuous transformation (CT) learning a competitive network using an associative synaptic modification rule learns to respond to an initial view of an object, and then similar views activate the same post-synaptic neuron through the strengthened synapses. As the object transforms continuously, the different views become associated onto the same post-synaptic neurons, as illustrated in Fig. 2. The CT learning effect can operate even when there are large separations of time between the presentation of views of the same object, and even if views of different stimuli are presented during this intervening time period in an interleaved training condition (Stringer et al., 2006). Spatial continuity in the context of continuous transformation learning is the property that the different views of an object are sufficiently similar that after one view has been learned, an adjacent view will have sufficient overlap of the active inputs to activate the same neuron, as illustrated in Fig. 2. In topologically mapped systems, these adjacent (overlapping) inputs will be spatially close, but need not be in a non-topologically mapped system.

In this paper, we compare computer simulations with psychophysical studies using the same set of stimuli to investigate the relative contributions of temporal continuity and spatial continuity in the learning of view invariant representations of objects in the brain.

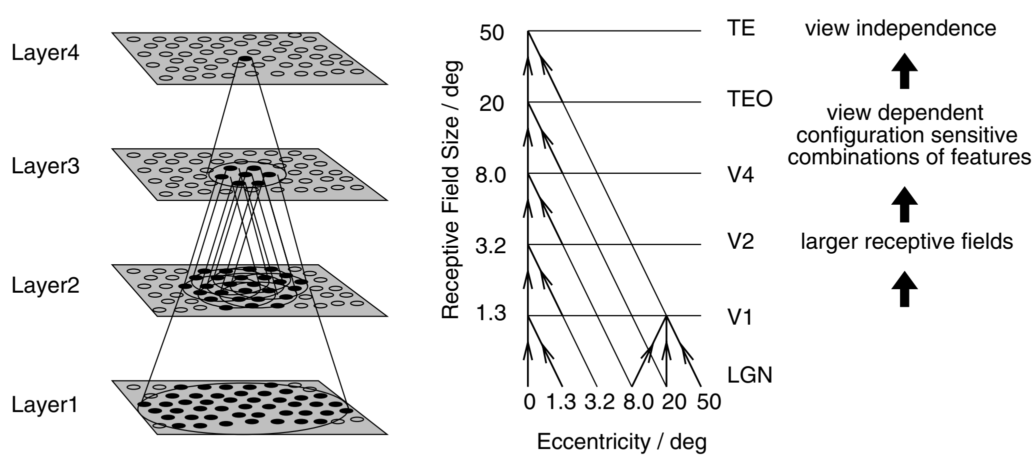
First, we test how closely predictions of the temporal vs spatial continuity theories are met in a hierarchical model of the ventral visual stream, VisNet (Wallis & Rolls, 1997; Rolls & Milward, 2000) illustrated in Fig. 1, in which the parameters can be precisely controlled. Use of the model helps to show the type of result expected if the system is trained with temporal trace vs spatial continuous transformation paradigms.

We then use the same realistically difficult set of objects in a psychophysical experiment with humans to investigate whether humans' learning reflects the use of short-term temporal correlations vs spatial continuity in the different transforms of each object.

2. Methods

In order to investigate the roles of temporal vs spatial continuity in human invariance learning in the context of the temporal trace and continuous transformation theories of invariance learning, human performance and a network model trained with the same set of stimuli, were compared with a range of training stimulus presentation paradigms. Key predictions of the continuous transformation (CT) vs temporal trace theories tested are that CT but not temporal trace learning can self-organize invariant representations when the views of different objects are interleaved, and that CT learning but not necessarily temporal trace learning will perform poorly if the spacing between the closest views become larger, thus breaking the spatial continuity in the images seen by the network.

In the 'interleaved' training condition, an initial view of the first object was shown, followed by an initial view of the second object and then an initial view of each of the remaining objects in order. Once a view of each object had been shown the next image in clockwise (viewed from above) sequence of the first object was presented followed by the next image in sequence of the second object, then the third and so on. This procedure ensured that in the interleaved condition two views of the same object did not occur close together in time. It is a prediction of the temporal trace hypothesis (and any model that uses temporal continuity to learn invariance) that training in this manner should cause invariance learning to be impaired (as views of different objects could become associated together



	Dimensions	# Connections	Radius
Layer 1	32x32	100	12
Layer 2	32x32	100	9
Layer 3	32x32	100	6
Layer 4	32x32	272	6
Retina	128x128x32	-	-

Fig. 1. (Left) Schematic diagram of the four layer hierarchical competitive network, VisNet. Convergence through the network is designed to provide fourth layer neurons with information from across the entire input retina. (Right) Convergence in the visual system. V1, visual cortex area V1; TEO, posterior inferior temporal cortex; TE, inferior temporal cortex (IT). (Bottom) Network dimensions showing the number of connections per neuron and the radius in the preceding layer from which 67% are received.