Binocular Fusion and Invariant Category Learning of Natural Objects due to Predictive Remapping during Scanning of a Depthful Scene with Eye Movements

Karthik Srinivasan, Stephen Grossberg, & Arash Yazdanbaksh
Center for Adaptive Systems, Graduate Program in Cognitive and Neural Systems,
Center for Computational Neuroscience and Neural Technology, and
Department of Mathematics
Boston University, 677 Beacon Street, Boston, MA 02215
skarthik@bu.edu, steve@bu.edu, yazdan@bu.edu

How does the brain maintain stable fusion of 3D scenes when the eyes move? Every eye movement causes each retinal position to process a different set of scenic features, and thus the brain needs to binocularly fuse new combinations of features at each position after an eye movement. Despite these breaks in retinotopic fusion due to each movement, previously fused representations of a scene in depth appear stable. This is illustrated by moving the eyes after fusing binocular or monocular ("Magic Eye") stereograms. The 3D ARTSCAN neural model proposes how the brain does this by unifying concepts about how multiple cortical areas in the What and Where cortical streams interact to carry out 3D boundary and surface perception, spatial attention, invariant object category learning, predictive eye movements, and learned coordinate transformations. Data from single neuron studies and also from psychophysical studies of covert visual attention shifts prior to eye movements (Cavanagh et al., 2010; Duhamel and Goldberg, 1992; Gottlieb, 1992; Gottlieb and Snyder, 2010; Melcher, 2007; Rolfs et al., 2011) are clarified by the model. The model also proposes how perceptual, attentional, and cognitive interactions among multiple brain regions (e.g., LGN, V1, V2, V3A, V4, MT, MST, PPC, LIP, ITp, ITa, SC) may accomplish predictive mapping as part of the process whereby view-invariant object categories are learned. This model builds upon the FACADE neural model of 3D vision and figure-ground separation (e.g., Fang and Grossberg, 2009; Grossberg 1994; Grossberg and Yazdanbakhsh, 2005; Kelly and Grossberg, 2000) and the ARTSCAN and pARTSCAN models of how invariant object categories are learned as the eyes freely scan a scene (Cao, Grossberg, and Markowitz, 2011; Fazl, Grossberg, and Mingolla, 2009; Grossberg, 2007, 2009). One key process concerns how an attended object's surface representation generates a form-fitting distribution of spatial attention, or attentional shroud, in parietal cortex that enables invariant object category learning to occur as the eyes freely scan a scene. Another key set of processes concern how predictive remapping signals, mediated by gain fields acting at multiple levels of visual cortex, enable binocular fusion of attended objects to be preserved, and the shroud of the attended object to remain stable, during saccadic eye movements that explore the object's surface. These competences are tested on objects from the Caltech 101 database.
Supported in part by CELEST, an NSF Science of Learning Center (NSF SBE-0354378) and by the SyNAPSE program of DARPA (HR0011-09-C-0011)