Neural Networks 22 (2009) 1383–1398

Contents lists available at ScienceDirect

Neural Networks

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

Cortical dynamics of navigation and steering in natural scenes: Motion-based object segmentation, heading, and obstacle avoidance☆

N. Andrew Browning, Stephen Grossberg*, Ennio Mingolla

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

ARTICLE INFO

Article history:
Received 4 December 2008
Received in revised form 7 May 2009
Accepted 18 May 2009

Keywords:
Optic flow
Navigation
MT
MST
Motion segmentation
Object tracking
Steering

ABSTRACT

Visually guided navigation through a cluttered natural scene is a challenging problem that animals and humans accomplish with ease. The ViSTARS neural model proposes how primates use motion information to segment objects and determine heading for purposes of goal approach and obstacle avoidance in response to video inputs from real and virtual environments. The model produces trajectories similar to those of human navigators. It does so by predicting how computationally complementary processes in cortical areas MT^-/MSTv and MT^+/MSTd compute object motion for tracking and self-motion for navigation, respectively. The model's retina responds to transients in the input stream. Model V1 generates a local speed and direction estimate. This local motion estimate is ambiguous due to the neural aperture problem. Model MT^+ interacts with MSTd via an attentive feedback loop to compute accurate heading estimates in MSTd that quantitatively simulate properties of human heading estimation data. Model MT^- interacts with MSTv via an attentive feedback loop to compute accurate estimates of speed, direction and position of moving objects. This object information is combined with heading information to produce steering decisions wherein goals behave like attractors and obstacles behave like repellers. These steering decisions lead to navigational trajectories that closely match human performance.

1. Introduction

The ViSTARS (Visually-guided Steering, Tracking, Avoidance, and Route Selection) model demonstrates how the primate magnocellular pathway may generate sufficient information for reactive navigation, route selection, and target tracking tasks (Fig. 1). When immersed in a realistic visual world, ViSTARS is capable of human-like steering behaviors towards goals and around obstacles in response to realistic visual scenes.

ViSTARS is a synthesis and further development of two previous models: The STARS model of Elder, Grossberg, and Mingolla (2005, in press) is capable of reactive steering towards goals and around obstacles, and accurately simulates human navigational data of Fajen and Warren (2003), among others. However, the STARS model did not directly process visual scenes. Rather, it used the Longuet-Higgins and Prazdny (1980) equations that describe the scenic geometry as model inputs. Browning, Mingolla, and Grossberg (in press) showed how ViSTARS could build upon STARS to directly process visual data, notably virtual world animations and driving video sequences of realistic visual scenes, as well as random dot displays, to compute accurate heading, or direction of travel (Hildreth, 1992), estimates at human-like accuracies. To accomplish this, the motion processing front end of ViSTARS adapted a biological motion perception model, called the 3D FORMOTION model, that has been progressively developed to explain and predict large perceptual and neurobiological data bases about motion perception (Baloch & Grossberg, 1997; Berzhanskaya, Grossberg, & Mingolla, 2007; Chey, Grossberg, & Mingolla, 1997; Grossberg, Mingolla, & Viswanathan, 2001). The current model adapts the cortical motion processing stream in the 3D FORMOTION model, but not its form-to-motion interactions, which are important when objects are incompletely defined and partially occlude one another. The current extension of ViSTARS shows, in addition, how heading estimates can be joined to STARS navigational mechanisms to achieve reactive navigation and object tracking estimates in response to realistic visual scenes.

This synthesis clarifies how the brain exploits computationally *complementary* processes for navigation and object tracking, respectively (Grossberg, 2000). As will be seen in greater detail below, the processing stream through cortical areas MT^+/MSTd is specialized for visually based navigation, whereas the parallel

☆ CAS/CNS Technical Report CAS/CNS-TR-08-007.

* Corresponding author at: Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA 02215, United States. Tel.: +1 617 353 7858/7; fax: +1 617 353 7755.
E-mail address: steve@bu.edu (S. Grossberg).

0893-6080/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2009.05.007

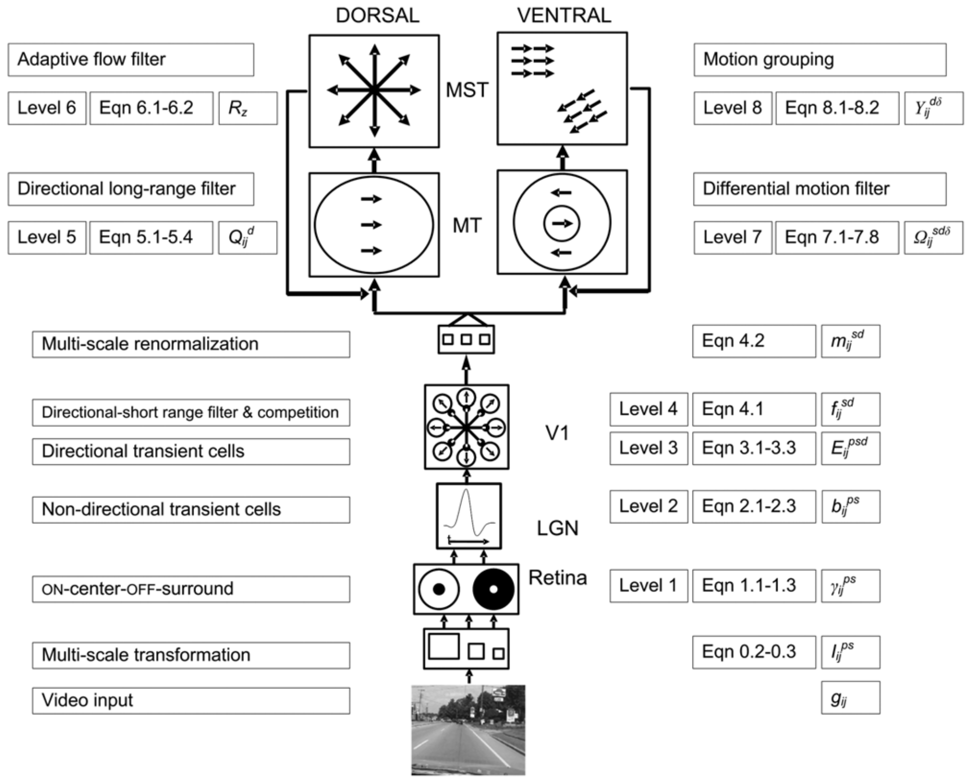

Fig. 1. Model overview: pictorial representation of each processing stage coupled with functional description, model level number, corresponding equation number, and output variable labels. See text for model description.

processing stream through cortical areas MT^-/MSTv is specialized for visual tracking of moving objects. In particular, navigating a body moving with respect to the world uses *additive* processing, whereas tracking an object moving with respect to that body uses *subtractive* processing.

Processing in cortical areas MT^+ and MSTd is described, and the ViSTARS model performance analyzed with respect to human heading data, in Browning et al. (in press). The current article focuses on how the cortical areas MT^- and MSTv estimate object position and motion, and additionally demonstrates how interactions of heading estimates from model MT^+ and MSTd, and object position estimates from model MT^- and MSTv, are sufficient to allow the STARS (Elder et al., in press) steering component in ViSTARS to steer around obstacles towards goals in a human-like fashion.

By showing navigational competence in realistic settings, ViSTARS provides an example of how real-time adaptive control systems can accomplish visually-based autonomous robotic navigation. By linking identified brain regions and cell types to navigational behaviors, the model clarifies how the brain accomplishes visually-based navigation and object tracking. Previous models typically contribute to one of these goals, but not both.

ViSTARS processing levels correspond to brain regions from retina through cortical areas V1, MT, and MST. Before describing the model, a short summary of pertinent experimental data will be given.

Neurophysiology. The early primate visual system consists of two distinct pathways: the parvocellular (P) pathway is concerned with high resolution, color, static information, whereas the magnocellular (M) pathway is concerned with low resolution, monochromatic, transient information (Kandel, Schwartz, & Jessell, 2000). The parvocellular pathway processes object form and identity in the What, or ventral, cortical processing stream. The magnocellular pathway processes motion, object location, and action in the Where, or dorsal, cortical processing stream (Mishkin, Ungerleider, & Macko, 1983; Schneider, 1967). Although form processing influences motion perception and navigational tasks through form-motion interactions (Baloch & Grossberg, 1997; Berzhanskaya et al., 2007; Grossberg et al., 2001; Ponce, Lomber, & Born, 2008), the results reported herein focus on the magnocellular pathway.

Magnocellular retinal cells respond with a burst of activation when presented with a step input (Benardete & Kaplan, 1999; Cleland, Dubin, & Levick, 1971; Enroth-Cugell & Robson, 1966; Kaplan & Benardete, 2001; Valois, Albrecht, & Thorell, 1982). M pathway retinal cells project to lateral geniculate nucleus (LGN) layers 1 and 2 and then to primary visual cortex (V1) (Callaway, 2005). V1 cells are directionally selective, responding more vigorously to motion in a preferred direction at a preferred speed, and are disparity selective (Hubel & Wiesel, 1959, 1962, 1968; Livingstone, 1998; Livingstone & Hubel, 1987; Schiller, Finlay, & Volman, 1976).

V1 projects to area MT (middle temporal cortex, or V5) which, in turn, projects to area MST (medial superior temporal cortex) (Albright, 1984; Born & Bradley, 2005). Cells in MT respond preferentially to motion in a particular direction within a range of speeds and depths (Albright, 1984; Born & Bradley, 2005). Macaque MT has two main sub-divisions: MT^+ consists of cells with large *additive* receptive fields that project primarily to dorsal MST (MSTd); MT^- consists of cells with *subtractive* (that is, ON-center OFF-surround opponent-motion) receptive fields that project primarily to ventral MST (MSTv) (Allman, Miezin, & McGuinness, 1985; Born, 2000; Born & Tootell, 1992). The MT^+/MSTd stream carries out visually-guided navigation, including heading estimation (Born & Tootell, 1992; Duffy, 1998; Duffy & Wurtz, 1995, 1997). The MT^-/MSTv stream carries out object-based segmentation and tracking (Born & Tootell, 1992; Duffy, 1998). In order to realize their complementary tracking and navigation functions, ventral and dorsal MST cells have different