Available online at www.sciencedirect.com

SCIENCE DIRECT®

Computer Vision and Image Understanding 100 (2005) 3–40

Computer Vision and Image Understanding

www.elsevier.com/locate/cviu

Attending to visual motion

John K. Tsotsos [a,b,*], Yueju Liu [a,b], Julio C. Martinez-Trujillo [c], Marc Pomplun [d], Evgueni Simine [a,b], Kunhao Zhou [a,b]

[a] *Department of Computer Science and Engineering, York University, Toronto, Canada*
[b] *Centre for Vision Research, York University, Toronto, Ont., Canada M3J 1P3*
[c] *Department of Physiology, McGill University, Montreal, Canada*
[d] *Department of Computer Science, University of Massachusetts, Boston, MA, USA*

Received 6 February 2004; accepted 5 October 2004
Available online 26 July 2005

Abstract

Visual motion analysis has focused on decomposing image sequences into their component features. There has been little success at re-combining those features into moving objects. Here, a novel model of attentive visual motion processing is presented that addresses both decomposition of the signal into constituent features as well as the re-combination, or binding, of those features into wholes. A new feed-forward motion-processing pyramid is presented motivated by the neurobiology of primate motion processes. On this structure the Selective Tuning (ST) model for visual attention is demonstrated. There are three main contributions: (1) a new feed-forward motion processing hierarchy, the first to include a multi-level decomposition with local spatial derivatives of velocity; (2) examples of how ST operates on this hierarchy to attend to motion and to localize and label motion patterns; and (3) a new solution to the feature binding problem sufficient for grouping motion features into coherent object motion. Binding is accomplished using a top-down selection mechanism that does not depend on a single location-based saliency representation.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Attention; Visual motion analysis; Feature binding; Selective tuning; Affine motion

* Corresponding author. Fax: +1 416 736 5857.
E-mail address: tsotsos@cs.yorku.ca (J.K. Tsotsos).
URL: http://www.cs.yorku.ca/~tsotsos (J.K. Tsotsos).

1077-3142/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2004.10.011

1. Introduction

The Selective Tuning model is a proposal for the explanation at the computational and behavioral levels of visual attention in humans and primates. Key characteristics of the model, all previously detailed in [1,2] include: (1) a top-down coarse-to-fine winner-take-all (WTA) selection process, (2) a unique WTA formulation with provable convergence properties, (3) a WTA that is based on region rather than point selection, (4) a task-relevant inhibitory bias mechanism, (5) selective inhibition in both spatial and feature dimensions for elimination of signal interference that leads to a suppressive surround for attended items, and (6) a task-specific executive controller. These characteristics lead to an extensive set of biological predictions many of which have now been supported by experiment. The bulk of the paper will focus on attention to visual motion. Past work will be summarized showing how this is not a well-studied issue. A new model of motion processing is presented and it is demonstrated how ST operates on this representation, with no changes to its previously described definition. In this way three points are made: first, that the weaknesses of previous demonstrations of ST have been remedied; second, that the original statement of ST has generality for a wide variety of visual processing representations; and third, examples of how feature binding can be solved using ST for complex motion patterns.

It had been suggested that previous demonstrations of the Selective Tuning model were neither biologically plausible nor very useful. In order to demonstrate that ST can indeed operate with realistic representations, the motion domain is chosen because enough is known about motion processing to enable a reasonable attempt at defining the feed-forward pyramid. Moreover, the effort is unique because it seems that no past model has presented a motion hierarchy plus attention to motion [3–12].

The layout of the remainder of this presentation is as follows. The next section will detail the feed-forward motion-processing network. Earlier versions of this network appear in [13,14]. Following this, an overview of ST is provided because this structure is imposed upon the feed-forward network. ST has been detailed several times in the past; here only a brief presentation is given and the reader is referred to [1,15,2,16–18] for further details. Section 4 will show several examples of the operation of the entire network including feed-forward and feedback components as well as a new solution to the feature-binding problem. A concluding discussion rounds out the paper.

2. Feed-forward motion processes

The motion representations and processes that are modeled are informed by current knowledge of motion analysis in the primate cortex. Although the literature is large on the topic, selected experimental observations are used here in order to simplify the models. It is generally accepted that motion processing in the monkey cortex goes through a series of stages, with neural representations in areas V1, MT,

Download English Version:

https://daneshyari.com/en/article/9669557

Download Persian Version:

https://daneshyari.com/article/9669557

Daneshyari.com