Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Computer Vision and Image Understanding 100 (2005) 64–106
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

The emergence of attention by population-based inference and its role in distributed processing and cognitive control of vision

Fred H. Hamker *

Allgemeine Psychologie, Psychologisches Institut II, Westf. Wilhelms-Universität, Fliednerstrasse 21, 48149 Münster, Germany

Received 8 January 2004; accepted 29 September 2004
Available online 21 July 2005

Abstract

Technologies such as video surveillance and vision guided robotics require flexible vision systems that interpret the scene according to the current task at hand. Attention has been suggested to play an important role in the process of scene understanding by prioritizing relevant information. However, the underlying processes that allow cognition to guide vision have not been fully explored. Our procedure has its origin in current findings of research in attention. We suggest an approach in which high-level cognitive processes are top-down directed and modulate stimulus signals such that vision is a constructive process in time. Prior knowledge is combined with the observation taken from the image by a population-based inference in order to dynamically update the conspicuity of each feature. Any decision, such as object detection, is based on these distributed conspicuities. We demonstrate this concept on a goal-directed object detection task in natural scenes.

Keywords: Attention; Natural scenes, Object detection; Object recognition; Cognitive control; Top-down inference; Computational neuroscience

* Fax: +49 251 8334180.
E-mail address: fhamker@uni-muenster.de.

1077-3142/$ - see front matter
doi:10.1016/j.cviu.2004.09.005

1. Introduction

Emerging new technologies require vision systems to flexibly focus on the relevant information in a visual scene. The processing of a full scene in parallel up to a high level description has turned out to be problematic and not fully successful in computer vision [1–3]. Attention might be the solution. The idea is to control the information flow and thus to improve vision by focusing the resources merely on aspects of the whole visual scene. Early, preattentive, and parallel vision modules compute a set of basic features from the scene, which are then attentively integrated and further processed. Such attentive processing has been described as a 'spotlight of attention' [4] that highlights an area of interest by routing that information into higher areas for further processing [5–7]. The guidance of an attentional focus can be implemented by a winner-takes-all process within a saliency map which indicates potentially relevant locations [6]. Based on this paradigm, sophisticated models of information control have been developed, in which the complex problem of scene understanding is transferred into a sequential analysis of image parts. Such spatial selection is computationally efficient [1,2]. However, we have to consider other crucial issues of efficiency as well. First of all, selection should be effectively guided by the task at hand. It would be problematic if we had to scan several salient items before focusing the relevant item. Thus, we have to elaborate mechanisms that integrate high level knowledge into the selection process. Second, the mechanisms of attention must result in a representation that facilitates further processing. For example, object recognition in natural scenes would not benefit a lot if we simply determined a point in space by some competitive mechanism. Even a region of interest can be problematic if it does not sufficiently cover the object of interest. Thus, we need forms of selection that enhance the features of an object in space.

The most crucial issue of attention deals with the integration of information from different modules (or brain areas). This could be implemented as a central process that collects the information from different modules and then controls those by a single attentional signal. For example in the Guided Search framework [8] a bottom-up map is combined with a top-down map in order to determine the activity in a master map of locations. The location of the highest activity could then be used to control a single attentional focus. An alternative has been outlined by the integrated competition hypothesis [9], in which different specialized modules (or brain areas) have to coordinate themselves to let a distributed system operate on the same event. We present an approach that follows this idea. At its core is a population code that encodes in a dual coding principle a feature and its respective conspicuity. The term conspicuity here reflects stimulus-driven saliency as well as task relevance and relates to the probability that a feature is present in the scene. We developed a population-based inference approach to continuously update the conspicuity using prior knowledge in form of generated expectations.

The idea is that all mechanisms act directly on the processed variables and modify their conspicuity. Attending a certain feature or a region in space enhances the probability of a feature being detected. In this respect, attention emerges in the vision process in order to serve in a flexible manner the needs of the task at hand.