Chapter 1
Influence of Movement Expertise on Visual Perception of Objects, Events and Motor Action: A Modeling Approach

Kai Essig
Bielefeld University, Germany

Oleg Strogan
Bielefeld University, Germany

Helge Ritter
Bielefeld University, Germany

Thomas Schack
Bielefeld University, Germany

ABSTRACT
Various computational models of visual attention rely on the extraction of salient points or proto-objects, i.e., discrete units of attention, computed from bottom-up image features. In recent years, different solutions integrating top-down mechanisms were implemented, as research has shown that although eye movements initially are solely influenced by bottom-up information, after some time goal driven (high-level) processes dominate the guidance of visual attention towards regions of interest (Hwang, Higgins & Pomplun, 2009). However, even these improved modeling approaches are unlikely to generalize to a broader range of application contexts, because basic principles of visual attention, such as cognitive control, learning and expertise, have thus far not sufficiently been taken into account (Tatler, Hayhoe, Land & Ballard, 2011). In some recent work, the authors showed the functional role and representational nature of long-term memory structures for human perceptual skills and motor control. Based on these findings, the chapter extends a widely applied saliency-based model of visual attention (Walther & Koch, 2006) in two ways: first, it computes the saliency map using the cognitive visual attention approach (CVA) that shows a correspondence between regions of high saliency values and regions of visual interest indicated by participants’ eye movements (Oyekoya & Stentiford, 2004). Second, it adds an expertise-based component (Schack, 2012) to represent the influence of the quality of mental representation structures in long-term memory (LTM) and the roles of learning on the visual perception of objects, events, and motor actions.

DOI: 10.4018/978-1-4666-2539-6.ch001
INTRODUCTION

We evaluated our modeling approach by investigating a simple task, where participants had to look at and grasp 12 known and unknown objects. Unlike existing approaches our model can adapt to differences in participants’ gaze behavior that results from better LTM created through a learning interactive phase. Knowledge about the cognitive and learning principles of action-based perception and the selection process of action relevant information from the steady flow of ongoing events is of great importance for the establishment of biologically inspired visual systems and the development of humanoid robots and intelligent systems.

We live in a dynamic environment and have multimodal information inflow in the form of seeing, hearing, and haptic contact. Since the human brain is too limited to process all this information, we have to focus our attention to scene relevant details – like the spotlight in a theatre (Posner, 1980). When reorienting the gaze to a new location, the focus of attention first has to be disengaged from the current location before it can be shifted towards the new location (Vickers, 2007). The scene is then explored by successively directing the focus to the relevant areas (Frintrop, Rome & Christensen, 2010; Duchowski, 2007). Through eye movements, humans can control the duration and the temporal and spatial order of fixations and thus, which image regions fall into the foveal field. The order in which a scene is investigated is determined by the mechanisms of selective attention (Frintrop, Rome and Christensen, 2010). When perceiving static objects, horizontal and vertical eye movements have an amplitude between 1 and 60 minutes of arc. During a fixation the eye is not completely still - there are different types of micro movements to move the image on the retina by several receptors, providing the photoreceptive cells with a constant flow of new stimulus (Holmqvist et al., 2011; Martínez-Conde, Macknik, & Hubel, 2004; Rötting, 2001).

Although the main components of a scene can be relatively quickly processed and an object can also be recognized in the periphery, a close inspection of an object requires a shift of attention towards it: the focus of attention is directed toward the region of interest, followed by a gaze shift enabling the further perception at higher resolutions (Frintrop, Rome & Christensen, 2010). Deubel and Schneider (1996) argue for an obligatory and selective coupling of saccade programming and visual attention to one common target object. In this context it is worth mentioning that humans are able to attend simultaneously to multiple regions of interest, usually between 4 and 5 regions (McMains & Somers, 2004). When resources are shared, for example in mobile robots, focusing on the relevant data is even more important than scene viewing (Frintrop, Rome & Christensen, 2010). Different modules have to be flexibly prioritized and coordinated in order to fulfill the respective needs of the mobile robots. This fact becomes even more prominent considering the fact that nowadays many robots are expected to act in complex environments and have to interact with human partners. In order to cope with these requirements, computational systems have been developed over the past 5-10 years to investigate how the concepts of human selection mechanisms can be exploited for object recognition, robot localization, and human-machine interaction (Frintrop, Rome & Christensen, 2010).

In order to enhance our understanding of the complex visual processes, a quasi-formal model (i.e., algorithmic descriptions of the suggested interpretations of the experimental data) can be constructed that attempts to simulate the empirical observations and thus can help us to better understand the complex processes underlying human behavior. The comparison of the results given by the empirical data and the model reveals information about the adequacy of the computational simulation. Computer simulations can usually be parameterized in order to test the influence and the interactions of the different parameters on the
Related Content

Fuzzy Clustering-Based Approaches in Automatic Lip Segmentation from Color Images

A Comparative Study on Diabetic Retinopathy Detection Using Texture-Based Feature Extraction Techniques

A Feature Selection Approach for Network Intrusion Classification: The Bi-Layer Behavioral-Based

Line Parameter based Word-Level Indic Script Identification System