Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computer Vision and Image Understanding
Computer Vision and Image Understanding 100 (2005) 107–123
www.elsevier.com/locate/cviu

Assessing the contribution of color in visual attention

Timothée Jost [a], Nabil Ouerhani [a,*], Roman von Wartburg [b], René Müri [b], Heinz Hügli [a]

[a] Institute of Microtechnology, University of Neuchâtel, CH-2000 Neuchâtel, Switzerland
[b] Department of Neurology, University of Bern, Inselspital, CH-3010 Bern, Switzerland

Received 15 January 2004; accepted 17 October 2004
Available online 23 May 2005

Abstract

Visual attention is the ability of a vision system, be it biological or artificial, to rapidly detect potentially relevant parts of a visual scene, on which higher level vision tasks, such as object recognition, can focus. The saliency-based model of visual attention represents one of the main attempts to simulate this visual mechanism on computers. Though biologically inspired, this model has only been partially assessed in comparison with human behavior. Our methodology consists in comparing the computational saliency map with human eye movement patterns. This paper presents an in-depth analysis of the model by assessing the contribution of different cues to visual attention. It reports the results of a quantitative comparison of human visual attention derived from fixation patterns with visual attention as modeled by different versions of the computer model. More specifically, a one-cue gray-level model is compared to a two-cues color model. The experiments conducted with over 40 images of different nature and involving 20 human subjects assess the quantitative contribution of chromatic features in visual attention.

Keywords: Visual attention; Saliency map; Human perception; Eye movements; Color vision

* Corresponding author.
E-mail address: nabil.ouerhani@unine.ch (N. Ouerhani).

1077-3142/$ - see front matter
doi:10.1016/j.cviu.2004.10.009

1. Introduction

Visual attention is the ability of a vision system, be it biological or artificial, to rapidly detect potentially relevant parts of a visual scene, on which higher level vision tasks, such as object recognition, can focus.

It is generally agreed nowadays that under normal circumstances human eye movements are tightly coupled to visual attention. This can be partially explained by the anatomical structure of the human retina, which is composed of a high resolution central part, the fovea, and a low resolution peripheral one. Visual attention guides eye movements in order to place the fovea on the interesting parts of the scene. The foveated information can then be processed in more detail. Thanks to the availability of sophisticated eye tracking technologies, several recent works have confirmed this link between visual attention and eye movements [1–3]. Hoffman et al. suggested in [4] that saccades to a location in space are preceded by a shift of visual attention to that location. Using visual search tasks, Findlay and Gilchrist concluded that when the eyes are free to move, no additional covert attentional scanning occurs, and most search tasks will be served better with overt eye scanning [5]. Maioli et al. [6] agree that "There is no reason to postulate the occurrence of shifts of visuospatial attention, other than those associated with the execution of saccadic eye movements." Thus, eye movement recording is a suitable means for studying the temporal and spatial deployment of visual attention in most situations.

Like in human vision, visual attention can play a fundamental role in computer vision, given the high computational complexity of typical tasks [7]. Thus, the paradigm of computational visual attention has been widely investigated during the last two decades, and numerous computational models of visual attention have been suggested [8–13]. For a more complete overview on existing computational models of visual attention, the reader is referred to [14].

Most of these models rely on the feature integration theory presented in [15]. The saliency-based model, which relies on this principle, has first been presented in [16], and has given rise to numerous software and hardware implementations [17–19]. The model starts with extracting a number of features from the scene, such as color, intensity, and orientation. Each of the extracted features gives rise to a conspicuity map which highlights conspicuous parts of the image according to this specific feature. The conspicuity maps are then combined into a final map of attention named saliency map, which topographically encodes stimulus saliency at every location of the scene. Note that the model is purely data-driven and does not require any a priori knowledge of the scene. This model has been used in a number of computer vision applications, including image compression [20], color image segmentation [21], and object tracking in dynamic environments [22].

However, and despite the fact that it is inspired by psychophysical studies, only few works have addressed the biological plausibility of the saliency-based model [23]. Recently, Parkhurst et al. [24] presented for the first time a quantitative comparison between the computational model and human visual attention. Using eye movement recording techniques to measure human visual attention, the authors report a relatively high correlation between human attention and the saliency map, especially