Vision Research 97 (2014) 31–44

Contents lists available at ScienceDirect

Vision Research

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

Artificially created stimuli produced by a genetic algorithm using a saliency model as its fitness function show that Inattentional Blindness modulates performance in a pop-out visual search paradigm

Massimiliano Papera *, Richard P. Cooper, Anne Richards

Mace Experimental Research Laboratories in Neuroscience (MERLiN), Psychological Sciences, Birkbeck College, University of London, UK

ARTICLE INFO

Article history:
Received 9 April 2013
Received in revised form 12 January 2014
Available online 5 February 2014

Keywords:
Saliency
Genetic algorithm
Inattentional Blindness
Visual search
Pop-out stimuli

ABSTRACT

Salient stimuli are more readily detected than less salient stimuli, and individual differences in such detection may be relevant to why some people fail to notice an unexpected stimulus that appears in their visual field whereas others do notice it. This failure to notice unexpected stimuli is termed 'Inattentional Blindness' and is more likely to occur when we are engaged in a resource-consuming task. A genetic algorithm is described in which artificial stimuli are created using a saliency model as its fitness function. These generated stimuli, which vary in their saliency level, are used in two studies that implement a pop-out visual search task to evaluate the power of the model to discriminate the performance of people who were and were not Inattentionally Blind (IB).

In one study the number of orientational filters in the model was increased to check if discriminatory power and the saliency estimation for low-level images could be improved. Results show that the performance of the model does improve when additional filters are included, leading to the conclusion that low-level images may require a higher number of orientational filters for the model to better predict participants' performance. In both studies we found that given the same target patch image (i.e. same saliency value) IB individuals take longer to identify a target compared to non-IB individuals. This suggests that IB individuals require a higher level of saliency for low-level visual features in order to identify target patches.

1. Introduction

Inattentional Blindness (IB) occurs when someone fails to notice a stimulus when it unexpectedly appears in front of them. This phenomenon is more likely to occur when the person is engaged in a task that consumes resources (Dehaene & Changeux, 2005; Hannon & Richards, 2010; Mack & Rock, 1998; Most et al., 2001, 2005; Richards, Hannon, & Derakshan, 2010; Richards, Hannon, & Vitkovitch, 2010). Understanding the IB phenomenon may give insight into the functioning of the attentional system. For example, one hypothesis is that IB is due in part to a processing failure, that is, when working memory resources are fully involved in another task, there are insufficient resources remaining for processing of the unexpected stimulus. Another possibility is that the stimulus may be processed but because it is irrelevant to the primary task it is inhibited and therefore does not reach awareness (Morey & Cowan, 2004; Richards et al., submitted for publication). Inattentional Blindness may have important implications for safety procedures such as those related to flying aeroplanes (Green, 2003), air traffic control or for eye witness accounts of crimes occurring a few metres away (Chabris et al., 2011).

There are individual differences in the propensity to be IB as, given the same physical environment and conditions, some people will notice the unexpected stimulus whereas other will not. An unexpected stimulus is more likely to be detected if it is salient (Wickens et al., 2001), and therefore one possible contributing factor in individual propensity to IB in a visual task is how sensitive people are to detect saliency differences in visual scenes.

The attentional system could be viewed as a seeking-features mechanism where what we perceive depends on what the mechanism is focused upon (Driver, 2001). Therefore some details of the visual input may not be processed when the system does not attend them. However, there are some visual aspects that automatically modulate our attention towards salient stimuli (e.g., face stimuli have the power to attract attention over other stimuli; see Mack et al., 2002), although even salient stimuli may go unnoticed if they are not relevant/expected to the task at issue; this may lead to Inattentional Blindness (Mack & Rock, 1998).

* Corresponding author.
E-mail address: m.papera@bbk.ac.uk (M. Papera).

http://dx.doi.org/10.1016/j.visres.2014.01.013
0042-6989/

In a typical sustained IB task, participants are asked to track a series of white Ls and Ts as they move around the screen and to silently count how many times these letters (targets) hit the frame on the screen but to ignore a similarly moving series of black Ls and Ts (distractors). Several seconds after subjects have started this primary task, a red cross appears on the right hand side of the screen and moves across the centre to the left hand side. Participants who, when questioned at the end of the task, report seeing the red-cross are classified as non-IB, whereas those who fail to report having seen it are classified as IB. This is one possible dynamic task to address this phenomenon (see Most et al., 2001; Simons, 2003). One limiting factor in IB research is that subjects are categorized into one of just two groups (i.e., IB and non-IB groups, Inattentionally and Non-Inattentionally Blind, respectively) on the basis of a one-trial task. This is a general problem in the literature related to this psychophysical phenomenon (Hannon & Richards, 2010). However, several alternatives are present in the literature; see for example Kuhn and Findlay (2010) for the relationship between IB and misdirected attention, or Simons and Chabris (1999) for IB in dynamic events.

Unfamiliar objects/targets are more readily detected than familiar objects/targets if the unfamiliar item is displayed along with familiar ones (Levin et al., 2002; Treisman & Souther, 1985; Wolfe, 2001). One way to control for this effect is to create a set of stimuli (both target and distracters) that are completely unfamiliar to the subject. To do this, a saliency model based on that of Verma and McOwan (2009) was used to create stimuli whose representations do not induce detection as a result of the possible confounding effect of their familiarity. A genetic algorithm (GA) uses the saliency model as its fitness function to perform an artificial process of selection in order to achieve certain levels of saliency for stimuli. Although the model reported here is very similar to that presented by Verma and McOwan (2009), changes were made to the pipeline of events (e.g., changes to the drawing algorithm and the way stimuli are coded in chromosomes). Verma and McOwan (2009) showed that visual searching behaviour was modulated by the saliency of the scene, namely high saliency portions of an image were inspected by the subjects more readily than low saliency areas. This affected the time taken to identify a change in that changes made in high saliency regions were noticed much faster than those made in low saliency portions of the image. This was also confirmed when the saliency of the region was reversed (e.g. when a low saliency region that presents a change is manipulated to become a high saliency region, the time a subject takes to identify the same change is shortened; see Verma & McOwan, 2010).

We report a genetic algorithm that uses two versions of this saliency model as its fitness function to create a series of stimuli that are then used in two studies. Both studies test whether a low-level saliency model (e.g. bottom-up processing based) is able to discriminate two different trends in searching behaviour: given the same levels of saliency participants classified as IB subjects may be slower to detect a target in target-present images, compared to those participants showing quicker responses in terms of reaction times (i.e., non-IB subjects).

2. The saliency model

The model developed by Verma and McOwan (2009) is based on an earlier model (Itti, Koch, & Niebur, 1998) which computes a saliency map of an image from feature contrasts derived from spatial filters, colour filters, a luminance filter and orientational filters, as well as modelling top-down factors. However, the model presented in Verma and McOwan (2009) does not make use of either top-down factors or features such as flicker and motion (Itti & Baldi, 2008). Our goal is achieve a reasonably good saliency estimation that allows us to predict human behaviour and discriminate between IB and non-IB subjects, rather than mirroring a large number of attentional mechanisms. (Several implementations of saliency models can be found in Desimone & Duncan, 1995; Field, 1987; Itti, Koch, & Niebur, 1998; Koene & Zhaoping, 2007; Li & May, 2007; Milanese, 1993; Peters et al., 2005.)

For an image I, the model provides a global saliency value and a saliency map (Koch & Ullman, 1985). The saliency estimation is achieved through the computation of orientation and luminance Scales, which are then further combined to form Sub-Features Maps and Feature Maps (see Fig. 1). Since our saliency model is inspired by the one described in Verma and McOwan (2009), only differences and crucial details of our approach are discussed.

The model makes use of a hierarchical structure which was inspired by Marr's (1982) model of visual processing, and in particular, by the so-called *primal sketch* of a given visual scene that employs feature extraction of basic components including regions, edges, textures, etc. The outcome of the model can be defined as:

$$S_{\mathrm{I(Map,Val)}} = W_i(L_{(\mathrm{I})}, O_{(0,40,80,120,160,200,240,280,320)}) \quad (1)$$

S_{I} is the saliency model that returns two main outputs: Val, a global saliency value estimated on the basis of the saliency map (i.e., Map) obtained from a given image I.

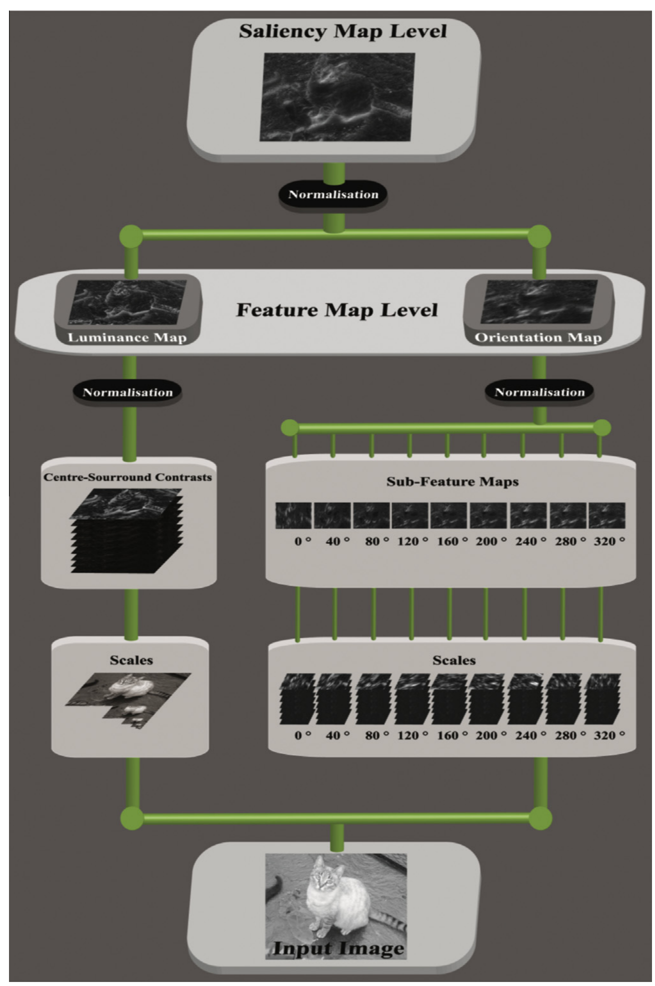

Fig. 1. The diagram depicts the decomposition of a given image for the features analysed by the saliency model. Orientation scales and luminance contrasts are extracted (see Scales/Contrasts level), combined together (i.e. sub-feature map and feature map levels), normalised and then further combined to form the global saliency map.