Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Vision Research 45 (2005) 643–659
Vision Research
www.elsevier.com/locate/visres

Visual correlates of fixation selection: effects of scale and time

Benjamin W. Tatler [a,*], Roland J. Baddeley [b], Iain D. Gilchrist [b]

[a] Sussex Centre for Neuroscience, School of Life Sciences, University of Sussex, Brighton BN1 9QG, United Kingdom
[b] Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, United Kingdom

Received 15 June 2003; received in revised form 9 November 2003

Abstract

What distinguishes the locations that we fixate from those that we do not? To answer this question we recorded eye movements while observers viewed natural scenes, and recorded image characteristics centred at the locations that observers fixated. To investigate potential differences in the visual characteristics of fixated versus non-fixated locations, these images were transformed to make intensity, contrast, colour, and edge content explicit. Signal detection and information theoretic techniques were then used to compare fixated regions to those that were not. The presence of contrast and edge information was more strongly discriminatory than luminance or chromaticity. Fixated locations tended to be more distinctive in the high spatial frequencies. Extremes of low frequency luminance information were avoided. With prolonged viewing, consistency in fixation locations between observers decreased. In contrast to [Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42* (1), 107–123] we found no change in the involvement of image features over time. We attribute this difference in our results to a systematic bias in their metric. We propose that saccade target selection involves an unchanging intermediate level representation of the scene but that the high-level interpretation of this representation changes over time.

Keywords: Saccadic selection; Image features; Spatial scale; Time course; Intermediate representation

1. Introduction

The way that our visual system samples world is both temporally and spatially constrained; sampling takes place during periods of fixation that typically occur at a frequency of 3–4 per second and is spatially constrained by sampling limits imposed by the retina. Given these constraints the visual system is unable to sample completely and uniformly the complex visual environment. Indeed, it is clear that during activities of daily life there are large proportions of the visual surroundings that we do not direct our eyes toward (e.g. Ballard et al., 1992; Land & Hayhoe, 2001; Land, Mennie, & Rusted, 1999). When viewing paintings and images, visual complexity is greatly reduced; the scene is constrained to two dimensions and spatially limited to a relatively small proportion of the observer's field of view. However, even under these conditions sampling is not complete or uniform, with some regions of the scenes receiving many more fixations than others (Buswell, 1935).

What are the processes that underlie this non-uniform sampling of the environment? Most researchers would argue that eye movement targeting involves a combination of bottom up and top down guidance factors. Some emphasise bottom up processes: implying that the most important factor in non-uniform sampling is the non-uniform distribution of "salience" in the world (e.g. Braun & Sagi, 1990; Kowler, Anderson, Dosher, & Blaser, 1995; Nakayama & Mackeben, 1989). The activity in low-level feature maps has been

* Corresponding author. Address: Department of Psychology, University of Dundee, Dundee DD1 4HN, United Kingdom. Tel.: +44 1382 348260; fax: +44 1382 229993.
E-mail address: b.w.tatler@dundee.ac.uk (B.W. Tatler).

0042-6989/$ - see front matter
doi:10.1016/j.visres.2004.09.017

proposed to underlie saccade targeting (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998; Niebur & Koch, 1996; Olshausen, Anderson, & Vanessen, 1993; Parkhurst, Law, & Niebur, 2002; Treisman, 1988; Wolfe & Gancarz, 1996). As evidence for the contribution of salience, the differences between the image statistics of fixated and non-fixated locations in scenes are emphasised; for example, Reinagel and Zador (1999) showed that fixated locations have higher contrast than non-fixated locations.

Other researchers emphasise the contribution of top down processes: implying that the non-uniform sampling is due mainly to high-level task demands. Pelz and Canosa (2001) suggested that "look ahead" fixations (checking objects that will be manipulated several seconds in the future) provide strong evidence that at least these types of eye movements are not salience driven, but rather are task dependent and driven by top down control. Shinoda, Hayhoe, and Shrivastava (2001) similarly stressed the importance of top down control, finding that detection of traffic signs in a driving simulator was modulated by visual scene context and task instructions.

While evidence that fixated and non-fixated locations differ in their statistics may be seen initially as evidence for the relative importance of low-level salience, this may not be the case. A predominantly top down selection mechanism may also result in non-random selection of low-level features. Most tasks require fixations on a specific set of objects and these objects tend to be distinguished by differences in luminance, colour, contrast and the occurrence of edges. Under this view, differences in image statistics at fixation could be an artefactual result of people fixating objects, which tend to differ from the background. Therefore, simply looking at the statistics at fixated and non-fixated locations cannot differentiate high- and low-level accounts.

One possible source of evidence is to investigate whether any quantifiable characteristics of eye movements change over viewing time. Both Buswell (1935) and Yarbus (1967) found that over time, the consistency between observers in where they fixated decreased. While this was primarily a qualitative observation, if confirmed quantitatively, it could place constraints on the interaction between top down and bottom up processes. Specifically, in the current study we measure not only the consistency of fixation locations, but also the inferred salience at these locations over time. This allows four possible frameworks to be distinguished. We call these four frameworks (1) *salience divergence*, (2) *salience rank*, (3) *random selection with distance weighting* and (4) *strategic divergence*.

The *salience divergence* model proposes that the balance between top down and bottom up control of saccade target selection changes over time. Specifically, the bottom up component is more influential early in viewing, but becomes less so as viewing progresses; this was suggested by Parkhurst et al. (2002). Such a framework could account for an observed decrease in between-participant consistency over time. In addition to a decrease in consistency, this framework predicts that the difference between saliency at fixated locations and at non-fixated locations will be greatest early in viewing.

A second possibility is that there is no change in either the top down or bottom up components of saccadic targeting over time. In the *salience rank* model, locations in the scene are ranked according to their visual salience and the oculomotor system selects targets sequentially according to this ranking; Itti and Koch's model uses a system for selecting successive targets for attention based upon decreasing salience (Itti & Koch, 2000). In any scene it is likely that there will be few locations of high salience, many of medium salience and even more of low salience, if salience is simply related to the output of filters (Field, 1987). Therefore the *salience rank* model predicts a decrease in consistency between participants, and a decrease in the salience of fixated locations over time.

The *random selection with distance weighting* framework for target selection (Melcher & Kowler, 2001) suggests that targets are selected using a proximity-weighted random walk process. This proposes that fixation locations are essentially random with respect to both bottom up and top down processes. The *random selection with distance weighting* proposal predicts that given a common starting location, the between-observer consistency of saccades will decrease over time, but that there should be no systematic change in the visual saliency at fixation.

A fourth possibility is *strategic divergence*. Here the influence of low-level visual feature salience on saccadic targeting does not change during viewing. Instead, the *strategic divergence* account proposes that the strategies chosen by observers have the same bottom up frame of reference for eye movements, but over time observers use different top down strategies. This could predict an increase in the variability of fixation locations, but no change in the saliency at fixation over time.

As can be seen, the four models predict both an increase in between-observer variability over time and different patterns of change in salience over time. We therefore quantified changes in the between-observer consistency in fixation locations as a function of viewing time. Explicitly, we estimated the probability distribution of fixation locations for individual observers. We then used an information theoretic measure (Kullback–Leiber divergence) to quantify the differences between these probability distributions. This quantity was estimated both as a function of fixation number and viewing time.

In order to quantify any difference in the visual saliency of fixated and non-fixated locations, we extracted