Vision Research 102 (2014) 41–51

Contents lists available at ScienceDirect

Vision Research

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

Deriving an appropriate baseline for describing fixation behaviour

Alasdair D.F. Clarke [a,b], Benjamin W. Tatler [c,*]

[a] Institute of Language, Cognition and Computation, School of Informatics, University of Edinburgh, UK
[b] School of Psychology, University of Aberdeen, UK
[c] School of Psychology, University of Dundee, UK

ARTICLE INFO

Article history:
Received 17 February 2014
Received in revised form 16 May 2014
Available online 29 July 2014

Keywords:
Eye tracking
Central tendency
Salience
Fixation location

ABSTRACT

Humans display image-independent viewing biases when inspecting complex scenes. One of the strongest such bias is the central tendency in scene viewing: observers favour making fixations towards the centre of an image, irrespective of its content. Characterising these biases accurately is important for three reasons: (1) they provide a necessary baseline for quantifying the association between visual features in scenes and fixation selection; (2) they provide a benchmark for evaluating models of fixation behaviour when viewing scenes; and (3) they can be included as a component of generative models of eye guidance. In the present study we compare four commonly used approaches to describing image-independent biases and report their ability to describe observed data and correctly classify fixations across 10 eye movement datasets. We propose an anisotropic Gaussian function that can serve as an effective and appropriate baseline for describing image-independent biases without the need to fit functions to individual datasets or subjects.

When we view complex scenes, where we look is influenced by a combination of low-level scene statistics (Itti & Koch, 2000), higher-level interpretation of the scene (Ehinger et al., 2009; Einhäuser, Spain and Perona, 2008), task goals (Buswell, 1935; Yarbus, 1967) and behavioural biases (Tatler & Vincent, 2009). If we are to understand the relative contributions of these different sources of guidance in scene viewing then techniques are required for quantifying the extent to which decisions about where to look can be attributed to each source.

At present, existing techniques can be categorised broadly into two approaches. First, the statistical properties at the centre of gaze can be quantified in order to measure how strongly a particular feature is associated with where gaze is directed (e.g., Pomplun, 2006; Reinagel & Zador, 1999). Second, locations that are likely to be fixated can be predicted based upon the distribution of statistical properties across an image and then the correspondence between the distribution of human fixation locations and the regions predicted as likely to be fixated from the statistical distribution can be assessed (e.g., Torralba, Oliva, Castelhano, & Henderson, 2006).

Both approaches can be used to assess the potential correspondence between a variety of low- or high-level features and fixation selection: provided that the feature under investigation can be quantified at each location in the scene, it is possible to quantify the strength of that feature at fixation or its distribution over the image. However, both approaches require a baseline measure in order to consider whether the association between the feature under test and fixation is greater than that expected by chance. Typically, a randomly generated set of locations is used to sample either the strength of the feature or the probability of selecting locations that fall within the regions predicted as likely to be fixated on the basis of the feature. The extent to which the control locations and the fixated locations correspond with the feature under test can then be used to assess whether any association between the feature and fixation is greater than would be expected by chance. A powerful and commonly used approach for making this judgment is to use the signal detection theoretic measure of the area under the receiver-operating-characteristics curve (see Green & Swets, 1966). The manner in which the random locations used as the baseline for such assessments are generated has important implications for the manner in which findings can be interpreted and indeed can significantly impact on the results (Henderson, Brockmole & Castelhano, 2007; Tatler, Baddeley & Gilchrist, 2005).

One approach is to use a uniform distribution for selecting control locations (e.g., Einhäuser, Spain & Perona, 2008; Parkhurst, Law & Niebur, 2002; Reinagel & Zador, 1999). Using such an approach means that any association between fixation and the feature under test that is beyond that found in the baseline comparison can be interpreted as suggesting that the feature is selected more

* Corresponding author. Fax: +44 (0) 1382 229993.
E-mail address: b.w.tatler@dundee.ac.uk (B.W. Tatler).

http://dx.doi.org/10.1016/j.visres.2014.06.016
0042-6989/

than would be expected if the eyes were directed randomly around a scene.

However, the existence of behavioural biases in how we view scenes (Tatler, 2007; Tatler & Vincent, 2009) suggests that a uniform random baseline may misrepresent selection with respect to features. That is, if the baseline comparison uses a uniform random distribution for generating control locations, any association found between fixation and features that extends beyond that in the baseline condition is likely to reflect a combination of selection based on image properties and image-independent biases in fixation behaviour. A more appropriate baseline for evaluating the association between an image feature and fixation placement is to select control locations from a distribution that reflects any image-independent biases in viewing behaviour. The most prominent and well-characterised image-independent bias in scene viewing is the central fixation bias: humans preferentially fixate the centre of the scene in a manner that is almost independent of the scene displayed to observers (Tatler, 2007; Tseng et al., 2009). As a result, control fixations can be drawn from distributions that reflect this central bias (see Tatler, 2007; Tatler, Baddeley & Gilchrist, 2005, for discussion of this issue).

There exist a number of ways that are typically used to construct a centrally-weighted distribution used in the baseline condition. One approach is to use a centred Gaussian to approximate the central bias and this may be fitted to the overall distribution of fixation locations in a dataset (Zhao & Koch, 2011), or scaled to the aspect ratio of the images presented (Judd, Durand & Torralba, 2012). Alternatively, these control distributions may be generated in ways that are aimed to maximise the chance of capturing any individual viewing biases that participants display when viewing scenes. There exist two main ways of attempting to capture individual viewing biases in baseline samples of features. First, the (x,y) locations of fixations on the test image can be used to sample features at the same locations in another (randomly selected) image (Parkhurst & Niebur, 2003). Second, (x, y) locations of fixations made by the same participant but when viewing different images can be used to sample features on the test image (e.g., Tatler, Baddeley & Gilchrist, 2005; Tatler & Vincent, 2009).

At present, it is unclear whether and how these different approaches to creating a baseline distribution vary in their suitability. The present study compares distributions of fixations across multiple existing datasets of eye movements in order to consider whether a single common distribution might be an appropriate baseline across studies and individuals or whether it is necessary to tailor the baseline distribution to each study and individual.

Being able to capture the statistics of the baseline condition appropriately is necessary for three reasons. First, if we wish to consider the relative importance of any feature in decisions about where to look, it is desirable to be able to quantify the unique variance associated with the particular feature after removal of variance associated with other factors that may contribute to decisions about where to look. In this way, any assessment of the importance of visual information (low- or high-level) to fixation selection should partial out variance that is associated with any image-independent biases in looking behaviour. Thus, if we compare the feature of interest to an appropriate baseline that accounts for image-independent biases, then we are better able to characterise associations between that feature and fixation behaviour. This principle extends beyond simply evaluating low-level salience models to any domain in which it is desirable to be able to characterise the contribution of a particular source of information to inspection behaviour. For example, in visual search paradigms, it is also useful to be able to remove any component of the behaviour that is driven by looking biases that are unrelated to the stimuli displayed.

Second, we can use this baseline as a benchmark for evaluating models of eye movement behaviour in scene viewing, as employed by Judd, Durand and Torralba (2012). Models should at least be able to outperform a baseline model based on image-independent biases such as looking at the centre of the screen. In their extensive comparison of a range of different salience models, Judd, Durand and Torralba (2012) found that only two models managed to outperform an image independent central bias baseline constructed using an aspect ratio-scaled Gaussian distribution. As there appears to be no empirical basis for this exact baseline, this may in fact underestimate the amount of variance that can be explained, and hence over-estimates the performance of the salience models.

Third, we can treat any image-independent bias as a factor in eye movement control itself. Thus, if we can computationally model these biases and derive appropriate characterisations of these biases we can use these as a component of models of fixation selection. That is, we can produce models with modules for low-level information, high-level information and image-independent biases. Given the strength of the central bias and its ability to predict human fixations, it is surprising that it is not more commonly incorporated into computational models. Indeed in their review, Judd, Durand and Torralba (2012) found only three studies that explicitly included a central bias in their model: Parkhurst and Niebur (2003) use the "shuffle" method; Zhao and Koch (2011) fitted Gaussians to their data, but restricted their baseline to an isotropic central bias, i.e., they fitted a covariance matrix with equal horizontal and vertical variance; and Judd et al. (2009) used an isotropic Gaussian fall-off that was stretched to match the aspect ratio of the image. Other examples in the literature include Clarke, Coco and Keller (2013) who used Euclidean distance from the centre of the image, and Spain and Perona (2011) who used a wide range of distance functions based on the Euclidean metric. Appropriate characterisation of image-independent biases therefore will allow appropriate and effective additions to existing models of fixation selection.

In the present study we evaluated different approaches to characterising baselines for understanding fixation behaviour when viewing scenes. Using ten eye movements datasets, we compared four ways of characterising image-independent biases in fixation selection: (1) fitting an isotropic Gaussian to the data (as in Zhao & Koch, 2011), (2) fitting a Gaussian scaled to the aspect ratio of the images (as in Judd, Durand & Torralba, 2012), (3) anisotropic Gaussians where the vertical and horizontal variances were fitted to each dataset, and (4) anisotropic Gaussians where the vertical and horizontal variances were fitted to each participant within each dataset. The final two approaches attempt to capture any experiment-specific (approach 3) or subject-specific (approach 4) differences in image-independent biases and as such conform to the recommendations made in previous discussions of this issue (Borji, Sihite & Itti, 2013a, 2013b; Tatler, Baddeley & Gilchrist, 2005). By comparing across these four approaches we were able to consider the relative ability of each approach for describing the data effectively and also the impact that each approach has upon our ability to classify fixated and control locations using each approach. One potential problem with the subject-level fitting (approach 4) is that this is likely to be sensitive to the sample size of eye movements used to construct the baseline distributions. This is a particular issue in studies with small numbers of trials or short presentations times (hence few fixations per image). As a result we considered how these approaches for describing the baseline are influenced by small n. In all of these approaches an empirical fit of the data is required to produce the baseline. We considered whether this is really necessary or whether a general purpose function can be employed that can be used irrespective of the subject or experiment under investigation. Here we used