A Saliency Model Predicts Fixations in Web Interfaces

Jeremiah D. Still
Missouri Western State University
Department of Psychology
jstill2@missouriwestern.edu

Christopher M. Masciocchi
Iowa State University
Department of Psychology
cmascioc@iastate.edu

ABSTRACT

User interfaces are visually rich and complex. Consequently, it is difficult for designers to predict which locations will be attended to first within a display. Designers currently depend on eye tracking data to determine fixated locations, which are naturally associated with the allocation of attention. A computational saliency model can make predictions about where individuals are likely to fixate. Thus, we propose that the saliency model may facilitate successful interface development during the iterative design process by providing information about an interface's stimulus-driven properties. To test its predictive power, the saliency model was used to render 50 web page screenshots; eye tracking data were gathered from participants on the same images. We found that the saliency model predicted fixated locations within web page interfaces. Thus, using computational models to determine regions high in visual saliency during web page development may be a cost effective alternative to eye tracking.

Author Keywords

Saliency, Interface Development, Design, Model

ACM Classification Keywords

H.1.2 User/Machine Systems; I.2.10 Vision and Scene Understanding

INTRODUCTION

Saliency, Search and Design

Some visual designs guide users to the locations of important information, while others mislead users. Visual saliency, inherent in a complex interface, cues users to certain spatial regions over others. If employed correctly by designers, salient cues may reduce information search times and facilitate task completion [cf. 18] by implicitly communicating to users where they ought to start their visual search [16]. In order to be considered salient, a feature must be visually unique relative to its surroundings. For example, text that is underlined amongst non-underlined text "pulls" the reader's attention to it. However, many interfaces, like web pages, are rich with visual media, such as text, pictures, logos and bullets, making the determination of salient features a complicated task. Given this complexity, designers are often left making best guesses about which spatial regions are salient within an interface. Previous research on visual search in web pages defines entry points as regions within a page where users typically begin their visual search. In this article, we will argue that these entry points are heavily influenced by visual saliency, that is, users will often begin searching web pages at the location of highest saliency. In related research examining cognitive processing these implicit and low level cues that guide a viewer's visual search are referred to as stimulus-driven properties – certain characteristics of the stimulus quickly "drive", or direct attention to certain locations over others. Currently, no consensus has been reached as to which visual characteristics, or stimulus-driven properties, make for effective entry points.

Measuring Overt Attention through Eye Tracking

Given the over abundance of visual information in our environments and our working memory limitations, attention must be selective, only allowing a limited amount of information into consciousness, for our cognitive system to function properly [8]. It has been suggested that the programming of eye movements has a direct and natural relationship with visual attention in that attention is often directed to whichever item is fixated [10]. Only information that falls directly on the fovea during a fixation is encoded with high resolution and only a limited amount of this high resolution information is processed, while the rest falls into rapid decay [see 4]. Thus, it is critical that users fixate on relevant visual information or that content will not reach users' awareness.

It is no surprise then, that designers often monitor eye movements to evaluate a web page's saliency, or entry points. Eye tracking systems allow designers to test whether their web pages actually guide users' fixations to important locations. However, eye tracking has a number of recognized costs. Eye tracking systems are often expensive,

not easily accessible, time consuming to employ and they gradually lose calibration [1, 2, 7, 15].

Stimulus and Goal Driven Searches

In this article we investigate the influence of stimulus-driven saliency on attention within the context of a web page. Stimulus-driven saliency guides attention quickly and without explicit intention, thus some might question its role during a purposeful search on a web page. There is ample evidence to suggest that goals do influence the guidance of attention. For example, web page eye tracking research has shown that changing the task (or goal) during a search, or seeking navigational or informational indicators, changes observers' fixation patterns [3]. Additional research has shown that, given enough time, expectations can cause a consistent pattern of fixations – F-shaped pattern or reading patterns (e.g., left-right/top-bottom) [14]. However, these goal-driven effects interact with stimulus-driven effects, making the stimulus-driven influences more difficult to examine [cf. 11]. Also, it is often the case that only a few seconds are spent on a web page (even with a goal in mind) making the understanding of stimulus-driven processing, which is believed to influence attention very rapidly, critical. For instance, when searching for information observers often only skim through approximately 18 words, and spend 4 to 9 seconds, per web page [2, 12]. One way to investigate the pure influence of stimulus-driven guidance is to use a computational saliency model designed to make predictions about what properties or features of a web page attention ought to select within complex media, or scenes.

Predicting Fixations through a Saliency Model

Visually salient items often draw observers' attention. To better understand the influences of saliency, or stimulus-driven selection, on attention, Koch and Ullman (1985) developed a model to compute an image's visual saliency without any semantic input (i.e., meaning of objects). Their model is based on the assumption that eye movement programming is driven by local image contrast leading to logical serial searches through complex spatial environments. These serial searches are guided by low level primitives extracted from a scene. The saliency model was developed under the pretense that low level visual features (i.e., color, light intensity, orientation) are processed pre-attentively in humans and, in turn, rapidly influence overt attention. Thus, the underlying assumption is that visual saliency is used to guide the fovea to unique areas within a scene that might provide the most efficient processing [5].

The computational model is implemented on a computer using digital pictures as stimuli to produce a pre-attentional or "saliency" map [9]. To create a saliency map, the model receives input from pixels within a digital picture. Then, it extracts three feature channels – color, intensity, orientation – at eight different spatial scales. These three channels are normalized and differences of center-surround are calculated for each separate channel. The separate channels are additively combined to form a single saliency map. An image's saliency map provides predictions of where spatial attention should be deployed [for detailed explanations refer to 6, 13]. In essence, the model makes predictions about which regions in an image have the most and least likely chance to be attended based purely on stimulus-driven properties. The saliency model is available for download from <SaliencyToolbox.net> as a collection of Matlab functions and scripts [17].

Testing a Saliency Model within Web Pages

Designers recognize the need to predict and identify where users' attention will be guided on a web page. For example, it is well known that one should avoid using poor designs that increase the likelihood of users missing important interface features such as branding, navigational or informational symbols. But, using an eye tracking system to monitor guidance of attention – as is traditional – can be expensive, difficult to employ and time consuming within the context of a practical iterative design process. Thus, we investigated the utility of a computational saliency model in predicting the guidance of attention in web page screenshots. This new method is benchmarked and compared to another set of data in which participants' eye movements were tracked while they viewed the same web page screenshots.

METHOD

Participants

The data from eight undergraduate participants are examined. All participants reported extensive web site experience.

Stimuli and Equipment

The images were 50 screenshots of various web pages. Each participant saw each screenshot only once.

Participants' eye movements were recorded by an ASL eye tracker with a sampling rate of 120 Hz. Screenshots were shown on a Samsung LCD monitor, which had a viewing area of approximately 38.0 cm × 30.0 cm. A chin rest maintained a viewing distance of approximately 80 cm. Images subtended approximately 26.7^0 x 21.2^0 visual angle.

Procedure

Participants first read and signed an informed consent document, and were then seated in front of the monitor with their chin in the chin rest. The experiment began and concluded with a 9-point calibration sequence to calibrate the eye tracker and estimate the amount of tracking error.

Participants were told that they would view a series of web page screenshots, and that they should, "look around the image like you normally would if you were surfing the internet." A fixation cross was presented at the center of the screen to signal the beginning of a trial. After a delay of approximately 1 second, a randomly selected web page screenshot was presented for 5 seconds. The fixation cross then reappeared to signal the beginning of the next trial.

The experiment took approximately 15 minutes to complete.

Figure 1. Two examples of web page screenshots and their corresponding saliency maps.

Creation of saliency maps

Saliency maps were created using the algorithms developed by Itti, Koch, and Niebur (1998). The model was run on each image individually and the output was normalized by dividing all values by the maximum value for that map, and multiplying all values by 100. To simplify data analysis, the size of the saliency maps was increased to be identical to the size of the screenshots (1024 x 768 pixels). As described in the Introduction, these saliency maps are 2-D representations of areas in the screenshot that show the relative saliency of locations in the image. Figure 1 shows an example of two web page screenshots and their corresponding saliency maps. Low values (dark areas in the image) indicate regions of the image that are low in saliency, while high values (light areas in the image) indicate regions high in saliency.

RESULTS

We used a similar technique to Parkhurst, Law, and Niebur (2002) to determine whether salient regions in web pages were fixated more often than would be expected by chance. Specifically, the values of the saliency map at the location of each participant's first ten fixations were extracted. For example, the x, y coordinates of the first fixation for each participant was determined for every screenshot and the value at the same location in the corresponding saliency map was extracted. This process was repeated for fixations two through ten. These values formed the *Observed Distribution* of participant responses (Figure 2).

To determine the likelihood that salient regions would be fixated by chance, we repeated the process used to find the Observed Distribution after rearranging the fixations and saliency maps for all screenshots. For example, the values from the saliency map for screenshots 2 to 50 were

extracted at the fixated locations from screenshot 1. The saliency values of all other screenshots were extracted at the location of the first ten fixations for all subjects for each screenshot. These values formed the *Shuffled Distribution*. The method used to create this distribution controls for spatial biases that may inflate correlations between fixations and salient regions. If the values of the Shuffled Distribution are larger than those of the Observed Distribution, it would indicate that participants fixated on regions that are lower in saliency than what is expected by chance. If, however, the values of the Observed Distribution are larger than those in the Shuffled Distribution, it would indicate that participants fixated regions that are higher in saliency than what is expected by chance.

Figure 2 shows the means for the Observed and Shuffled Distributions of the first ten fixations for each screenshot. An analysis of variance was conducted with fixation number (1-10) as a within-subjects variable and distribution (observed, shuffled) as a between-subjects variable, to determine whether any differences between the distributions varied as a function of fixation number. The main effect of fixation number was reliable, $F(9, 882) = 6.39$, $MSE = 19.03$, $p < .001$. Pairwise comparisons revealed that the values for the first fixation were higher than all other values, and that the values of the tenth fixation were lower than all other values. This indicates that early fixations tend to occur at regions of higher salience than those of later fixations. More importantly, the main effect of distribution was also reliable, $F(1, 98) = 4.86$, $MSE = 397.95$, $p < .05$, indicating that the values of Observed Distribution were larger than those of the Shuffled Distribution. This difference confirms that participants fixated regions higher in saliency than would be expected by chance, showing that the saliency model is effective at predicting fixations. Distribution x Fixation number was not significant, $F < 1$.
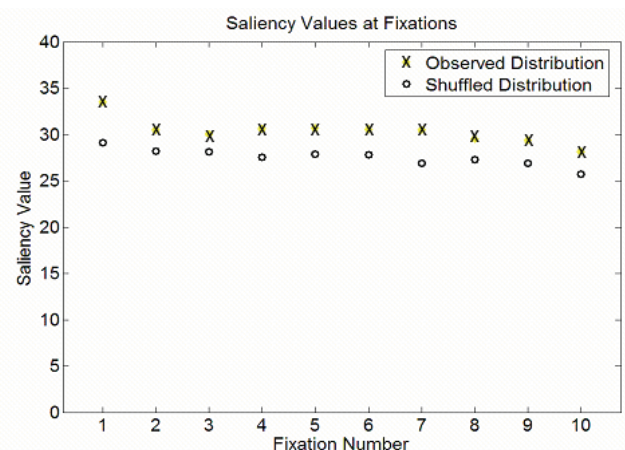

Figure 2. Mean saliency values for the observed ('X') and shuffled ('o') distributions for the first ten fixations.

DISCUSSION

Eye tracking is a commonly employed method for examining the guidance of overt attention within interfaces (e.g., web pages). However, it has several drawbacks. We propose that a web page's saliency, stimulus-driven properties, may be revealed through the use of a computational saliency model. Therefore, we compared the performance of the model to eye tracking data collected from human observers. We were able to demonstrate that, indeed, the saliency model predicts the deployment of overt attention within a web page interface.

Previous research has shown a modest correlation between saliency and eye fixations in natural and artificial scenes [13]. We have extended this research by showing that even in web pages, which may contain more semantic information (e.g., meaningful: text or images) than nature scenes, fixations are correlated with saliency. Specifically, participants were more likely to fixate on regions in the web pages with a higher saliency value than predicted by chance.

Our data suggest that saliency maps alone can provide reasonable predictions of overt attention. In addition, saliency maps can be generated quickly, and require no additional equipment or participants. Even with these positive attributes, one may be hesitant to abandon eye tracking altogether. Our recommendation to designers is to choose the method most appropriate for your project given your constraints and needs. It is often the case that developing effective interfaces requires many levels of analysis. For example, during the early formative testing process it would be appropriate to begin by using the saliency model to ensure that regions identified as being important are also visually salient. Then, during the 'final' prototype development stage, employ the eye tracking method to verify that your participants are actually looking at the critical elements in the design.

REFERENCES

1. Arroyo, E., Selker, T. & Wei, W. Usability tool for analysis of web designs using mouse tracks. *Computer-Human Interaction extended abstracts on human factors in computing systems* (2006), 484-489.
2. Chen, M., Anderson, J. R. & Sohn, M. What can a mouse cursor tell us more?: Correlation of eye/mouse movements on web browsing. *Computer-Human Interactions extended abstracts on human factors in computing systems* (2001), 281-282.
3. Cutrell, E. & Guan, Z. What are you looking for? An eye-tracking study of information usage in web search. *Proceedings of the SIGCHI conference on human factors in computing systems* (2007), 407-416.
4. Egeth, H. E. & Yantis, S. Visual attention: Control representation and time course. *Annual Review of Psychology* (1997), *48*, 269-297.
5. Itti, L. & Koch, C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12) (2000), 1489-1506.
6. Itti, L., Koch, C. & Niebur, E. A model of saliency-based fast visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11): 1254-1259, November 1998.
7. Johansen, S. A. & Hansen, J. P. Do we need eye trackers to tell where people look? *Proceedings of Computer-Human Interaction extended abstracts on human factors in computing systems* (2006), 923-928.
8. Johnston, W. A. & Dark, V. J. Selective attention. *Annual Review of Psychology* (1986), *37*, 43-75.
9. Koch, C. & Ullman, S. Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology* (1985), *4*, 219-227.
10. Kowler, E., Anderson, E., Dosher, B. & Blaser, E. The role of attention in the programming of saccades. *Vision Research* (1995), *35*, 1897-1916.
11. McCarthy, J. D., Sasse, M. A. & Riegelsberger, J. (2003). Can I have the menu please? An eyetracking study of design conventions. *Proceedings of Human-Computer Interaction*, 401-414.
12. Nielsen, J. (2008, May). *How little do users read?* Retrieved May 12, 2009 from http://www.useit.com/alertbox/percent-text-read.html.
13. Parkhurst, D., Law, K. & Niebur, E. Modeling the role of salience in the allocation of overt visual attention. *Vision Research* (2002), *42*, 107-123.
14. Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* (1998), *124*(3), 372-422.
15. Tarasewich, P., Pomplun, M., Fillion, S. & Broberg, D. The enhanced restricted focus viewer. *International Journal of Human-Computer Interaction* (2005), *19*(1), 35-54.
16. Treisman, A. M. Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance* (1982), *8*, 194-214.
17. Walther, D. & Koch, C. Modeling attention to salient proto-objects. *Neural Networks* (2006), *19*, 1395-1407.
18. Wolfe, J. M. Guided Search 4.0: Current Progress with a model of visual search. In W. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 99-119). New York: Oxford, 2007.