Research Article

Design and Validation of an Attention Model of Web Page Users

Ananya Jana and Samit Bhattacharya

Department of Computer Science & Engineering, IIT Guwahati, Assam 781039, India

Correspondence should be addressed to Samit Bhattacharya; samit3k@gmail.com

In this paper, we propose a model to predict the locations of the most attended pictorial information on a web page and the attention sequence of the information. We propose to divide the content of a web page into conceptually coherent units or objects, based on a survey of more than 100 web pages. The proposed model takes into account three characteristics of an image object: chromatic contrast, size, and position and computes a numerical value, the attention factor. We can predict from the attention factor values the image objects most likely to draw attention and the sequence in which attention will be drawn. We have carried out empirical studies to both develop and determine the efficacy of the proposed model. The study results revealed a prediction accuracy of about 80% for a set of artificially designed web pages and about 60% for a set of real web pages sampled from the Internet. The performance was found to be better (in terms of prediction accuracy) than the visual saliency model, a popular model to predict human attention on an image.

1. Introduction

In this age of information, web pages play an important role. They provide an interface to the vast repository of information known as the World Wide Web (WWW). The web page design and display technologies (e.g., web 2.0) make it possible to provide the information in a "usable" form (by allowing us to change font, add colors, and insert animation, providing quick links, and so on). However, the focus of web page design so far has been to provide technology to make the entire web page usable. Implicit in this approach is the assumption that the user actually tries to see (or access) the content of the entire page. The assumption need not always hold though. Based on their behavior, we can broadly divide them into two types. Some users explore the entire contents of a web page (e.g., a WikiPedia user interested to know about some topics). We may call them "goal-oriented" to differentiate them from the other group who explore the web page content only partially at first. We may refer to the second group of users as "exploratory."

As an example, consider a personal home page put up by someone interested in changing jobs. The page contains educational qualification, experience, achievements, and contact address and may be a brief write-up about the person with a photograph. The designer would certainly like the viewer to see all the contents in detail. However, a potential employer, who may have to go through hundreds of such web pages, may not be interested to explore everything in detail at first. Instead, he may glance through the page for a short duration and try to acquire the "necessary information" within this short span of time. If the information is found to be worthwhile, he may explore the content further. Otherwise, he may simply move to some other page. The potential employer belongs to the exploratory group. For a designer, it is a challenge to design web pages for such users, because there is no way to predict if the information the designer wants to highlight will actually be "seen" by the user or not. As an illustration, consider the above example again. What is the *worthwhile information* here? As the page was put up by someone interested to change jobs, the designer should assign more importance to the information on "experience" than the information on "education," which in turn is more important than "contact" and so on. However, within the early exploration phase, will the user be able to see the "experience" content or move away to other pages without seeing it?

Thus, we need some ways to assign importance to certain contents with respect to the others on a web page. A designer can take recourse to options such as highlighting, font size

(a)   (b)

FIGURE 1: Examples of web page design problems. It is not clear if the user will get to see the worthwhile information at all in either of the cases.

increase, and color and contrast, in accordance with the available guidelines. However, guidelines are not followed in most of the cases, which include "official" web pages as well as those developed by amateurs (such as blogs and home pages). The problem arises because the responsibility of conforming to the guidelines lies solely with the designer. As is the case, most designers either ignore those or are not aware of their existence. Therefore, it becomes difficult to know whether the most important content is actually visible to the user. Two examples of such bad designs are shown in Figure 1.

Thus, if a web page is not properly designed, the usability aspects of the design may not be of much use. So, in addition to the standard aspects of usability, we propose to add one more dimension to it: the "visibility" of the information contained in a web page. We may reasonably assume that some information on the page is more important than the others (from the designer's point of view). Naturally, a designer likes to see that the important information gets the user's attention before other information. It is also important that the important information draws the attention early, before the user loses interest and moves away. We define "visibility" as the degree to which a piece of information on a web page is able to draw user attention before the other information. The challenge is to *predict* the visibility of information on a web page.

One solution is to use *eye tracking systems* [1–3]. However, most web designers cannot afford such systems due to high cost. An alternative approach is to semiautomate the web page design process. At present, web pages are designed using either of the several available tools such as Microsoft FrontPage, Microsoft Publisher, and DreamWeaver, none of which can check "if the most important information has higher visibility than others." A semiautomatic design process can be envisaged as consisting of a tool to check automatically the source program of a web page (output of the existing tools) and predict the sequence in which the web page content is likely to be viewed by a user. Based on the predicted attention sequence, the different regions of the web page can

be ranked as per their attention drawing abilities. From this ranking, a designer can decide if the information that he feels most important are able to draw the user's attention early and can take corrective measures accordingly. In order to develop such an approach, we require two things.

First, we should be able to divide the web page content into *units* or *chunks* that represent *conceptually coherent* information from the designer's perspective. For example, the information about research activities in one's home page may be considered a unit that is different from another unit of information detailing education. Therefore, we introduce the notion of *objects* to divide the pictorial and textual information contained in a web page. The second thing we need is a predictive model of human attention, which works on the objects. With the predictive model, we can estimate the sequence of user attention on the objects. On the basis of this information, we rank the objects as per their attention drawing capabilities. The ranks will allow us to infer about the visibility of objects. In this work, we propose one such model for the image objects (i.e., objects representing pictorial information).

The proposed approach is developed from empirical data. We conducted *three* major user studies for empirical data collection. The first study was a detailed survey of about 100 web pages sampled from the Internet to determine the types of objects used in web page design as well as their relative significance. In the second study, eye gaze data were collected from participants to develop the proposed attention model. We again collected eye gaze data of web page users in the third study, to estimate the prediction accuracy of the proposed model.

The paper is organized as follows. Section 2 presents works related to human attention modeling. The survey and data analysis for identification of object types and their significance are presented in Section 3. The proposed model is described in Section 4. Parameters of the model were estimated from an empirical study, which is described in Section 5. The empirical studies carried out for model

validation are presented in Section 6. The strengths and weaknesses of the proposed model are discussed in Section 7. Section 8 concludes the paper.

2. Related Works

Human attention modeling is a well-researched and active area, of which computational attention modeling is an important part [4–6]. We found two such models relevant to our work, namely, (a) the guided search model and (b) the visual saliency model.

2.1. Guided Search Model. The guided search model (GSM) [7, 8] represents the human search behavior in a visual field. The search behavior is characterized by both *parallel* and *limited capacity* visual processes. Our visual system initially processes the whole input in parallel. Information collected from the parallel processes is then used by limited capacity processes to perform other operations that are limited to one location at the same time.

During the parallel processing of input, two *feature maps* are created for *color* and *orientation* features. Each of these features is quantized with four values. The values for color are red, yellow, green, and blue (represented by some numerical values, e.g., red = 10, etc.). For orientation, the values are given in degrees with respect to the vertical direction (tilted right > $0°$, left < $0°$). The four orientations are: (a) steep ($-45°$ < angle < $45°$), (b) shallow ($-90°$ < angle < $-45°$ or $45°$ < angle < $90°$), (c) right ($0°$ < angle < $90°$), and (d) left ($-90°$ < angle < $0°$). From the feature maps, an *activation map* for the input is created. The activation map takes care of the fact that there are two components of attention: stimulus driven, bottom-up activation, and user driven, top-down activation. Bottom-up activation is a measure of how different an item is from its neighboring items. It does not depend on the user's knowledge of the specific search task. It guides the attention towards the distinctive items in the visual field but not towards a desired item by the user. To incorporate user's desire, top-down activation is needed. Therefore the activation of a particular item is a combination (a weighted sum) of both the top-down and the bottom-up activation [9]. The likelihood of an item to be attended by the user depends on the activation value. The higher the activation value is, the more will be the chance to get attention from the user. Search time to locate an item also depends on the activation value. The higher the activation value is, the lesser time it will take to locate that item.

Although the model can compute the likelihood of an item in a visual field to capture user attention (using the bottom-up activation value of the item), it has one important limitation from implementation point of view. The model has used four different values for colors (red, green, blue, and yellow) and orientations (steep, shallow, right, and left). However, these values are experiment specific; that is, values differ with the different application of the model. For instance, if we use such model for designing web pages, the values may be different [9]. The onus is on the implementer to try, exhaustively, different sets of values for these properties.

Thus, the model lacks portability. In the next subsection, we discuss another relevant model of human visual attention called the visual saliency model, which is essentially an extension of the GSM.

2.2. Visual Saliency Model. Visual salience is a distinct quality which makes some items stand-out from their neighbors [10]. Our attention is attracted to visually salient items. The visual saliency model (VSM) finds the most salient location within a visual scene [11–14]. In VSM, feature maps are created using *center-surround mechanism*, which involves comparison of the visual properties of each location with its surrounding. Three feature maps for color, intensity, and orientation are constructed using the mechanism. These maps are then combined to obtain a unique *saliency map* (i.e., a representation of saliency at every location in the visual field) from which the most salient locations are determined. A normalization process is used for feature combination, where the feature values are mapped to a fixed range in order to eliminate modality dependent differences. The normalization process has one more function. If any feature is uniform throughout the image, the weight for that feature is reduced; otherwise, the weight for that feature is increased.

By definition, at any given time, the maximum of the saliency map represents the most salient image location to which *focus of attention* should be directed. This maximum is detected by a winner-takes-all network inspired from biological architectures. Winner-takes-all is a phenomenon used to show that the most salient location gets the user's attention at a particular instant. To create dynamic shifts of the focus of attention, rather than permanently attending to the most salient location, the VSM uses two mechanisms:

(1) The focus of attention is shifted so that its center is at the attended location.

(2) The intensity of attended location is decreased to direct the attention to other locations too.

Clearly, VSM offers a way to compute human attention behavior. It has also been tried to predict fixations on web pages [15]. However, it is constrained by the resolution of the input image [14]. It has been found that it can efficiently find the salient locations in an image of resolution 800 × 400 (http://www.saliencytoolbox.net/). For images of higher resolution, the existing implementation asks for resizing the image to 800 × 400. Though the implementation can return the salient regions on the original image, the results are different than the resized image. The main problem of the VSM in the present context, however, is to make it work with web page objects. As we mentioned before, we propose to view a web page as a set of objects (conceptually coherent units of information). Although a web page can be considered as an image and we can apply VSM to identify salient locations on it, the approach fails to rank objects as we found out in our work. This is because the VSM distinguishes objects only in terms of their image properties without taking into account other factors such as size and position.
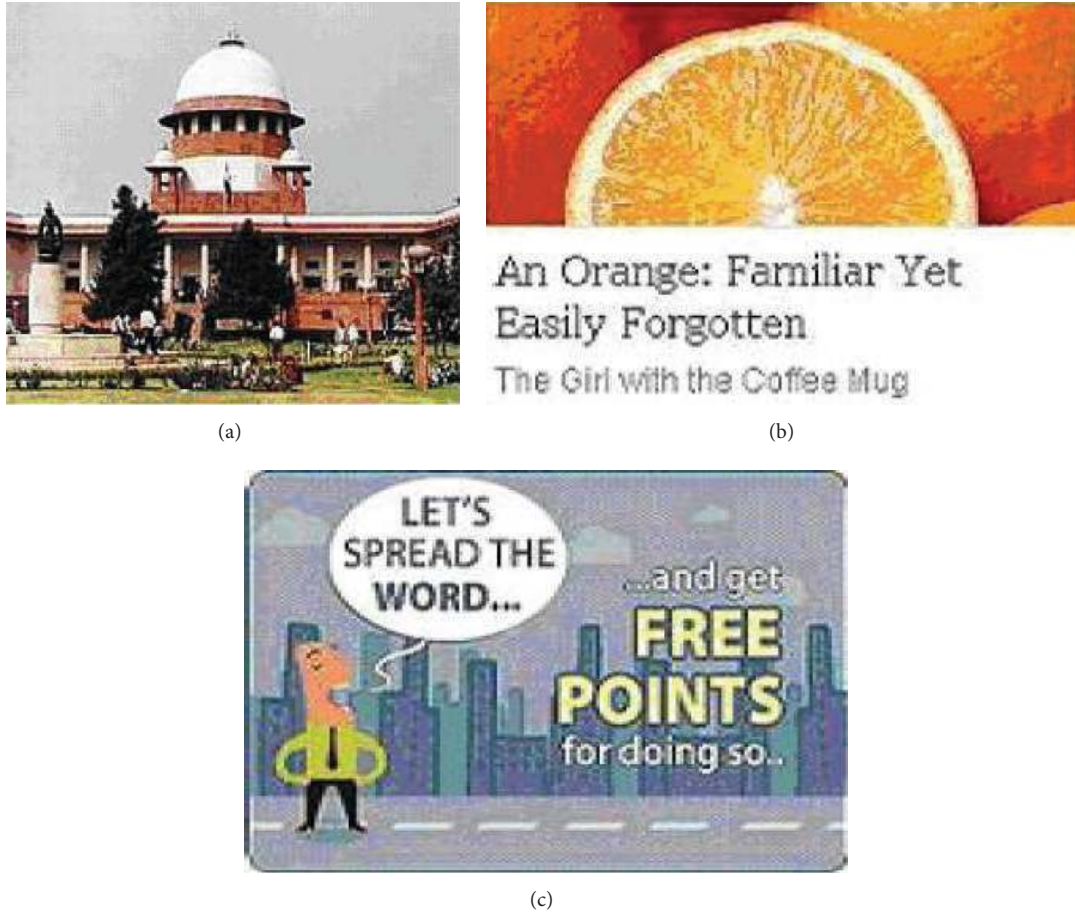
(a)

(b)

(c)

Figure 2: Examples of image objects. In (a), a pure image is shown. An image object with a caption is shown in (b) while in (c) an image with embedded text is shown.

3. Web Page Objects: Types, Definition, and Importance

In order to understand our proposed "object-view" of web pages, we have studied about 100 web pages sampled from the Internet. The survey helped us to identify the types of objects used in web page design. Only those web pages that did not have any animation were considered in the study. We adopted the following approach to identify the samples.

We approached a group of 25 volunteers (undergraduate and graduate students) who were regular Internet users. They belonged to the age group of 20–25 years. Among them, 15 were male and the rest female. Each of them was asked to submit a list of 10 web pages without animation, which they visited often. In this way, we collected a list of 250 web pages. The methodology allowed us to collect information about popular web pages, which in turn led us to identify objects that are likely to be present in most of the web pages. We then checked for any duplication in the entries and removed those. We also checked each of the remaining entries ourselves to see if any web page contained animations and removed those

from the list. After these steps, we were left with about 100 web pages (103, to be precise). Based on the thorough investigation of each of the web pages in the final list, we identified the following seven types of objects that are used in web page design. *(1) Images.* A single image or an image with embedded text or image with caption (i.e., textual description of the image placed very near to it) is an object. Examples of such objects are shown in Figure 2. To categorize an object having an image with a caption, we assume that the caption is close to the image (maximum distance to be 1 cm on a 1280 × 1024 screen) and the maximum number of words is 15 (decided based on our study of popular web pages provided by the volunteers).

(2) Text. A piece of text with/without heading, even if it contains variation in fonts, styles (bold, italics, etc.), or color, as shown in Figure 3.

(3) List. A list under its heading together forms an object. One such example is shown as follows.

Table 1: An example of a table object.

Name	Holds	Size (Megs)
UA_Root	Root files	5 M
UA_Applications	Applications directory	65 M
UA_4BSD	4BSD directory	371 M
UA_PDP11	PDP-11 directory	479 M
UA_PDP11_Trees	PDP-11/trees directory	97 M
UA_VAX	VAX directory	5 M
UA_Other	Other directory	6 M

Figure 3: A text object with heading. Sometimes such objects may occur without a heading also.

A List Object. FTP access to the Unix Archive is available from

(i) ftp.math.utah.edu in the US,

(ii) ftp.medienfuzzis.com in Germany,

(iii) ftp.autistici.org in Italy,

(iv) ftp.lug.udel.edu in the US,

(v) ftp.tux.org in the US,

(vi) ftp.gcu-squad.org in France,

(vii) sunsite.icm.edu.pl in Poland,

(viii) ftp.uvsq.fr in France,

(ix) pdp11.org.ru in Russia,

(x) ftp.darwinsys.com in Canada,

(xi) ftp.ics.es.osaka-u.ac.jp in Japan,

(xii) ftp.cs.tu-berlin.de in Germany,

(xiii) mirror.interoute.net in England,

(xiv) minnie.tuhs.org in Australia,

(xv) ftp.win.tue.nl in Netherlands,

(xvi) ftp.tuhs.org.ua in Ukraine.

(4) Tables. Tables constitute separate objects from texts, based on their attention drawing capabilities. An example is shown in Table 1. They may/may not be accompanied by a heading.

Figure 4: Example of dangling headings.

Figure 5: Examples of menu objects.

Figure 6: Examples of interactive objects.

(5) Dangling Headers. Sometimes, a heading appears without any accompanying text, list, or tables, as shown in Figure 4.

(6) Menu. We propose to treat menus, which may appear on a side or the top of a web page, as objects separate from text or list objects. Menu object examples are shown in Figure 5.

(7) Interactive Items. This is any kind of interactive item like search box, text box, button, radio button, check box, and so forth, through which users can provide input. Some examples are shown in Figure 6.

We computed the area occupied by these objects in the sample web pages, as shown in Figure 7. It may be noted in Figure 7 that the image objects occupy most of the area (more than 40%). This is followed by the text objects (about 16%). Each of the other types of objects occupies less than 9% area of the sample web pages. In this work, we attempted to propose a model of user's attention for the image objects.

4. Proposed Model

The key concept in our proposed model is the *attention factor* (af) of an object, which is the measure of the degree of attention a user is likely to pay to the object. We have used four features of an object, namely, the *intensity contrast*, the *chromatic contrast*, the *size*, and the *position*, to determine af.

4.1. Intensity Contrast. Intensity contrast defines the difference in luminance of the object and the background with
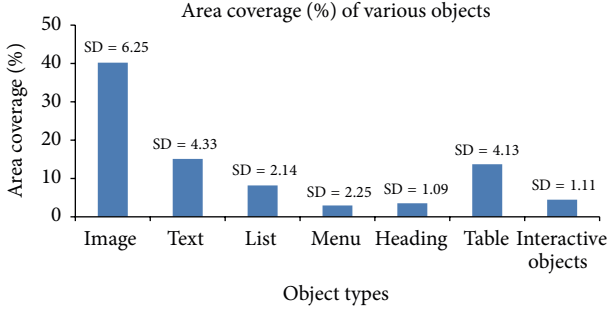
FIGURE 7: The histogram showing the average area occupied by the different objects in the sample web pages.
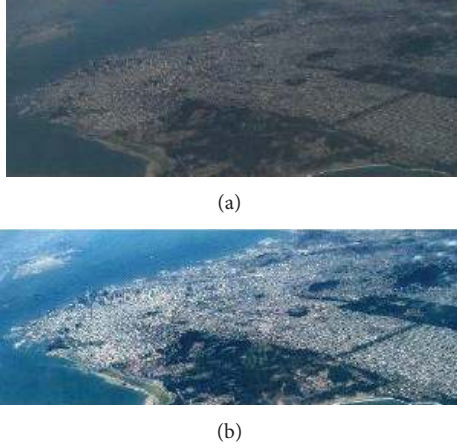

(a)

(b)

FIGURE 8: Intensity contrast example. (a) shows a low contrast image whereas (b) shows the same image with high intensity contrast.

respect to the other objects present in a scene. Intensity contrast plays an important role in affecting the human visual attention as it makes an object distinguishable from the background and other objects. Figure 8 shows the contrast difference between the two images.

Two definitions are commonly used to calculate the contrast [16], namely, the Michelson contrast and the Weber contrast. In this work, we are discriminating the objects with respect to the other objects in the web page and not with the background. Hence, we have used the Michelson contrast in our work. The expression to calculate intensity contrast of an object is shown in (1), where obj_i^{ic} is the intensity contrast of the ith object, int_{wp} is the intensity of the whole web page, and int_{obj} is the intensity of the object:

$$\text{obj}_i^{\text{ic}} = \frac{\text{abs}\left(\text{int}_{\text{wp}} - \text{int}_{\text{obj}}\right)}{\text{int}_{\text{wp}}}. \tag{1}$$

The individual intensity values are taken from the HSI model [17] and are calculated as follows:

$$\text{int} = \frac{R + G + B}{3 \times 255}. \tag{2}$$

4.2. Chromatic Contrast. There is a biological phenomenon known as the Color Double Opponency [18]. In the human retina, there are three types of cone photoreceptors that respond to different wavelengths (long (L), medium (M), and short (S), also referred to as red (R), blue (B), and green (G) [19]). In the center of the receptive fields, neurons are excited by one color and inhibited by another. Such chromatic opponency exists for the red/green, green/red, blue/yellow, and yellow/blue color pairs in the human visual cortex. Therefore we need to consider chromatic contrast separately from intensity contrast.

We have used the formulation reported in Itti et al. [13] to compute the chromatic contrast. The computation consists of two steps. In the first step, we calculate the chromatic contrast for both the object and the web page.

Step 1. Consider

$$r_{\text{abs}} = rr - \left(\frac{(gg + bb)}{2}\right), \tag{3a}$$

$$g_{\text{abs}} = gg - \left(\frac{(rr + bb)}{2}\right), \tag{3b}$$

$$b_{\text{abs}} = bb - \left(\frac{(rr + gg)}{2}\right), \tag{3c}$$

$$y_{\text{abs}} = \left(\frac{(rr + gg)}{2}\right) - \left(\frac{\text{abs}(rr - gg)}{2}\right) - bb \tag{3d}$$

$$\text{op}_{\text{img}} = \frac{(r_{\text{abs}} + g_{\text{abs}} + b_{\text{abs}} + y_{\text{abs}})}{4}. \tag{3e}$$

As a second step, we find the difference ratio as shown in (4).

Step 2. Consider

$$\text{obj}_i^{\text{cc}} = \frac{\text{abs}\left(\text{op}_{\text{img}} - \text{op}_{\text{obj}}\right)}{\text{op}_{\text{img}}}, \tag{4}$$

where rr, bb, and gg are the red, blue, and green components of the web page, respectively, r_{abs}, g_{abs}, b_{abs}, and y_{abs} are the RGBY (red, green, blue, and yellow) values after the double opponency is applied to the web page and op_{img} is the mean of all the RGBY values obtained from the web page. Similarly, we calculate op_{obj}, which is the mean of all the RGBY values obtained from the object obj on the web page. From these values, the chromatic contrast obj_i^{cc} for the ith object is obtained as a ratio of the absolute difference between the op_{img} and op_{obj} to the value of op_{img}.

4.3. Size. We are familiar with the fact that bigger objects in a visual scene take less time to be located. We have taken this
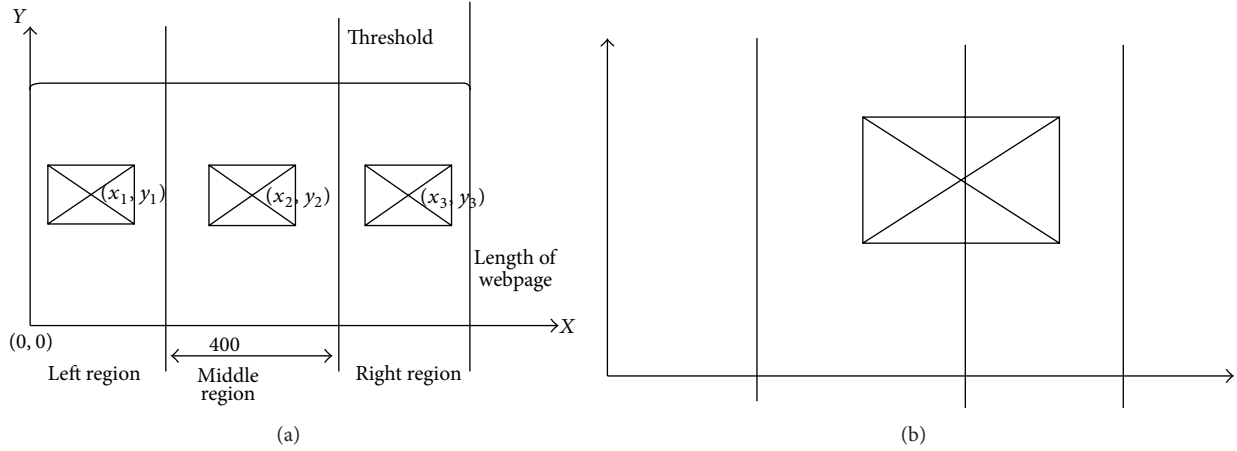
FIGURE 9: The partitioning of the screen into the three regions. In (a), three objects are shown as examples of object positions. Note that in this work, we assume an object lies entirely in a region. Thus situations such as the one shown in (b), where objects overlap two or more regions, are not considered.

fact into account by considering the ratio of the area occupied by the object to the total area of the web page, as one of the factors in our proposed model, as shown in

$$\text{obj}_i^{\text{size}} = \frac{s_{\text{obj}}}{s_{\text{wp}}}, \tag{5}$$

where s_{obj} and s_{wp} are the area of object under consideration and the web page, respectively, and $\text{obj}_i^{\text{size}}$ is the size of the object obj with respect to the web page wp.

4.4. Position. It has been found that the human attention is initially drawn to the middle of an image [20]. Therefore, position of an object plays an important role in drawing attention. The inclusion of position in the model is achieved through the use of two *position factors*, namely, the X position factor x_p and the Y position factor y_p.

We divide the screen into three regions, as shown in Figure 9. Assuming a resolution of at least 800×400, the 400 pixels in the middle of the screen form the *middle* region. Objects placed in this region are supposed to draw the most attention. The portions of the screen on the left and right of the middle region are the *left* and *right* regions, respectively. The position of an object is represented by its center (x_c, y_c). Depending on the location of the center, we decide where the object lies. In this work, we have only considered objects that completely lie in only one region; that is, the object is not spread across two or more regions. In the same Figure 9, three objects placed in three regions are shown as examples.

The X position factor x_p represents the importance of the X coordinate of the object's center, in relation to the object's attention drawing capability. Therefore, the objects in the middle region should have the highest x_p. The objects on the left or right regions should have lower x_p values. We can treat the right boundary line of the middle region (i.e., 2/3 of the screen width) to represent a *threshold* position $x_{\text{threshold}}$. The x_p of any object that lies on the left side of this threshold

is simply the X coordinate of the object's center, as the X coordinate increases linearly from the leftmost bottom point of the screen to the threshold line. We want to reduce the position factor for the objects lying on the right-side of the threshold. Thus, we propose to calculate x_p of the objects on the right region as $X - x_c$, where $X \times Y$ is the screen resolution.

In one of our empirical studies, we observed the eye movement behavior of 20 participants on web pages with an eye tracker (details discussed later). We found that the objects were seen most of the time in top to bottom order. Thus, we propose the y component y_p to be simply y_c.

Equation (6) summarizes our formulation of the position factors of an object, where $\text{obj}_i^{x_p}$ and $\text{obj}_i^{y_p}$ denote the X and Y position factors of the object, respectively:

$$\text{obj}_i^{x_p} = \begin{cases} x_c & x_c \leq x_{\text{threshold}} \\ X - x_c & x_c > x_{\text{threshold}}. \end{cases} \tag{6}$$

As Itti et al. [13] explained, if an object is salient in an image, it gets the attention from the user. Saliency defines how different an object is with respect to the other objects and the background. If an object is different in color or intensity, no matter whether it is more or less, from other objects, it attracts the user's attention. To implement such phenomena, we have taken the difference ratio in the first two factors. In case of the size and position factors, such phenomena need not apply. Therefore we only considered the simple ratio for size and no ratio for the position factors.

4.5. Proposed Attention Factor. We first normalize the individual factors to a fixed range [0-1], as shown in the following

FIGURE 10: An example page used in the study.

(the symbols used are self-explanatory and k is the total number of objects in the web page):

$$I_i = \frac{\mathrm{obj}_i^{\mathrm{ic}}}{\mathrm{Max}\left(\mathrm{obj}_1^{\mathrm{ic}}, \mathrm{obj}_2^{\mathrm{ic}}, \ldots, \mathrm{obj}_k^{\mathrm{ic}}\right)},$$

$$C_i = \frac{\mathrm{obj}_i^{\mathrm{cc}}}{\mathrm{Max}\left(\mathrm{obj}_1^{\mathrm{cc}}, \mathrm{obj}_2^{\mathrm{cc}}, \ldots, \mathrm{obj}_k^{\mathrm{cc}}\right)},$$

$$S_i = \frac{\mathrm{obj}_i^{\mathrm{size}}}{\mathrm{Max}\left(\mathrm{obj}_1^{\mathrm{size}}, \mathrm{obj}_2^{\mathrm{size}}, \ldots, \mathrm{obj}_k^{\mathrm{size}}\right)}, \qquad (7)$$

$$X_i = \frac{\mathrm{obj}_i^{x_p}}{\mathrm{Max}\left(\mathrm{obj}_1^{x_p}, \mathrm{obj}_2^{x_p}, \ldots, \mathrm{obj}_k^{x_p}\right)},$$

$$Y_i = \frac{\mathrm{obj}_i^{y_p}}{\mathrm{Max}\left(\mathrm{obj}_1^{y_p}, \mathrm{obj}_2^{y_p}, \ldots, \mathrm{obj}_k^{y_p}\right)}.$$

We propose af to be the linear weighted combination of the above factors, as shown in the following, where a, b, c, d, and e are the weights whose values range from 0 to 1:

$$\mathrm{af}_i = a \cdot I_i + b \cdot C_i + c \cdot S_i + d \cdot X_i + e \cdot Y_i. \qquad (8)$$

Our proposed model works as follows. Given any input web page, we first calculate the factor values for each object using (1), (4), (5), and (6). From the factor values, af_i for each object

is computed using (8). On the basis of af_i, the objects are ranked (the higher the af, the higher the rank, that is, the lower the rank value with rank 1 implying the highest rank). In order to find the weights in (8), we carried out an empirical study as discussed in the next section.

5. Data Collection for Estimation of the Weights

In the empirical study, eye gaze data were collected from 16 participants for a set of image objects and analyzed.

5.1. Experimental Setup and Participants. The Tobii x50 Eye tracker (http://www.tobii.com/) was used to collect the gaze data. Tobii is a stand-alone eye tracker which can be integrated to any monitor. We used a $17''$ LG Flatron screen with 1280×1024 resolution for the experiments. The gaze data obtained from the eye tracker were analyzed with the ClearView 2.7 by Tobii Technology.

We designed 16 *pages* for the experiment using Web Page Maker v2. Figure 10 shows an example page used in the study. Each of these pages contained 6 image objects only. The objects were numbered $O_1, O_2, \ldots, O_6$ from left to right and top to bottom manner as shown in Figure 10. All the pages contained the same 6 objects, but with different visual properties. The brightness, contrast, size, and position of the objects were varied using the online image editor

TABLE 2: The 16 arrangements of the PPTs designed for data collection.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	P_{16}
S_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
S_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1
S_3	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1	2
S_4	4	5	6	7	8	9	10	11	12	13	14	15	16	1	2	3
S_5	5	6	7	8	9	10	11	12	13	14	15	16	1	2	3	4
S_6	6	7	8	9	10	11	12	13	14	15	16	1	2	3	4	5
S_7	7	8	9	10	11	12	13	14	15	16	1	2	3	4	5	6
S_8	8	9	10	11	12	13	14	15	16	1	2	3	4	5	6	7
S_9	9	10	11	12	13	14	15	16	1	2	3	4	5	6	7	8
S_{10}	10	11	12	13	14	15	16	1	2	3	4	5	6	7	8	9
S_{11}	11	12	13	14	15	16	1	2	3	4	5	6	7	8	9	10
S_{12}	12	13	14	15	16	1	2	3	4	5	6	7	8	9	10	11
S_{13}	13	14	15	16	1	2	3	4	5	6	7	8	9	10	11	12
S_{14}	14	15	16	1	2	3	4	5	6	7	8	9	10	11	12	13
S_{15}	15	16	1	2	3	4	5	6	7	8	9	10	11	12	13	14
S_{16}	16	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

(http://www.online-image-editor.com/). While designing the web pages, we ensured that that the feature values are not too close. We did this through trial and error, that is, designed many pages and only used those that satisfied the criteria.

Sixteen volunteers took part in the study. There were 5 female and 11 male participants, with an average age of 22.71 yrs (in the age group of 21–25 years). Among them, 6 participants wore spectacles. None of the participants was color blind. The participants on average had 4.65 years of computer experience. There were 3 participants who were already familiar with Tobii.

5.2. Procedure and Result. Power point presentations (PPTs) were made from the 16 pages, with each page put into a slide. In between two pages (slides), a blank slide was inserted to make the participant reset his/her eye position.

Each participant was asked to view one PPT. Therefore, there were 16 PPTs. The order of the pages in each PPT was varied following the *Latin square method* [21, 22] (to account for the learning bias). The 16 arrangements (denoted by S_i, where $1 \leq i \leq 16$) of the pages (numbered 1–16) for each participant (denoted by P_i, where $1 \leq i \leq 16$) are shown in Table 2.

Each participant was asked to view the pages through a slideshow of his/her assigned PPT. There was a fixed time interval (5 secs) between two successive slides in all the PPTs. Except for the 3 participants familiar with the Tobii eye tracker, the working of the eye tracker was explained to the rest of the participants.

After this briefing and training session, each participant was taken to the experimental setup in isolation and his/her eye movements were calibrated with the eye tracker. Subsequently, the participant was shown the slides corresponding to him/her (see Table 2). Every participant saw the slideshow continuously without any interruption. It took about 2.5 minutes for each participant to complete the tasks. The gaze plots of the participants for each page was recorded in this

TABLE 3: The rank of the six objects of Figure 10, obtained from the gaze plots of the 16 participants.

	O_1	O_2	O_3	O_4	O_5	O_6
P_1	2	1	3			
P_2	1	3	4		2	5
P_3	2	3	4	1	6	5
P_4		2	1		3	
P_5	2	6	5	3	4	1
P_6	3		5	4	2	1
P_7	4	1	5		2	3
P_8	2	1		3	4	
P_9				2	1	
P_{10}	5	2	3	1	4	
P_{11}	2	3	4	1	6	5
P_{12}	1			2		
P_{13}	1	2	3	4		
P_{14}				2	1	
P_{15}	3	2	1	4	5	6
P_{16}	5	6	1	4	2	3

session, which were obtained using the ClearView software. In total, we had collected 16 × 16 = 256 gaze plots. Figure 11 shows an example of the gaze plots recorded.

5.3. Result Analysis. From the eye gaze data, we assigned a *rank* to each object according to the order the participants followed to attend the objects (which we got from the time stamps recorded by the system). For example, if an object was attended first by a participant, it was ranked 1. The last attended object was given the highest rank (which is the total number of objects). There were some objects which were not attended by some participants. Those unattended objects did not get any rank. In order to assign ranks, we discarded all the fixations where time spent was less than 240 ms as these fixations are known to be unconscious [23]. In this
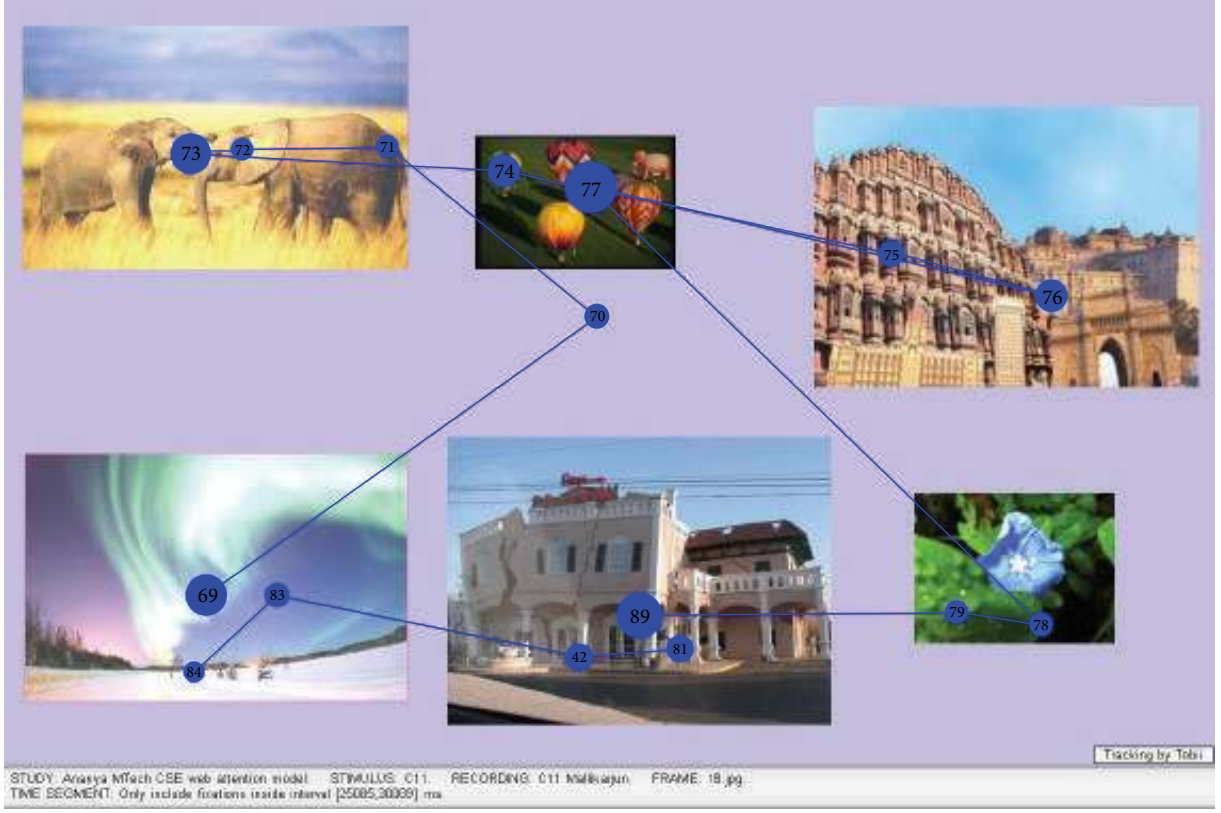
FIGURE 11: An example gaze plot of a participant for the page shown in Figure 10. The blue circles show the point of gaze. The larger the circle, the larger the gaze duration at the point.

way, we constructed an *observed rank table* for each of the 16 pages (obtained from the gaze data of the 16 participants). An example is given in Table 3, which shows the ranks observed from all the participants (denoted by P_1 to P_{16}) for the objects of the page shown in Figure 10.

It may be observed from Table 3 that the same objects got different ranks from participants. We needed to combine these ranks for further analysis. We adopted the following approach based on the *preferential voting scheme* [24]. From the observed rank table of a web page, we constructed a second table. Each row in this table corresponds to a particular object and the ith column represents the total number of the ith ($1 \leq i \leq n$, where n is the total number of objects) ranks obtained by that object (which is derived from the observed rank table). The table for the page of Figure 10, derived from Table 3, is shown in Table 4. As an explanation, consider the 5th row and 4th column of Table 4. The entry is 3, which indicates that the total number (from all participants) of the 4th rank obtained by the 5th object, that is, O_5, was 3.

In order to combine the observed ranks from different users, an index called the *combined index* [24] was formed. The combined index, denoted by z, was computed using the expression shown in (9), where v_{ij} denotes the total number of jth rank received by the ith object and w_j is the weight:

$$z = \sum_{j=1}^{k} w_j v_{ij}. \tag{9}$$

TABLE 4: Objects versus ith ranks for the page shown in Figure 10, which is derived from Table 3.

Object	Rank					
	1	2	3	4	5	6
O_1	3	5	2	1	2	0
O_2	3	4	3	0	0	2
O_3	3	0	3	3	3	0
O_4	3	3	2	4	0	0
O_5	2	4	1	3	1	2
O_6	2	0	2	0	3	1

The weights are assumed to form a monotonically decreasing sequence with $w_j - w_{j+1} \geq d(j, \varepsilon)$ for $j = 1, 2, \ldots, k - 1$ and $w_k \geq d(j, \varepsilon)$. The function $d(j, \varepsilon)$ is the *discrimination intensity function* and ε is called the *discriminating factor*. The function $d(j, \varepsilon)$ is nonnegative and nondecreasing in ε. In this work, we have used $d(j, \varepsilon) = \varepsilon/j$ with $\varepsilon = 0.01$. (We have actually experimented with three discrimination functions with different values of the discriminating factor. The particular values we used turned out to give the best results.)

We made one important observation in our study—users fixate upon at least two locations. From this observation, we can conclude that the object which gets a relatively higher number of first two ranks has higher chances of being seen by

the user early. So, the object which gets maximum number of first two ranks is the most important object.

Thus, we check the $(v_{i1}+v_{i2})$s for all the objects and choose the object for which $(v_{i1} + v_{i2})$ is maximum. The optimization problem corresponding to that object is solved and we get the values of the weights w_i from the solution. Using these weights, we then calculate the z values of all the other objects. The higher the z value, the lower the rank. In this way, we can rank all the objects in a web page. This is the *empirical* rank of the object.

As an example of the procedure, consider the page of Figure 10. From Table 4, we can see that O_1 gets the maximum number (8) of the first two ranks. Hence, the optimization problem corresponding to the object O_1 is chosen for maximization. The formulation of the optimization problem along with the set of constraints for the objects of the page shown in Figure 10 is shown as follows:

Maximize z, where

$$z = 3w_1 + 5w_2 + 2w_3 + 1w_4 + 2w_5 + 0w_6 \qquad (10)$$

subject to

$$3w_1 + 5w_2 + 2w_3 + 1w_4 + 2w_5 + 0w_6 \leq 1,$$

$$3w_1 + 4w_2 + 3w_3 + 0w_4 + 0w_5 + 2w_6 \leq 1,$$

$$3w_1 + 0w_2 + 3w_3 + 3w_4 + 3w_5 + 0w_6 \leq 1,$$

$$3w_1 + 3w_2 + 2w_3 + 4w_4 + 0w_5 + 0w_6 \leq 1,$$

$$2w_1 + 4w_2 + 1w_3 + 3w_4 + 1w_5 + 2w_6 \leq 1,$$

$$2w_1 + 0w_2 + 2w_3 + 0w_4 + 3w_5 + 1w_6 \leq 1,$$

$$w_1 - w_2 \geq 0.01,$$

$$w_2 - w_3 \geq 0.01,$$

$$w_3 - w_4 \geq 0.01,$$

$$w_4 - w_5 \geq 0.01,$$

$$w_5 - w_6 \geq 0.01,$$

$$w_6 \geq 0.01.$$

$$(11)$$

We can use the simplex method [25] to solve the above problem. We used the simplex solver tool (http://www.zweigmedia.com/RealWorld/simplex.html) and got the optimal solution as $z = 1$ with the weights $w_1 = 0.0923077$, $w_2 = 0.0823077$, $w_3 = 0.0723077$, $w_4 = 0.0623077$, $w_5 = 0.0523077$, and $w_6 = 0.01$. Using these weights, we calculated the combined indices for other objects (using (9) and Table 4). From these combined indices, we calculated the empirical ranks of the objects (on the basis of the fact that the higher the z value, the lower the rank). Following this approach, we found the empirical ranks of the objects $O_1, O_2, O_3, O_4, O_5,$ and O_6 of the page of Figure 10 to be 1, 4, 5, 2, 3, and 6, respectively.

TABLE 5: The Pearson product moment correlation coefficients for each factor in the visibility measure (8) and the empirical ranks of the objects.

Factor	Correlation coefficient
Intensity contrast	0.09
Chromatic contrast	−0.17
Size	−0.44
X-position	−0.47
Y-position	−0.1

5.4. Estimation of Model Parameters. The proposed attention factor is a weighted linear combination of the five components (corresponding to the four factors) as shown in (8). We checked the correlation between each of the components and the empirical ranks obtained from the eye tracking data. For this purpose, we calculated the Pearson product-moment correlation coefficient [26]. The correlation coefficients are shown in Table 5.

It may be noted in Table 5 that the correlation coefficient between the intensity contrast and the empirical ranks was 0.09. A correlation coefficient in the range [0–0.09] implies that there is not enough correlation between the two variables and we can ignore the effect of the factor. Consequently, we eliminated intensity contrast from the expression of the visibility measure.

In order to determine the weights in (8), we first scaled each of the correlation coefficients as $\text{coeff}_{\text{scaled}} = |\text{coeff}|/(0.17 + 0.44 + 0.47 + 0.1 = 1.18)$ where $|\text{coeff}|$ denotes the absolute value of the corresponding correlation coefficient. Then, we set the weights equal to the corresponding scaled correlation coefficients, which were found to be (approximately) 0.14, 0.37, 0.4, and 0.09, respectively. The attention factor with the weights is shown in

$$\text{af}_i = 0.14C_i + 0.37S_i + 0.4X_i + 0.09Y_i. \qquad (12)$$

6. Empirical Studies for Model Validation

In order to ascertain the efficacy of our proposed model, we carried out another study. We had designed 20 pages, each containing six image objects (different from the pages used in the previous study). We collected eye gaze data of 20 participants using an identical setup and procedure as before (Section 5). Among the participants, there were 12 males and 8 females (age group: 21–25 years, mean age: 23.8 yrs, all having normal vision without any color blindness). They were graduate students and regular computer users. The total number of gaze plots collected was $20 \times 20 = 400$.

Using the model of af_i (12), we calculated the ranks for the 120 objects (predicted rank). From the empirical data, we determined the empirical ranks, following the procedure described in Section 5.3. Whenever these two ranks were the same for an object, we termed it as an *exact match*. When predicted and actual rank varied by ±1, we called it as a *partial match*. In terms of these two types of matches, the accuracy

TABLE 6: Results of the validation studies.

	Proposed model			Visual saliency model		
	Exact match	Partial match	Accuracy (%)	Exact match	Partial match	Accuracy (%)
Validation study 1 (120 objects)	55	40	79.16	23	42	54.16
Validation study 2 (122 objects)	24	30	59.01	23	24	38.52

of prediction was calculated as shown in (13), where N is the total number of objects:

$$\text{Accuracy} = \frac{\text{exact match} + \text{partial match}}{\text{total number of objects}} \times 100\%. \quad (13)$$

The results of the analysis of the gaze plots for both the sets in terms of the exact matches, partial matches, and accuracy are shown in Table 6 (the row for *validation study 1*). In order to compare the performance of our proposed model, we also calculated the ranks of the objects using the visual saliency model, which are also shown in Table 6. As the model predicts only the salient locations, we mapped those to the objects for prediction of object ranks.

One issue is the applicability of the model for general web page design. As we mentioned, we developed and evaluated the model with artificial pages containing only image objects. Typical web pages, however, contain different types of objects. Presence of these different object types with their unique properties (e.g., typographical properties of text objects such as length, width, spacing, style, and decoration) can affect attention drawing abilities of these as well as image objects placed in between. So, the question naturally arises, how realistic our proposed model is? In order to answer this question, we carried out a third empirical study involving 17 web pages (a subset of the 103 web pages mentioned in Section 3) containing 122 objects (all types of objects). The percentage area coverage of various objects in those images is shown in Figure 12.

Eye gaze data of 17 participants (3 of them common to the previous validation study) were collected for the 17 web pages following identical setup and procedure as in Section 5. In total, we collected $17 \times 17 = 289$ gaze plots. We calculated the predicted ranks and empirical ranks as before for the objects and determined accuracy, shown in Table 6 (the row for validation study 2). The corresponding results using the visual search model are also shown for comparison.

It may be noted that some of the 17 web pages contained more than 6 objects. For those web pages, we had a relatively larger optimization problem to solve. However, we observed that the participants hardly saw more than 6 objects in those pages within the 5 seconds. Therefore, we assumed that the empirical ranks would not vary much if we consider only up to the 6th rank for web pages having more than 6 objects. Based on this assumption, we had chosen to form optimization problems considering only up to the 6th rank, by setting the weights corresponding to the higher ranks to 0.
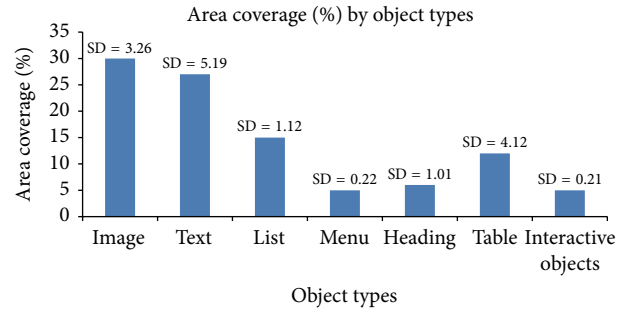

FIGURE 12: The histogram showing the average area occupied by the different objects in the web pages used in the second validation study.

7. Discussion

Our objective was to come up with an attention model of web page viewers, which should predict the likely sequence of viewing the web page objects. We aimed towards a simple and intuitive model: something that will be easier for designers to understand and use. This is important since our end objective is to use the model as a tool support in the design environment. The tool should be *usable* by designers who are not familiar with the cognitive theories of human attention. As may be noted, the model is based on *only* three features, out of which two (size and position) are very easy to understand. The features are linearly related, which again is easier to comprehend than some nonlinear relationships. These qualities are in contrast to other attention models (e.g., the saliency models), which involve many more features and complex relationships. The results of the validation study (Table 6) show that the proposed model was able to predict the ranks of objects with reasonably high accuracy (about 80% and 60% in the first and second studies, resp., with an overall average of about 70%). These accuracy rates indicate that the trade-off between complexity of the model and accuracy is suitably addressed by our proposed model.

We envisage a semiautomatic web page design environment with the model. We plan to implement this model as a tool in the existing design platforms. The tool shall allow a designer to *specify* objects on a page. Based on the selections, the tool shall compute ranks of those objects. As we mentioned before, the rank of an object indicates its likelihood of drawing the user's attention before other objects. The higher the rank (i.e., the lesser the rank value), the more likely the object is to draw attention first. For example, if two objects have ranks 3 and 2, the object having rank 2

will most likely be drawing the user's attention before the other object. Thus, from the predicted ranks, the designer can determine if the *important* objects will indeed be able to draw the attention before the other objects. With this knowledge, the designer can draw conclusion about the potential usability of the page: if there are too many objects (as in Figure 1) and the most important objects are getting lower ranks, then the design needs to be improved as the user is likely to move to the next page without viewing the most important content. Consequently, the designer can take corrective actions (i.e., reducing the number of objects or making the important objects more attractive by changing their visual properties such as contrast, size, and position).

As we found out in the validation studies, the proposed model outperforms the saliency model (about 26% and 22% improvements in the first and second validation studies, resp.). We are not aware of any other model of human attention that can be used to address the problem of predicting attention sequence on web page objects. Thus, we feel that our proposed model can provide a good alternative.

The first validation study indicates that the model accuracy is expected to be high if the web page contains mostly image objects. In the second validation study, the prediction accuracy was found to be about 60%. In the second study, we considered each object as an image object to compute attention factor. Since the attention drawing abilities of text and other object types are likely to be influenced by factors other than those considered for image objects, the accuracy dropped. Therefore, it may appear that the proposed model in its current form is only applicable in artificial situations, such as the first validation set. However, this is not necessarily the case. In our survey of web pages (Section 3), we found a significant proportion (about 43%) of the surveyed pages containing mostly image objects. The 17 pages we had chosen for the second validation set were more balanced and hence the drop in prediction accuracy. If we assume that the surveyed pages are representative of the reality, then we can conclude that the proposed model in its current form can be deployed in many practical design situations. Apart from that, the 60% accuracy for the second set also indicates that, although the model is developed primarily for image objects, it still can be used to predict user's attention sequence for web pages containing different object types with much better accuracy than the existing model (nearly 22% better).

Thus, we expect our proposed model to significantly enhance the present state of the art. However, the model performance for web pages not having majority of image objects can be improved in many ways. We need to identify and incorporate factors influencing attention drawing for nonimage objects, particularly for textual objects as they are found to be the next most important object types after image (Figure 7). Also, we have formulated the position factors in terms of the center of the objects. Implicit in this formulation is the assumption that the objects do not span more than one of the three regions (left, right, and middle), in which we divide a web page. Not all the web pages in the second validation set conformed to this requirement, which may be another reason for the drop in accuracy. Thus, the definition of the position factors can be modified. In addition, the presence of small animations, which are common in web pages nowadays, need to be considered. Undoubtedly, animations increase attention drawing. However, improper use of animations may lead to user annoyance, which may make the user move away from the page. Therefore, predicting the visibility of animation objects will be a useful extension to the proposed model.

8. Conclusion

In this work, we proposed a model to predict the attention behavior of exploratory web page users. In order to achieve the objective, we proposed the idea of dividing web page contents into objects. We developed a model to predict user's attention sequence on these objects. Although we considered image objects to develop the model, empirical results indicate the usefulness of the model in general web page design.

The present model can be improved in several ways. We can incorporate factors for other types of objects (in particular, the text objects) in the model. The position factor can be further enhanced and generalized. The utility of animation in increasing visibility can be modeled and incorporated. We plan to work on these issues in the future.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] A. Bojko, "Using eye tracking to compare web page designs: a case study," *Journal of Usability Studies*, vol. 1, no. 3, pp. 112–120, 2006.

[2] A. Bojko and A. Stephenson, "It's all in the eye of the user: how eye tracking can help answer usability questions," *User Experience*, vol. 4, no. 1, 2005.

[3] N. Webb and T. Renshaw, "Eye tracking in HCI," in *Research Methods for Human-Computer Interaction*, P. Cairns and A. L. Cox, Eds., pp. 35–69, 2011.

[4] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Reviews Neuroscience*, vol. 2, pp. 1–11, 2001.

[5] J. K. Tsotsos and A. Rothenstein, "Computational models of visual attention," *Scholarpedia*, vol. 6, no. 1, article 6201, 2011.

[6] J. K. Tsotsos, *A Computational Perspective on Visual Attention*, The MIT Press, Cambridge, Mass, USA, 2011.

[7] J. M. Wolfe, "Guided Search 2.0 A revised model of visual search," *Psychonomic Bulletin & Review*, vol. 1, no. 2, pp. 202–238, 1994.

[8] J. M. Wolfe, "Guided search 4.0: current progress with a model of visual search," in *Integrated Models of Cognitive Systems*, W. Gray, Ed., pp. 99–119, Oxford University Press, New York, NY, USA, 2007.

[9] B. Liu, G. Francis, and G. Salvendy, "Applying models of visual search to menu design," *International Journal of Human Computer Studies*, vol. 56, no. 3, pp. 307–330, 2002.

[10] L. Itti, "Visual salience," *Scholarpedia*, vol. 2, no. 9, p. 3327, 2007.

[11] Y. Hu, X. Xie, W. Y. Ma, L. T. Chia, and D. Rajan, "Salient region detection using weighted feature maps based on the human visual attention model," in *Advances in Multimedia Information Processing—PCM 2004*, vol. 3332 of *Lecture Notes in Computer Science*, pp. 993–1000, Springer, Berlin, Germany, 2005.

[12] L. Itti, C. Gold, and C. Koch, "Visual attention and target detection in cluttered natural scenes," *Optical Engineering*, vol. 40, no. 9, pp. 1784–1793, 2001.

[13] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[14] Q. Zhang and H. Xiao, "A hierarchical computational model of visual attention using multi-layer analysis," in *Proceedings of the 2nd International Conference on Communication Systems, Networks and Applications (ICCSNA '10)*, pp. 267–270, July 2010.

[15] J. D. Still and C. M. Masciocchi, "A saliency model predicts fixations in web interfaces," in *Proceedings of the 5th International Workshop on Model Driven Development of Advanced User Interfaces (MDDAUI '10)*, pp. 25–28, April 2010.

[16] E. Peli, "Contrast in complex images," *Journal of the Optical Society of America A*, vol. 7, no. 10, pp. 2032–2040, 1990.

[17] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, Mass, USA, 1992.

[18] J. H. Y. Lau and B. E. Shi, "Improved illumination invariance using a color edge representation based on double opponent neurons," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '08)*, pp. 2735–2741, Hong Kong, June 2008.

[19] T. W. Lee, T. Wachtler, and T. J. Sejnowski, "Color opponency is an efficient representation of spectral properties in natural scenes," *Vision Research*, vol. 42, no. 17, pp. 2095–2103, 2002.

[20] G. Buscher, E. Cutrell, and M. R. Morris, "What do you see when you're surfing? Using eye tracking to predict salient regions of web pages," in *Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems (CHI '09)*, pp. 21–30, ACM, New York, NY, USA, April 2009.

[21] A. Blandford, A. L. Cox, and P. Cairns, "Controlled experiments," in *Research Methods for Human-Computer Interaction*, P. Cairns and A. L. Cox, Eds., pp. 1–16, Cambridge University Press, 2011.

[22] J. Lazar, J. H. Feng, and H. Hochheiser, Eds., *Research Methods in Human-Computer Interaction*, John Wiley & Sons, 2010.

[23] H. Fengpei, "The studies of eye tracking and usability test," in *Proceedings of the 7th International Conference on Computer-Aided Industrial Design (CAIDC '06)*, pp. 1–5, IEEE, Hangzhou, China, November 2006.

[24] W. D. Cook and M. Kress, "A data envelopment model for aggregating preference rankings," *Management Science*, vol. 36, no. 11, pp. 1302–1310, 1990.

[25] F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research*, McGraw-Hill, 8th edition, 2004.

[26] J. L. Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *The American Statistician*, vol. 42, no. 1, pp. 59–66, 1988.