Exploring Human Eye Behaviour using a Model of Visual Attention

Oyewole Oyekoya and Fred Stentiford
University College London, Adastral Park Campus; ooyekoya, f.stentiford@ee.ucl.ac.uk

Abstract

It is natural in a visual search to look at any object that is similar to the target so that it can be recognised and a decision made to end the search. Eye tracking technology offers an intimate and immediate way of interpreting users' behaviours to guide a computer search through large image databases. This paper describes experiments carried out to explore the relationship between gaze behaviour and a visual attention model that identifies regions of interest in image data. Results show that there is a difference in behaviour on images that do and do not contain a clear region of interest.

1. Introduction

Understanding the movement of the eye over images is critical for improving our ability to manage and exploit image data. Eye tracking experiments have been performed for various purposes such as understanding the human visual process and improving access to digital data. In this paper the gaze behaviour of participants is compared with data obtained through a model of Visual Attention (VA) [2] to detect differences in behaviour arising from varying image content. Regions of Interest (ROI) are identified both by human interaction and prior analysis and used to explore aspects of vision. Images with and without obvious subjects were used in this work to accentuate any behaviour differences that might be apparent.

1.1. Related Work

The tracking of eye movements has been employed as a pointer and a replacement for a mouse [10], to vary the screen scrolling speed [9] and to assist disabled users [8]. However, this work has concentrated upon replacing and extending existing computer interface mechanisms rather than creating a new form of interaction. Indeed the imprecise nature of saccades and fixation points has prevented these approaches from yielding benefits over conventional human interfaces.

Privitera and Stark [3] compare algorithmically detected ROIs with human detected ROIs as a criterion for evaluating and selecting bottom-up, context free algorithms. Jaimes, Pelz et al [4] compare eye movement across categories and links category-specific eye tracking results to automatic image classification techniques. Dasher's fast hands-free writing [7] use a method for text entry based on inverse arithmetic coding that relies on gaze direction and which is faster and more accurate than using an on-screen keyboard.

2. Experimental Process

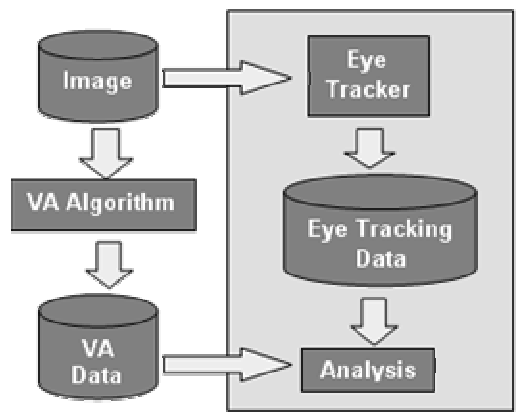

Figure 1. System diagram

For each image, the VA Algorithm is applied to identify regions of interest. The same image is viewed by a human participant using the EYEGAZE eye tracker. The eye tracking data and the VA data are combined and analysed by identifying the coordinates of the gaze points on the image and obtaining the scores from the corresponding VA data.

Results with four subjects on six images are reported below. Three images contained obvious regions of interest, and the remainder contained unclear or no regions of interests.

All participants had normal or corrected-to-normal vision and had no knowledge of the purpose of the study. Over the course of the experiment, participants were presented a series of images for 5 seconds each separated by displays of a blank screen followed by a central black dot on a white background (Figure 2). These images were displayed on a 15" LCD Flat Panel Monitor at a resolution of 1024x768 pixels. All participants were encouraged to minimise head movement and were asked to focus on the dot before each image was loaded.

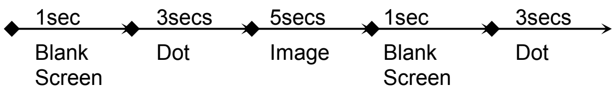

Figure 2. Display Sequence

The eye tracker allows for head movement of up to 1.5 inches (3.8cm) and uses the pupil centre corneal reflection method to determine gaze direction. Calibration is needed to measure the properties of each subject's eye before the start of each experimental run. The processing of information from the eye tracker is done on a 128MB Intel Pentium 3 system with a video frame grabber board.

2. Visual Attention Model

The model used in this work [2] assigns high values of visual attention to pixels when neighbouring pixel configurations do not match identical positional arrangements in other randomly selected neighbourhoods in the image. This means that textures and other features that are common in an image will tend to suppress attention values in their neighbourhood.

Let a set of measurements **a** correspond to a location

$$\mathbf{x} = (x_1, x_2) \text{ where } \mathbf{a} = (a_1, a_2, a_3)$$

Define a function **F** such that **a** = **F**(**x**). Consider a neighbourhood N of **x** where

$$\{\mathbf{x'} \in N \text{ iff } |x_i - x'_i| < \varepsilon_i \ \forall\ i\}$$

Select a set of m random points S_x in N where

$$S_x = \{\mathbf{x'}_1, \mathbf{x'}_2, \mathbf{x'}_3, ..., \mathbf{x'}_m\}.$$

Select another location **y**.

Define the set $S_y = \{\mathbf{y'}_1, \mathbf{y'}_2, \mathbf{y'}_3, ..., \mathbf{y'}_m\}$ where

$$\mathbf{x} - \mathbf{x'}_i = \mathbf{y} - \mathbf{y'}_i .$$

The neighbourhood of **x** is said to match that of **y** if

$$|F_j(\mathbf{x}) - F_j(\mathbf{y})| < \delta_j \text{ and } |F_j(\mathbf{x'}_i) - F_j(\mathbf{y'}_i)| < \delta_j \ \forall\ i,j.$$

A location **x** will be worthy of attention if a sequence of t neighbourhoods matches only a relatively small number of other neighbourhoods in the space. In the case of a two-dimensional still image, m pixels **x'** are selected in the neighbourhood of a pixel **x**. Each of the pixels might possess three colour intensities, so **F**(**x'**) = **a** = (r, g, b). Typically t = 100, m = 2, ε_i = 2 and δ_j =80 in rgb space.

The visual attention estimator has been implemented as a set of tools that processes images and produces corresponding arrays of attention values (VA scores). The attention values are thresholded and those above the threshold are displayed as a map using a continuous spectrum of false colours with the scores exceeding a certain threshold being marked with a distinctive colour.

3. Results

Figures 4 to 9 show the images used in the experiments together with their VA maps and graphs of four subjects. The saccades and fixations performed by the subjects on each of the images were recorded through the eyetracking system. The VA score that corresponded to the pixel at each fixation point was associated with the time of the fixation and plotted as graphs for study in units of 20ms. It was observed that there was considerable variation in behaviour over the four subjects, but all viewed the regions with the highest VA scores early in the display period.

The variance of the VA score (x) over time is given by ν where

$$\nu = \frac{n\sum x^2 - \left(\sum x\right)^2}{n(n-1)}$$

The variance ν measures the average spread or variability of the data series x. The variances of the VA scores for the duration of the display over the six images for each subject are shown in Table 1 and Figure 3.

		Subjects			
		1	**2**	**3**	**4**
Unclear ROI	**Image1**	325	193	333	532
	Image2	479	496	328	629
	Image3	389	175	365	197
Obvious ROI	**Image4**	741	687	1094	857
	Image5	1432	1453	1202	1466
	Image6	1246	1226	862	1497

Table 1. Variance of VA score against time

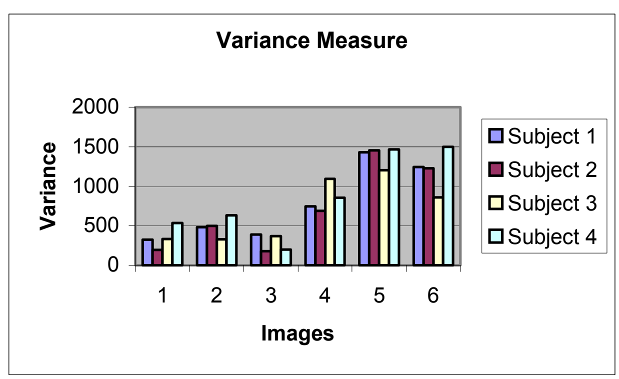

Figure 3. Variance histogram

4. Discussion

The goal of this study was to explore the relationship gaze behaviour and the Visual Attention model described in determining eye movement patterns over different stages of viewing. Results indicate that regions with high VA scores do attract eye gaze for those images studied. However, it was apparent that individual behaviours varied considerably and it was difficult to identify a pattern over such a small amount of data. Nevertheless the results did show that there was a higher variance in VA score over time on images with obvious ROIs due to gaze patterns shifting between areas of high visual attention and the background. This would seem reasonable in view of a natural inclination to make rapid visual comparisons between anomalous material and a relatively predictable background.

Privitera and Stark [3] evaluated 10 different algorithms for detecting regions of interest by comparing output of such algorithms to eye tracking data. They concluded that it was unreasonable to expect an algorithm to be able to predict the location of every region of interest. The framework employed in this work allows for further exploration of gaze behaviour and validation of attention models, hence leading to improved algorithmic detection of regions of interests.

During the experiment, some participants reported that eye blinking and blur (due to continuous screen-stare) were unavoidable. Hence, the eye tracking data for blinking, and off-image gaze points were discarded in the analysis.

5. Conclusions

A substantial part of the gaze of the participants during the first two seconds of exposure is directed at areas of high visual attention as estimated by the model. Many of the saccades for several participants are characterised by frequent movements to and from the areas of high visual attention, which is shown by high variances for images containing salient material. Several participants dwell on the subject areas for longer periods of time but still periodically scan background material. More work is necessary to obtain statistical significance across more images and participants.

The subjects were not given specific tasks when viewing the images in these experiments and this may have introduced some confounding influences. Future work will be focussed on specific retrieval tasks, which should reduce inter-subject variability and at the same time explore new interfaces for content-based image retrieval.

6. Acknowledgements

The authors acknowledge the support of BT Exact Technologies, SIRA and the Engineering and Physical Sciences Research Council in this work.

7. References

[1] O.K. Oyekoya and F.W.M. Stentiford, “Exploring the Significance of Visual Attention by Eye Tracking”, *Proceedings of the London Communications Symposium 2003*, UCL, London, 8-9 Sept, pp. 149-152.

[2] F.W.M. Stentiford, “An estimator for visual attention through competitive novelty with application to image compression,” *Picture Coding Symposium*, Seoul, 24-27 April, 2001.

[3] C.M. Privitera, L. W. Stark, "Algorithms for Defining Visual Regions of Interest: Comparison with Eye Fixations," *IEEE Transactions On Pattern Analysis and Machine Intelligence*, September 2000, Vol. 22, No 9, pp 970-982.

[4] A. Jaimes, J.B. Pelz, T. Grabowski, J. Babcock, and S.-F. Chang, "Using Human Observers' Eye Movements in Automatic Image Classifiers", *Proceedings of SPIE Human Vision and Electronic Imaging VI*, San Jose, CA, 2001.

[5] D. Parkhurst, K. Law, and E. Neibur, "Modelling the role of salience in the allocation of overt visual attention," *Vision Research*, vol. 42, pp. 107-123, 2002.

[6] M. Pomplun & H. Ritter, “A three-level model of comparative visual search,” In M. Hahn & S. C. Stoness, (Eds.), Proceedings of the Twenty First Annual Conference of the Cognitive Science Society, 543-548, 1999.

[7] D.J. Ward and D.J.C. MacKay, “Fast hands-free writing by gaze direction”, Nature 418 pp 838, Aug. 22 2002.

[8] F. Corno, L. Farinetti and I. Signorile, “A cost effective solution for eye-gaze assistive technology,” 2002 IEEE Int Conf. on Multimedia and Expo, August 26-29, Lausanne, 2002.

[9] T. Numajiri, A. Nakamura, and Y. Kuno, “Speed browser controlled by eye movements,” 2002 IEEE Int Conf. on Multimedia and Expo, August 26-29, Lausanne, 2002.

[10] J.P. Hansen, A.W. Anderson, and P. Roed, “Eye gaze control of multimedia systems,” Symbiosis of Human and Artifact (Y. Anzai, K. Ogawa, and H. Mori (eds), Vol 20A, Elsevier Science, pp 37-42, 1995.

[11] G. Marmitt, A.T. Duchowski, ``Modeling Visual Attention in VR: Measuring the Accuracy of Predicted Scanpaths", EuroGraphics 2002 (Short Presentations), September 2-6, 2002, Saarbrucken, Germany.

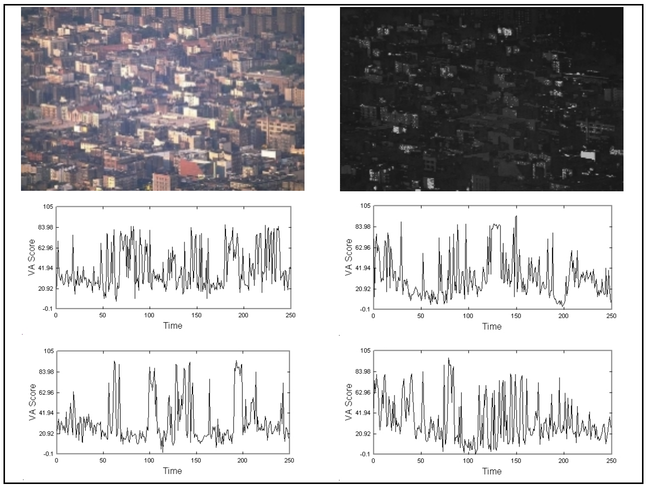

Figure 4. Image 1 with unclear ROI

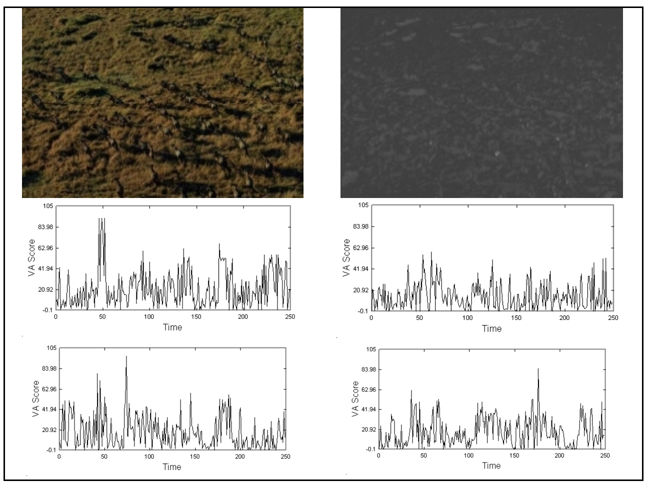

Figure 5. Image 2 with unclear ROI

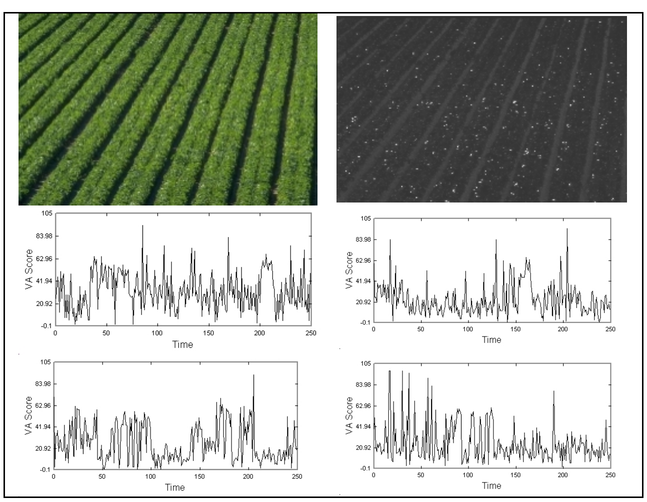

Figure 6. Image 3 with unclear ROI

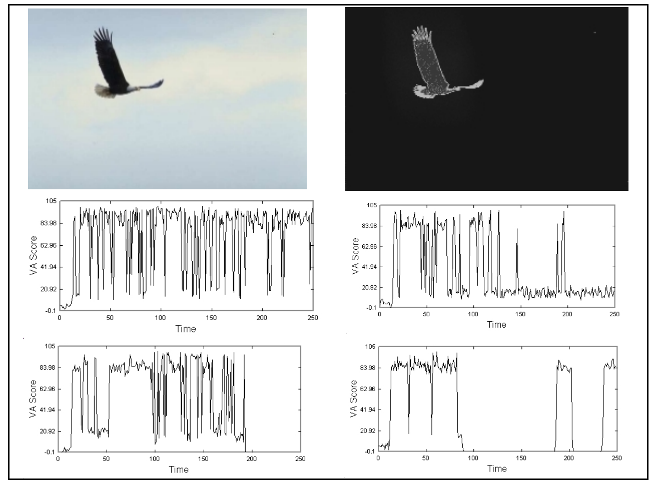

Figure 7. Image 4 with obvious ROI

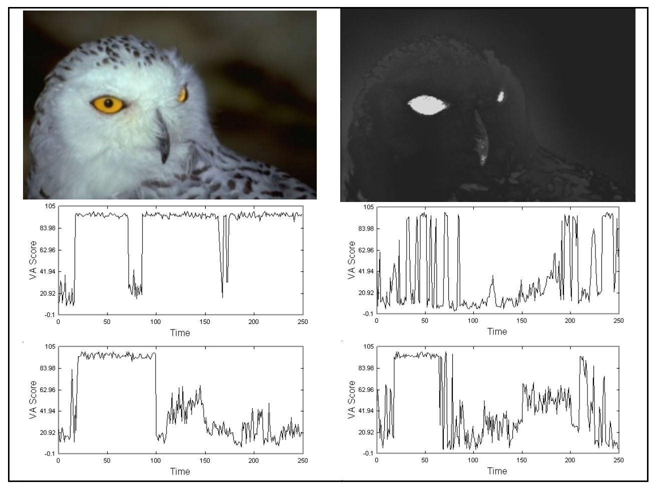

Figure 8. Image 5 with obvious ROI

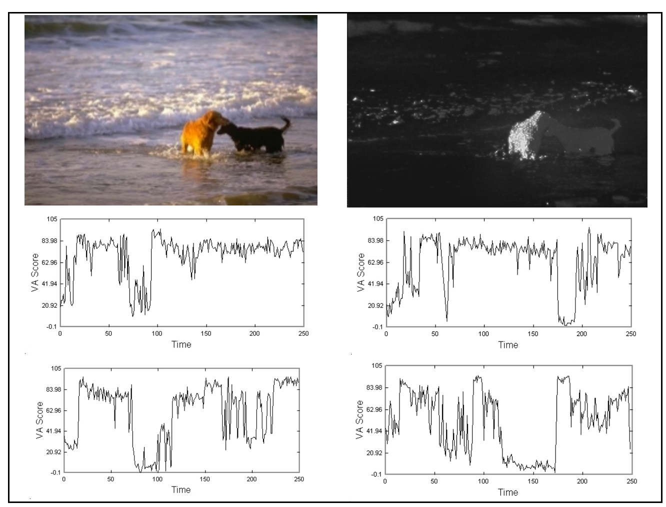

Figure 9. Image 6 with obvious ROI