

EXPLORING THE SIGNIFICANCE OF VISUAL ATTENTION BY EYE TRACKING

O K Oyekoya[†] and F W M Stentiford[‡]

Content Understanding Group, University College London, Adastral Park Campus, Ipswich UK.

ABSTRACT: Understanding the movement of the eye on areas of visual attention should increase our ability to manage and exploit image data tremendously. Our hypothesis is that most humans will look at high VA scoring regions of an image during different stages of viewing. This paper describes the problem and proposes an experiment to confirm our hypothesis. Initial results are presented and the final results will be presented in later papers after a series of experiments.

Keywords: eye tracking, regions of interest, content based image retrieval, visual attention.

1 INTRODUCTION.

In the last few years, research in the field of content based retrieval has focused on facilitating access to multimedia information (e.g., images, video, etc.) in large digital databases. In order to limit the amount of information to be processed, detection of Regions of Interests (ROIs) is necessary so that only regions that may be relevant to the problem at hand are selected for analysis. Understanding the human eye movement and the visual process is deemed to be useful in the construction of useful visual attention (VA) algorithms. Visual attention is the innate ability to spot anomalies in our environment and to take appropriate action as we think necessary. A new measure of visual attention was devised that can be used to identify regions of interest in many categories of images [1, 2, 3]. The measure is based upon the hypothesis that visual attention is to a certain extent dependent upon the disparities between neighbourhoods in the image. When we scan a visual scene (picture), our eyes alternate between rapid jumps (saccades) and brief stops (fixations). Although little information is processed during saccadic movement (because of the fast motion of images across the retina), eye fixations enable us to focus our attention like a spotlight. Therefore, one way of exploring what people pay attention to in any given situation is to use an eye-tracking system (Eyegaze) to record their visual attention strategies (scan paths) and the location and duration (typically in the region of 150 milliseconds) of their fixations.

1.1 Hypothesis

This experiment will test the movement of the eye within specially selected images using the eye tracker with the aim of confirming the hypothesis that most humans will look at high VA scoring regions of an image during various stages of viewing. First, we detect human identified Regions of Interests (hROIs) [4] using the eye tracker on specially chosen images. Then, we compute VA scores for each pixel in the image using a VA algorithm proposed by Stentiford [2, 3] to obtain algorithmically-detected ROIs (aROIs) [4]. The graphs produced by images with ROIs will be compared with the graphs of images without obvious ROIs in order to obtain a measure of success.

2. EXPERIMENTAL METHODS.

Ten research students will be asked to participate in the experiment. Each will participate in two half-hour sessions, with a break every five minutes. All participants should have normal or corrected-to-normal vision and will have little knowledge to the purpose of the study. Over the course of the

[†] ooyekoya@ee.ucl.ac.uk; <http://www.ee.ucl.ac.uk/~ooyekoya>

[‡] f.stentiford@ee.ucl.ac.uk; <http://www.ee.ucl.ac.uk/~fstentif>

experiment, participants are presented images. The images are obtained from digital libraries already gathered from various royalty-free sources. All images will be displayed on a 15" LCD Flat Panel Monitor at a resolution of 1024x768 pixels. Representative images (with and without obvious ROIs) are shown below.

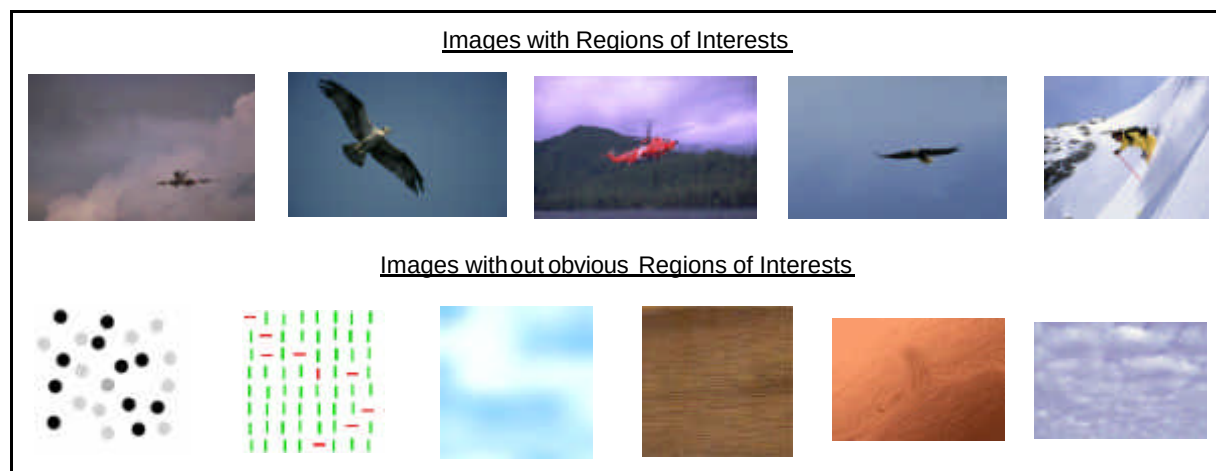


Fig. 1. Sample Images

Prior to beginning the experiment, participants will be asked to fixate on the centre of the canvas (i.e. the blank container on the Eyegaze application). At this time, an image is presented for a period of five seconds. Subsequently, the display will be blanked. Participants view each image once until they have viewed 20 images (phase 1). The process will be repeated with the same images in the same order (phase 2). At the beginning of each section of 20 trials, the eye tracker is re-calibrated.

The procedure to calibrate the Eyegaze System is robust yet fast and easy to perform. The calibration procedure takes approximately 15 seconds. The calibration procedure is fully automatic; no assistance from another person is required. The procedure adapts to the user speed by waiting for the user to fixate clearly on each calibration point before accepting it and moving on to the next point. The procedure accommodates interruptions from the user blinking or looking away from the computer screen. The procedure simply waits for a good fixation before moving to the next calibration point. After the original pass through the calibration points, the procedure tests that the eye was properly fixated on each point by checking that each gaze point prediction is consistent with all the other calibration points. It retakes any calibration points that are inconsistent with other points. The procedure does not accept the full calibration until the overall gaze prediction accuracy and consistency exceed desired thresholds. To achieve high gaze point tracking accuracy, the image processing algorithms in the Eyegaze System explicitly accommodate several common sources of gaze point tracking error such as nonlinear gaze point tracking equations, accommodating head range variation, accommodating pupil diameter variation and accommodating glint straddling pupil edge. A chair with head rest will provide support for chin and forehead in order to minimize the effects of head movements.

The Eyegaze system is an eyetracker designed to measure where a person is looking on a computer screen. Gaze point tracking measurements are made unobtrusively via a remote video camera mounted below the computer monitor. Nothing is attached to the subject. The Eyegaze System tracks the subject's gaze point on the screen automatically and in real time. Gaze point measurements are made at a 60 Hz rate. Image processing and gaze point calculations are performed in software (Trace) on a Windows NT/2000 computer. Gaze direction is determined using the pupil-centre-corneal reflection (PCCR) method.

3. TYPICAL EXPERIMENT RESULTS

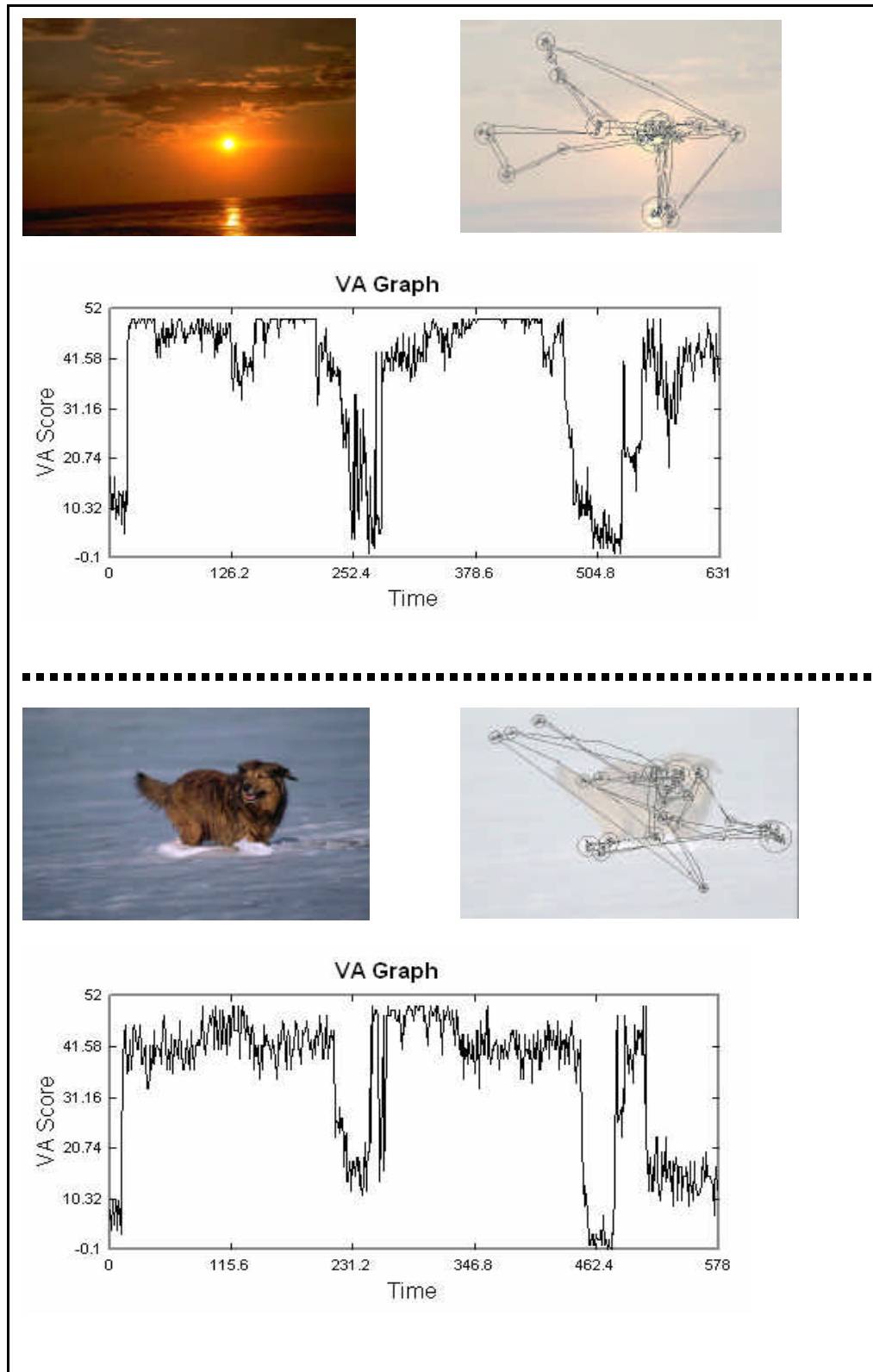


Fig. 2. The score for each pixel of the human identified ROIs is plotted against the viewing time (hundredth of a second). The pixel scores are determined by the VA algorithm. Maximum VA Score is 50. Eye tracking results on the top-left images is shown in the top-right images along with fixation and saccades. Note that most of the fixations are on the region(s) of high interest.

4. CONCLUSIONS & FUTURE WORK

The Trace program displays a user-prepared image on the computer monitor and passively collects the eyegaze activity as a subject observes the screen. After the data collection phase, the eyegaze history is played back both as a time history and as a trace superimposed on the original screen image (fig. 2). The trace may be paused, reversed and replayed at different speeds. The Visual C++ source code is available for modification and will be combined with custom-built data analysis software, which will be extremely useful for future eye-tracking research. It is intended to modify and combine the trace codes with the analysis code to automate the experimental process for future research.

An eye-tracking experiment has been presented that investigates the human eye movement in an image during display. The interesting phenomenon about experiments is that a positive result is just as good as a negative result. If we are able to confirm that humans do look at high regions of interest, then we would be able to enhance our VA algorithm [2, 3] using results obtained from the experiment. However, a negative result will mean that people tend to look at areas of low VA which will be surprising and certainly worthy of further investigation. The results will be carefully analysed to ensure that the data obtained is statistically significant.

It is hoped that this work will lead to a better understanding of visual search processes and new ways of using eye-tracking data to build new interfaces and improve algorithms for image retrieval systems.

ACKNOWLEDGMENTS.

The authors acknowledge the support from BT Exact Technologies and SIRA in this work.

REFERENCES.

- [1] F. W. M. Stentiford, "An attention based similarity measure with application to content based information retrieval," in *Storage and Retrieval for Media Databases 2003*, M. M. Yeung, R. W. Lienhart, C-S Li, Editors, Proc SPIE Vol. 5021, 20-24 Jan, Santa Clara, 2003.
- [2] F. W. M. Stentiford, "An evolutionary programming approach to the simulation of visual attention," Congress on Evolutionary Computation, Seoul, May 27-30, 2001.
- [3] 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.
- [4] Claudio M. Privitera, Lawrence 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, Pg. 970-982.
- [5] A. Jaimes, J.B. Pelz, T. Grabowski, J. Babcock, and S.-F. Chang, "Using Human Observers' Eye Movements in Automatic Image Classifiers" in proceedings of SPIE Human Vision and Electronic Imaging VI, San Jose, CA, 2001.
- [6] D. Parkhurst, K. Law, and E. Neibur, "Modeling the role of salience in the allocation of overt visualattention," *Vision Research*, vol. 42, pp. 107-123, 2002.