Eye tracking — a new interface for visual exploration

O K Oyekoya and F W M Stentiford

Eye-tracking technology offers a natural and immediate way of communicating human intentions to a computer. Eye movements reflect interests and may be analysed to drive computer functionality in games, image and video search, and other visual tasks. This paper examines current eye tracking technologies and their applications. Experiments are described that show that target images can be identified more rapidly by eye tracking than using a mouse interface. Further results show that an eye-tracking technology provides an efficient interface for locating images in a large database. Finally the paper speculates about how the technology may enter the mass market as costs decrease.

1. Introduction

Widespread interest in discovering information has created a demand for tools that capture users' intentions. The popularity of search engines (such as Google and Yahoo) has highlighted users' requirements for rapid and effortless access to relevant information. But there is now increasing research activity in the categorisation and retrieval of visual multimedia content for sharing and entertainment purposes as opposed to text-based mechanisms targeted at improving access to written material.

Whereas key words form a convenient feature for characterising documents, there is no such obvious attribute present in images and video material. In addition there is no agreement on what might constitute a universal syntax for images that could capture the meaning that we all see in images. In fact every user possesses a different subjective perception of the world and it is not therefore possible to capture this in a single fixed set of features and associated representations. In this way it is not possible to guarantee to anticipate a user's perception of the visual content and indeed users may change their minds in the middle of a retrieval operation.

The mouse and the keyboard dominate the types of interfaces found in computers today. Most people are happy to use them to interact with their machines, but they present mental and physical barriers to communication. The keyboard requires knowledge of a language of interaction and a chain of events involving vision, thought and muscular movement, all of which requires a judgement as to whether initiating the effort will be worthwhile or not. The mouse reduces keyboard interaction and enables simple visual selection, but still requires the same physical and conscious mental processes to take place. Eye tracking offers a valuable short cut in computer communications for visual tasks [1]. Gaze behaviour could provide information to the machine without the essential need for extra co-ordinated muscular movement and the associated effort. Indeed the reduced level of effort should allow users to convey more relevant factors more easily to the machine and in a shorter time. In addition, there is scope for identifying users' intentions from pre-attentive activity of which the user is not consciously aware and promises to yield extremely rapid search performances.

The following sections present a background to this work. This is followed by descriptions of our system and recent experiments in image retrieval through an eye-tracking interface. The final sections discuss and present some conclusions and an indication of how eye-tracking technology might enter the mass market.

2. Eye-tracking technologies

The first technologies up to the 1960s were invasive and required tampering directly with the eyes. The search coil method [2] offers high accuracy and large dynamic range but requires an insertion into the eye! Non-invasive methods, such as the Dual Purkinje Image eye tracker [3], require the head to be restricted and are relatively expensive. More recently systems have appeared that use video images with some using infra-red cameras. The eye has several key characteristics that makes gaze direction measurable from a video camera image. Eye pointing is precise because there is a centralised region in the retina where there is increasing image resolution towards its centre. LC Technology's Eyegaze system [4] uses the Pupil-Centre/Corneal-Reflection method to determine the eye's gaze direction. A video camera located below the computer screen remotely and unobtrusively observes the subject's eye. No attachments to the head are required in this set-up. A small, low-power, infra-red light emitting diode (LED) located at the centre of the camera lens illuminates the eye. The LED generates the corneal reflection and causes the bright pupil effect, which enhances the camera's image of the pupil. The accuracy of eye-tracking systems depends in large measure on how precisely the image processing algorithms can locate the relative positions of pupil centre and the corneal reflection. To achieve the bright-eye effect, light is shone into the eye along the axis of the camera lens. The eye's lens focuses the light that enters the pupil on to a point on the retina. Because the typical retina is highly reflective, a significant portion of that light emerges back through the pupil, and the eye's lens serendipitously directs that light back along the camera axis right into the camera. Thus the pupil appears to the camera as a bright disk, which contrasts very clearly with the surrounding iris. Specialised image-processing software in the Eyegaze computer identifies and locates the centres of both the pupil and corneal reflection. Trigonometric calculations project the person's gazepoint based on the positions of the pupil centre and the corneal reflection within the video image. Other systems (such as ASL [5], Smarteye [6], IBM's Almaden [7], Arrington's Viewpoint [8], SR's Eyelink [9] and CRS [10] eye trackers) have variations in the design of their respective algorithms for calculating gaze positions, with little or no difference in the basic infra-red technology. Some manufacturers have a head-mounted, as well as a remote, version of their eye tracker and prices have gone down considerably in the last few years.

Several methods have been proposed for improving the accuracy of estimating gaze direction and inferring intent from eye movement. Identification and analysis of fixations [11] and saccades in eye-tracking protocols has been shown to be important for understanding visual behaviour. Privitera et al [12] used ten image processing algorithms to compare human-identified regions of interest (ROIs) with regions of interest determined by an eye tracker and defined by a fixation algorithm. The comparative approach used a similarity measurement to compare two aROIs (algorithmically detected ROIs), two hROIs (human-identified ROIs) and an aROI plus hROI. The prediction accuracy was compared to identify the best-matching algorithms and different algorithms fared better under differing conditions. They concluded that aROIs cannot always be expected to be similar to hROIs in the same image because two hROIs produce different results in separate runs. This means that algorithms are unable in general to predict the sequential ordering of fixation points. Jaimes et al [13] compared eye movement across categories and linked category-specific eye-tracking results to automatic image classification techniques. They hypothesised that the eye movements of human observers differed for images in different semantic categories, and that this information could be effectively used in automatic content-based classifiers. The eye-tracking results suggested that similar viewing patterns occur when different subjects view different images in the same semantic category. Hence, while algorithms are unable to predict the sequential ordering of points of interest, similarity in viewing patterns over images in the same category is possible.

3. Applications

Eye-tracking equipment is used as an interface device in several diverse applications. The number of applications of eye tracking is increasing, as presented in Duchowski's review [14] of diagnostic and interactive applications based on off-line and real-time analysis respectively. Interactive applications have concentrated upon replacing and extending existing computer interface mechanisms rather than creating a new form of interaction. The tracking of eye movements has been employed as a pointer and a replacement for a mouse [15], to vary the screen scrolling speed [16] and to assist disabled users [17]. Schnell and Wu [18] applied eye tracking as an alternative method for the activation of controls and functions in aircraft. Dasher [19] used a method for text entry that relies purely on gaze direction. Nikolov et al proposed [20] a system for construction of gaze-contingent multi-modality displays of multi-layered geographical maps. Gaze-contingent multi-resolutional displays (GCMRDs) centre high-resolution information on the user's gaze position, matching the user's interest. In this system different map information is channelled both to the central and to the peripheral visual fields, giving real performance advantage.

In its diagnostic capabilities eye-tracking provides a comprehensive approach to studying interaction processes such as the placement of menus within Web sites and to influence design guidelines more widely [21]. However, the imprecise nature of saccades and fixation points has prevented these approaches from yielding benefits over conventional human interfaces. Fixations and saccades are used to analyse eye movements, but it is evident that the statistical approaches to interpretation (such as clustering, summation and differentiation) are insufficient for identifying interests due to the differences in human perception of image content.

Although eye tracking has not yet been implemented on mobile devices, research is under way on how the detection of ROIs that catch the eye can be used to improve the quality of images presented on small screens. In the future, eye trackers could automate this process for individual users. Xin Fan et al [22] proposed an image-viewing technique based on an adaptive attention-shifting model, which enabled the browsing of large images on limited and heterogeneous screen zones of mobile phones. Xin fan's paper focused on facilitating image viewing on devices with limited display sizes.

Nokia [23] conducted a usability evaluation on two mobile internet sites and identified a demand for search on mobile phones contrary to the initial hypothesis that users would be discouraged by the effort of keying inputs. The research also showed that customers preferred any interface that produced a successful search despite any extra effort required. The Collage Machine [24] is an agent of Web recombination. It deconstructs Web sites and re-presents them in collage form. It can be taught to bring media of interest to the user on the basis of the user's interactions. The evolving model provides an extremely flexible way of presenting relevant visual information to the user on a variety of devices.

Eye-tracking experiments have been conducted to investigate the informativeness of images and the speed of eye-tracking interfaces. Arising from this work, an eye-tracking interface has been developed which rapidly converges to target images. This work is described and discussed in the next sections.

4. System overview

The best interfaces are natural and easy to use. They are unobtrusive and provide relevant information quickly and in ways that do not interfere with the task itself. This system has been designed to provide an interface for searching visual digital data in an image database (see Fig 1). A pre-computed network of similarities between image regions in an image collection is traversed using eye tracking, always assuming that the users' gaze behaviours yield suitable information about their intentions. It is reasonable to believe that users will look at the objects in which they are interested during a search, and this provides the machine with the necessary information to retrieve plausible candidate target images for the user. Retrieved images will contain regions that possess similarity links with the previously gazed regions, and can be presented to the user in a variety of ways.

4.1 Eye-tracking equipment

The Eyegaze system [4] was used in the experiments to generate raw gazepoint location data at the camera field rate of 50 Hz (units of 20 ms). A clamp with chin rest provided support for chin and forehead in order to minimise the effects of head movements, although the eye tracker does accommodate head movement of up to 1.5 inches (3.8 cm). It was not essential to use the chin rest, but this removed a potential source of error and eliminated any variance in head movement across subjects. The system set-up is shown in Fig 2. Calibration is needed to measure the properties of each subject's eye before the start of the experiments. The images were displayed on a 15-inch LCD flat panel monitor at a resolution of 1024 × 768 pixels.

In the second experiment, the loading of 25 images in the 5 × 5 grid display took an average of 110 ms on a Pentium IV 2.4 GHz PC with 512 Mbit/s of RAM. In the

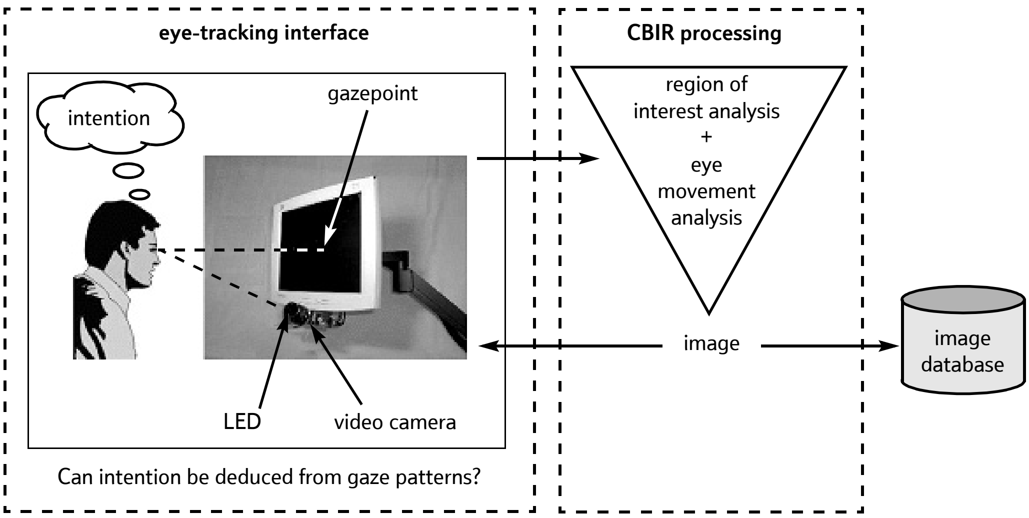

Fig 1 Proposed system architecture.

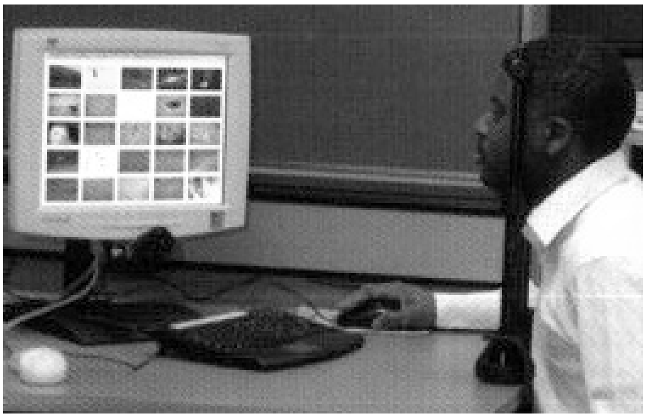

Fig 2 Eyegaze set-up.

third experiment, the loading of 16 images in the 4×4 grid display took an average of 100 ms on the same system. Gaze-data collection and measurement of variables were suspended while the system loaded the next set of images into memory. During this period the display remained unchanged and was updated instantaneously as soon as the contents of the next display had been composed.

The processing of information from the eye tracker is carried out on a 128 Mbit/s Intel Pentium III system with a video frame grabber board.

4.2 *Visual attention and similarity*

It has been shown that attention mechanisms can be directly related to similarity measures [25] and affect the strength of those measures. During a search the human eye is attracted to salient regions and those regions probably have most impact and contribute most towards recognition and user search strategies. This work makes use of both aspects:

- firstly, an attention model [25] is used to automatically identify candidate regions of interest for validation against eye-tracking data where we would expect most fixations to occur,
- secondly, an attention-based similarity metric is used to define visual relationships in a database of images for exploration with an eye-tracking interface.

The visual attention (VA) model used in this work employs an algorithm that assigns high attention scores to pixels where neighbouring pixel configurations do not match identical positional arrangements in other randomly selected neighbourhoods in the image. This means, for example, that high scores will be associated with anomalous objects, or edges and boundaries, providing those features do not predominate in the image. For display purposes the attention scores for each pixel are displayed as a map using a continuous spectrum of false colours with the scores being marked with a distinctive colour or grey level as in Figs 3 and 4.

The similarity measure [25] used in this work is not dependent upon intuitively selected features, but instead upon the notion that the similarity of two patterns is determined by the number of features in common. This means that the measure can make use of a virtually unlimited universe of features rather than a tiny manually selected subset that will be unable to characterise many unseen classes of images. Moreover, the features are deliberately selected from image regions that are salient according to the model and, if validated, reflect similarity as judged by a human.

4.3 *Experimental strategy*

A series of experiments was devised to establish the feasibility of an eye-gaze-driven search mechanism. The first experiment investigated whether users looked more frequently at salient regions as determined by the attention model and whether any other eye behaviour was apparent. A negative result would indicate a potential lack of information in gaze data relevant to image retrieval.

The second experiment investigated the effectiveness of an interface controlled by gaze behaviour when compared with other interfaces. In this experiment the speed of operation was compared with that of a mouse interface. Again a negative result would cast doubt on the benefits of using eye movement in such an interface.

Finally the proposed system was implemented with the aim of investigating whether eye tracking can be used to reach target images in fewer steps than by chance. The effect of the intrinsic difficulty of finding specific images and the time allowed for the consideration of successive selections were also investigated

5. Gaze behaviour

Participants were presented with a sequence of six images for 5 sec each, separated by displays of a blank screen, followed for 3 sec by a central black dot on a white background. Three of the images contained easily discernible subjects and three did not. All participants were encouraged to minimise head movement and were asked to focus on the dot before each image was displayed. The participants were not given any specific task apart from being asked simply to look at the images. All participants had normal or corrected-to-normal vision and had no knowledge of the purpose of the study. Participants included a mix of graduates and administrative staff.

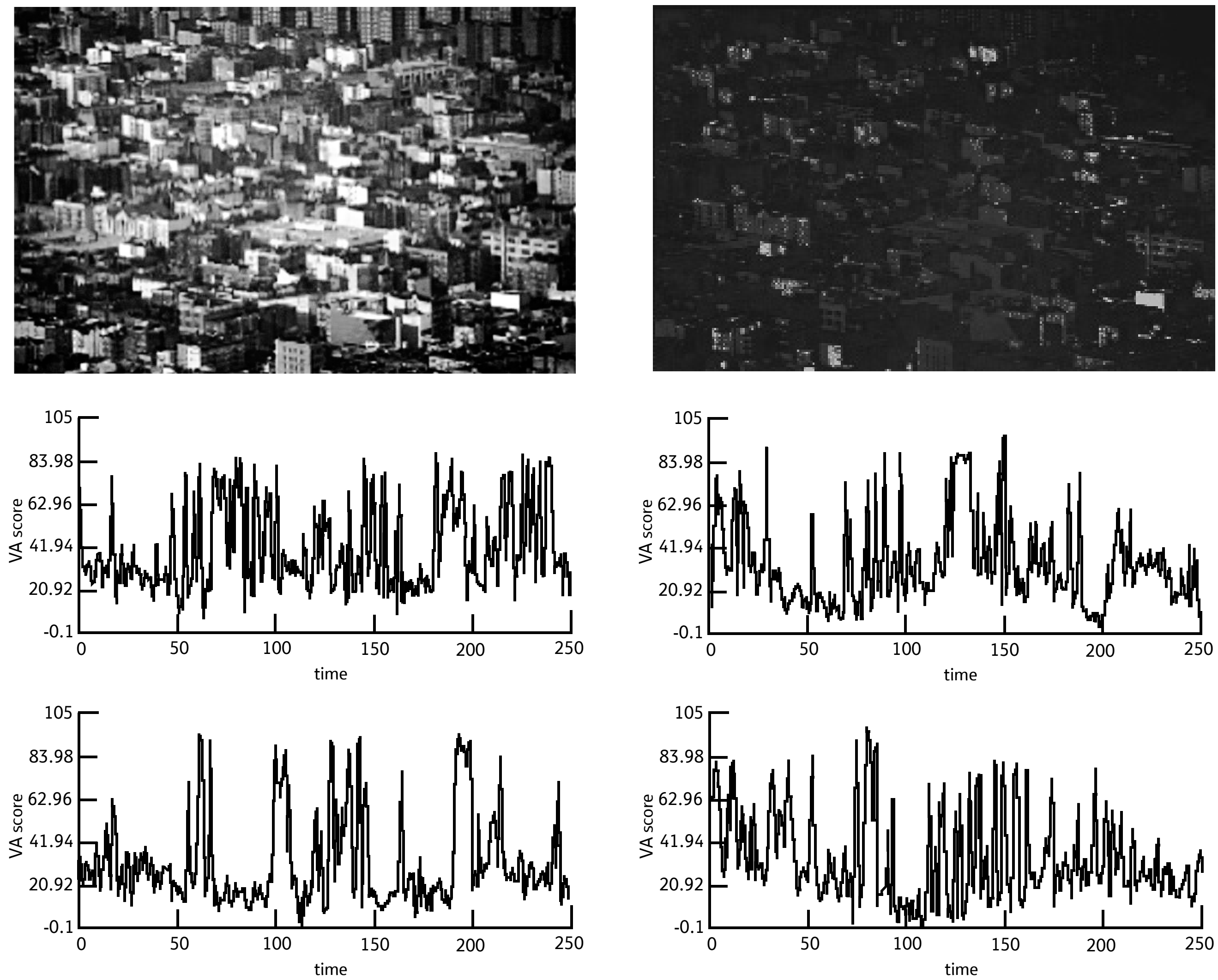

Fig 3 No obvious subject image, VA map and plots.

The locations of saccades and fixations performed by the subjects on each of the images were recorded by the eye-tracking system. The VA score that corresponded to the pixel at each fixation point was associated with the time of the fixation and plotted for study in units of 20 ms as illustrated in Figs 3 and 4. It can be seen that there was considerable variation in behaviour over the four participants, but all looked at regions with the highest VA scores early in the display period. Table 1 shows the total length of time, in ms, spent fixating on regions of high VA score for each participant on each image. This shows that in all cases a large proportion of the 5 sec exposure time was spent observing the salient regions rather than the background, if such a salient region was present in the image. Images without obvious subjects did not give such a pronounced result. This not only confirmed that the gaze of the users was attracted by regions of high VA score, but it also showed that the eye-tracking system was able to gather data related to users' interests and therefore that this information might be available for image retrieval through a suitable interface.

Table 1 Times (ms) spent fixating on regions of high VA score.

Images		Subjects			
		1	2	3	4
Unclear ROI	1	40	60	20	140
	2 (Fig 3)	580	420	500	400
	3	100	0	40	20
Obvious ROI	4	2820	2340	2420	1280
	5 (Fig 4)	3680	1480	2220	1960
	6	4240	980	1620	1240

6. Relative speeds of eye and mouse

A task-oriented experiment was conducted to compare the speed of the eye and the mouse as an input mode to control an interface. Participants were asked to find a target image in a series of displays with the aim of comparing the response times of searching and selecting the target image using the computer mouse and the eye under varying conditions.

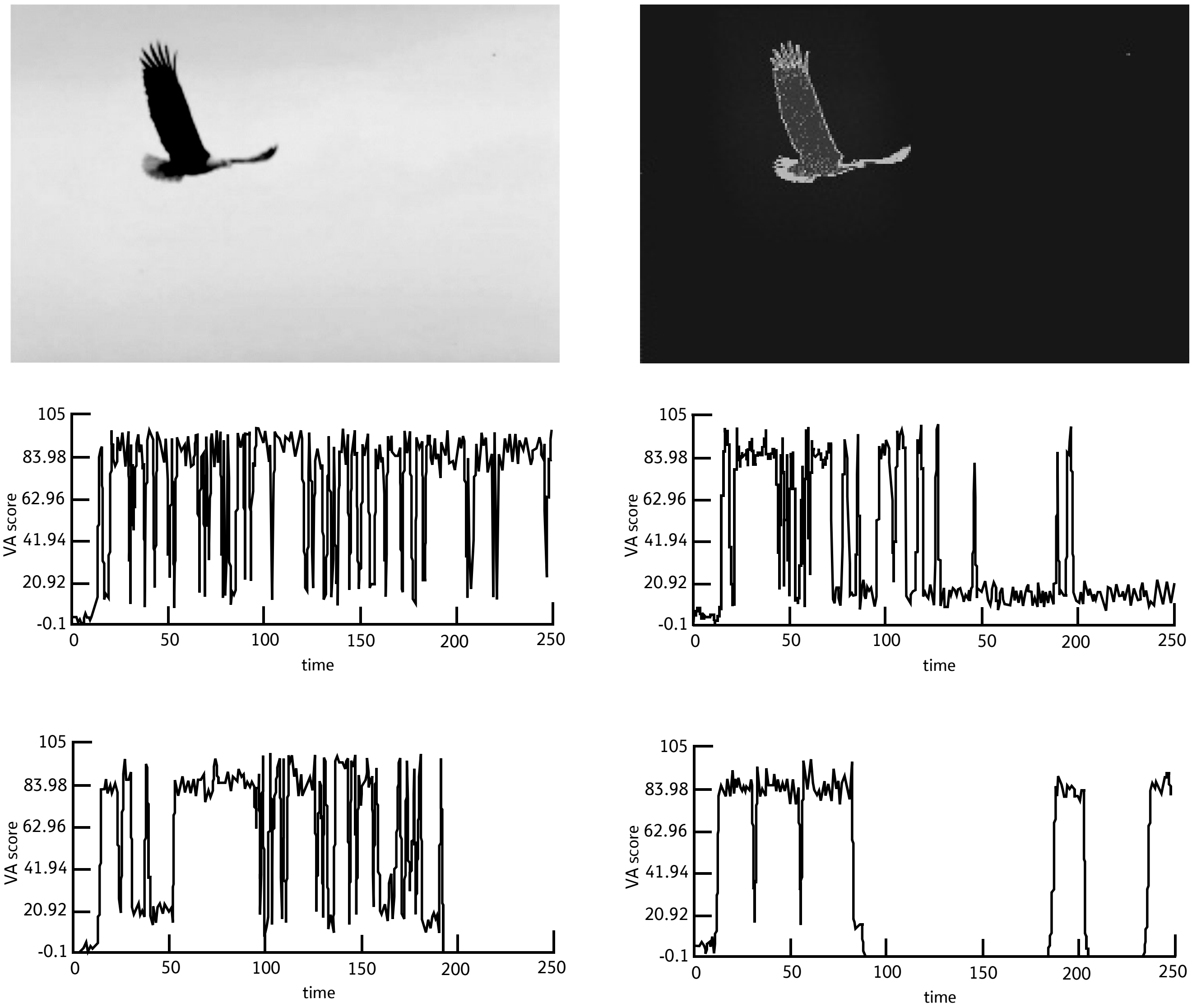

Fig 4 Obvious subject image, VA map and plots.

A total of 12 participants took part in the experiment. Participants included a mix of students and university staff. All participants had normal or corrected-to-normal vision and provided no evidence of colour blindness. Participants were asked to locate a target image in each of a series of 50 5 × 5 arrays of 25 thumbnail images. After finding the target the participants made a selection by clicking with the mouse or fixating on it for longer than 40 ms with the eye. The array was then re-displayed with the positions of the images rearranged with the target image appearing two times in every location during the 50 displays. Participants were randomly divided into two groups; the first group used the eye-tracking interface first and then the mouse, and the second group used the interfaces in the reverse order. This enabled any variance arising from the ordering of the input modes to be identified. Different sequences of the 50 target positions were also employed to remove any confounding effects arising from the ordering of the individual image search tasks. All participants experienced the same sequence of target positions as well as different sequences while using the two input modes. A typical participant in the mouse first group performed four runs: mouse (target sequence 1), eye (target sequence 1), mouse (target sequence 2) and eye (target sequence 3). There was a 1 min rest between runs.

There was a significant main effect of input, $F(1,10) = 8.72$, $p = 0.015$ with faster mean response times when the eye was used as an input (2.08 sec) than when the mouse was used (2.43 sec), as shown in Table 2. The main effect of the order was not significant with $F(1,10) = 0.43$, $p = 0.53$. The main effect of target positions was not significant, $F(1,10) = 0.58$, $p = 0.47$.

7. Image retrieval

This experiment was designed to explore the performance of an image retrieval interface driven by an eye tracker. Thirteen participants were asked to find

Table 2 Mean response times for target image identification task.

Order	Target positions	Input mode	Response time (sec)	
			Mean	Standard deviation
Mouse first (6 participants)	Same-sequence	Mouse	2.33	0.51
		Eye	1.79	0.35
	Different-sequence	Mouse	2.43	0.38
		Eye	1.96	0.42
Eye first (6 participants	Same-sequence	Mouse	2.35	0.82
		Eye	2.29	0.74
	Different-sequence	Mouse	2.59	1.44
		Eye	2.27	0.73

target images in a database and their performance measured.

1000 images were selected from the Corel image library. Images of 127 kilobytes and 256 × 170 pixel sizes were loaded into the database. The categories included boats, landscapes, vehicles, aircrafts, birds, animals, buildings, athletes, people and flowers. Four easy-to-find and four hard-to-find target images were selected for the experiment by using a random gaze strategy to explore the image database. Screens of thumbnail images were displayed as 229 × 155 pixels in 4 × 4 arrays. The initial screen is shown on the left of Fig 5 where the target image that the participant has to find is located at the top left with a dark (red) border.

Participants began by viewing the initial screen and endeavouring to find the target image among the other 15 images. The display automatically changed when the accumulation of all fixations greater than 80 ms on a specific image position exceeded a threshold. In this way the display would change relatively quickly if the participant concentrated on a relevant image, but would take longer if the gaze was less definite. This selected image determined the next 15 thumbnails to be displayed as indicated by the highest of the pre-computed similarity scores for other images in the database. The participant was presented with a succession of such screens until the target image was retrieved whereupon the run halted and the successfully found target was highlighted with a red border (shown as a dark border on the right of Fig 5). Each participant performed eight runs using both easy-to-find and hard-to-find images. The maximum number of screen changes was limited to 26.

Two fixation cumulative thresholds of 400 ms and 800 ms were employed as a factor in the experiment. Another factor was introduced to allow the display to include either one or no randomly retrieved images. It was thought that this would reduce the likelihood of displays repeating due to occasional incorrect similarity values. In this case one of the images was retrieved randomly from the database ('Randomly retrieved' = 1) rather than on the basis of similarity with the previously selected image. A random gaze-generation strategy in which images in each screen are selected randomly was then simulated for comparison with selection by gaze.

As measures of performance the number of steps to target, the time to target (F_1), and the number of fixations (F_2) of 80 ms and above were monitored and recorded during the experiment (Table 3). The results of

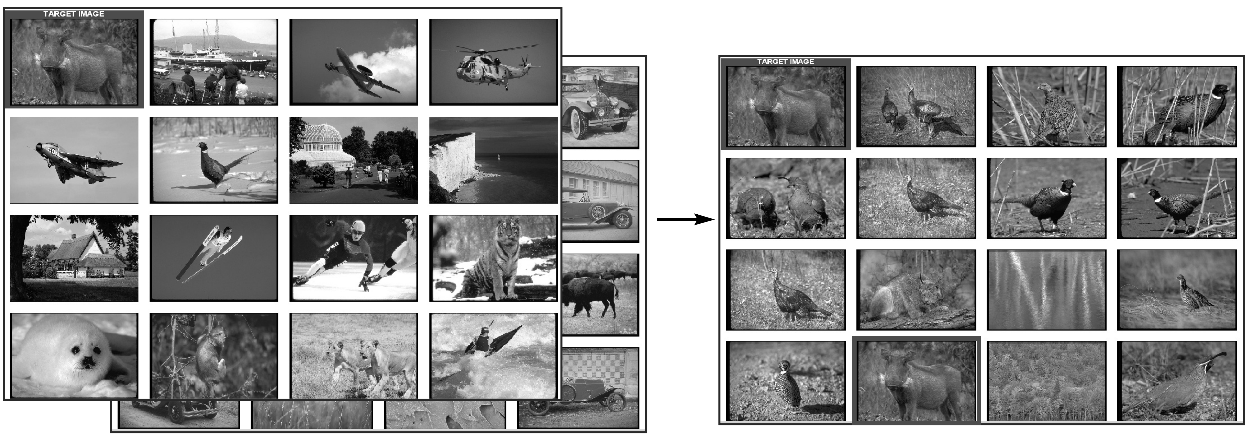

Fig 5 Initial screen leading to final screen with retrieved target.

Table 3 Analysis of human eye behaviour.

Image type	Fixation threshold	Randomly retrieved	Target not found (frequency)	Steps to target	Time to target (sec)	Fixation numbers
Easy-to-find	400 ms	0	38.5%	14	34.944	99
		1	53.8%	18	36.766	109
	800 ms	0	38.5%	14	55.810	153
		1	15.4%	11	51.251	140
Hard-to-find	400 ms	0	69.2%	23	52.686	166
		1	84.6%	23	50.029	167
	800 ms	0	92.3%	24	104.999	327
		1	69.2%	19	83.535	258

the analysis of variance (ANOVA) performed on the steps to target revealed a significant main effect of image type, $F(1,12) = 23.90$, $p < 0.0004$ with fewer steps to target for easy-to-find images (14 steps) than the hard-to-find images (22 steps). The main effect of the fixation threshold was not significant with $F(1,12) = 1.50$, $p < 0.25$. The main effect of randomly-retrieved was also not significant, $F(1,12) = 0.17$, $p < 0.69$. The analysis of the time to target produced similar results to the analysis of the number of fixations. There was a significant main effect of image type, $F_1(1,12) = 24.11$, $p < 0.0004$, $F_2(1,12) = 21.93$, $p < 0.0005$, with shorter time to target and fewer fixations for easy-to-find images (40.5 sec and 125 fixations) than the hard-to-find images (71.3 sec and 229 fixations). The main effect of the fixation threshold was also similarly significant with $F_1(1,12) = 18.27$, $p < 0.001$ and $F_2(1,12) = 16.09$, $p < 0.002$. The main effect of randomly retrieved was not significant, $F_1(1,12) = 1.49$, $p < 0.25$ and $F_2(1,12) = 0.76$, $p < 0.40$.

The same treatment combinations experienced by all participants were applied to the random-gaze generation tool to obtain steps to target under the same conditions (Table 4). In summary, the results of the ANOVA revealed a main effect of the selection mode, $F(2,23) = 3.81$, $p < 0.037$, with fewer steps to target when the eye gaze is used (18 steps) than when random selection is used (22 steps). There was also a main effect of image type, $F(2,23) = 28.95$, $p < 0.00001$ with fewer steps to target for easy-to-find images (16 steps) than the hard-to-find images (24 steps). Further analysis of the simple main effect revealed that there was a significant difference between the modes for the hard-to-find images, $F(2,23) = 3.76$, $p < 0.039$, as opposed to the easy-to-find images, $F(2,23) = 2.02$, $p < 0.16$.

8. Discussion

The first experiment tested whether users looked at regions declared salient by the visual attention model. The results showed that this was the case for the images and participants involved, but more images and a larger number of participants would be necessary to obtain statistical significance. This result also indicated that users fixate on foreground material in images and that this behaviour may be employed to drive a prototype search interface.

The second experiment went further to explore the speeds of visual processing involved in an image target identification task when compared with a conventional input device such as a mouse. The 25 stimuli presented to each participant, and the predetermined choice of image target, produced a difficult task, and the experiment imposed a high cognitive load. The participant had to search for the target and then make a selection. Our results indicated slower mouse responses and was supported by the significant main effect of input ($p = 0.015$), with the eye interface having faster response times than the mouse interfaces, and was

Table 4 Comparison of eye and random-gaze generation.

Selection mode	Image type	Randomly retrieved	Target not found (frequency)	Steps to target
Eye gaze	Easy-to-find	0	38.5%	14
		1	34.6%	15
	Hard-to-find	0	80.8%	23
		1	76.9%	21
Random gaze	Easy-to-find	0	57.7%	20
		1	38.5%	16
	Hard-to-find	0	96.2%	25
		1	92.3%	26

consistent with Ware and Mikaelian's conclusions [26]. When using the mouse the participant had to spend time locating both the cursor and the item to be selected, and then use the mouse to move the cursor to the item. On the other hand the eye tracker interface was quicker because only the selected item needed to be located. However, the speed difference was not just dependent on extra mouse movement because the eye tracker required the user to fixate on the target for longer than 40 ms before a screen change.

Finally in the image retrieval experiment the participants using the eye-tracking interface found the target in fewer steps than an automated random-gaze strategy ($p < 0.037$), and the analysis of the simple effect attributed the significant difference to the hard-to-find images. This meant that the probability of finding the hard-to-find images was significantly increased due to human cognitive abilities as opposed to the indiscriminate selection by the simulated random-gaze strategy using the same similarity information. The main effect of the fixation threshold was not significant which indicates that there is scope for using smaller thresholds than 400 ms. Future experiments, if successful, would indicate that unconscious pre-attentive vision may be playing a significant part in visual search. Additional discussion and results can be found in Oyekoya and Stentiford [27—30].

9. Conclusions and future directions

An eye-controlled interface can provide a more natural mode of retrieval as it requires a minimum of manual effort and cognitive load, and almost unconscious operation. It has been shown that the eye is attracted to image regions that are predicted to be salient by the attention model and that the eye-tracking system was able to gather data related to users' interests. Secondly, the eye-tracking interface yielded a significantly better speed performance than the mouse in a target location task. In an image-retrieval task users were able to successfully navigate their way to target images in a database using only eye gaze, with significantly better performance than randomly generated selections.

There is much research to be carried out before eye trackers can become as pervasive as keyboards and mice. The accuracy, cost and usability of equipment must improve before laboratory results can be reproduced on PCs, laptops, and even PDAs. We might expect cheap eye trackers to emerge in the games market where 'look and shoot' would give faster gratification than painful button pressing or joystick pushing. Small cameras embedded in monitors and laptop lids or glasses would be obvious locations for such devices. Gaze-contingent displays have great potential where additional information may be displayed dependent on eye movement. For example, larger scale maps may be offered at the focus of attention or additional details supplied related to an object being studied. Eye behaviour may also be used to drive PTZ cameras in ways that enable people to 'see' their way around remote locations. Eye trackers are already a great asset to the disabled, but only as an awkward and costly replacement for existing devices, and not as a computer interface to be used just as effectively as an able-bodied person. The results reported here indicate that eye trackers have the potential for eliciting human intentions extremely rapidly and may be applied to certain visual search tasks. It seems reasonable that reducing costs and advancing camera technology will mean that eye trackers will appear in many more applications within the next few years.

Acknowledgements

The authors will like to thank the EPSRC, British Telecommunications plc, SIRA and the Imaging Faraday Partnership for their support in this work. This research has been also conducted within the framework of the European Commission funded Network of Excellence 'Multimedia Understanding through Semantics, Computation and Learning' (MUSCLE) [31].

References

1 Jacob R J K: 'Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces', Advances in Human-Computer Interaction, 4, pp 150—190 (1993).

2 Robinson D A: 'A method of measuring eye movement using a scleral search coil in a magnetic field', IEEE Transactions on Biomedical Electronics, BME-10, pp 137—145 (1963).

3 Crane H D and Steele C S: 'Accurate three-dimensional eye tracker', Applied Optics, 17, pp 691—705 (1978).

4 http://www.eyegaze.com/

5 http://www.a-s-l.com/

6 http://www.smarteye.se/

7 http://www.almaden.ibm.com/cs/blueeyes/

8 http://www.arringtonresearch.com/

9 http://www.eyelinkinfo.com/

10 http://www.crsltd.com/

11 Salvucci D D and Goldberg J H: 'Identifying fixations and saccades in eye-tracking protocols', in Proceedings of the Eye Tracking Research and Applications Symposium, pp 71—78, New York, ACM Press (2000).

12 Privitera C M and Stark L W: 'Algorithms for Defining Visual Regions of Interest: Comparison with Eye Fixations', IEEE Transactions On Pattern Analysis and Machine Intelligence, 22, No 9, pp 970—982 (2000).

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

14 Duchowski A T: 'A Breadth-First Survey of Eye Tracking Applications', Behaviour Research Methods, Instruments and Computers (BRMIC), 34, No 4, pp 455—470 (2002).

15 Hansen J P, Anderson A W and Roed P: 'Eye gaze control of multimedia systems', in Anzai Y, Ogawa K and Mori H (Eds): 'Symbiosis of Human and Artifact', Vol 20A, Elsevier Science, pp 37—42 (1995).

16 Numajiri T, Nakamura A and Kuno Y: 'Speed browser controlled by eye movements', IEEE Int Conf on Multimedia and Expo (2002).

17 Corno F, Farinetti L and Signorile I: 'A cost effective solution for eye-gaze assistive technology', IEEE Int Conf on Multimedia and Expo, Lausanne (August 2002).

18 Schnell T and Wu T: 'Applying Eye Tracking As Alternative Approach for Activation of Controls and Functions in Aircraft', Proceedings of the 5th International Conference on Human Interaction with Complex Systems (HICS), Urbana, Illinois, USA, p 113 (2000).

19 Ward D J and MacKay D J C: 'Fast hands-free writing by gaze direction', Nature, 418, p 838 (2002).

20 Nikolov S G, Bull D R and Glichrist I D: 'Gaze-Contingent Multi-Modality Displays of Multi-Layered Geographical Maps', Proc of the 5th Intl Conf on Numerical Methods and Applications (NM&A02), Symposium on Numerical Methods for Sensor Data Processing, Borovetz, Bulgaria (2002).

21 McCarthy J, Sasse M A and Riegelsberger J: 'Could I have the menu please? An eye tracking study of design conventions', Proceedings of HCI2003, Bath, UK (September 2003).

22 Xin F, Xing X, Wei-Ying M, Hong-Jiang Z and He-Qin Z: 'Visual Attention Based Image Browsing on Mobile Devices', IEEE International Conference on Multimedia and Expo, I, pp 53—56, Baltimore, MD, USA (2003).

23 Roto V: 'Search on Mobile Phones', Best Practices and Future Visions for Search User Interfaces: A Workshop, CHI 2003 (2003).

24 Kerne A: 'Collage Machine: an interactive agent of Web recombination', Leonardo, 33, No 5, pp 347—350 (2000).

25 Stentiford F W M: 'Attention Based Similarity', Pattern Recognition, Elsevier (to appear) (2006).

26 Ware C and Mikaelian H: 'An evaluation of an Eye Tracker as a Device for Computer Input', Proceeding of SIGCHI+GI '87, Human Factors in Computing Systems (also published as a special issue of the SIGCHI Bulletin) (April 1987).

27 Oyekoya O K and Stentiford F W M.: 'Exploring Human Eye Behaviour Using a Model of Visual Attention', International Conference on Pattern Recognition, Cambridge UK (August 2004).

28 Oyekoya O K and Stentiford F W M: 'A Performance Comparison of Eye Tracking and Mouse Interfaces in a Target Image Identification Task', 2nd European Workshop on the Integration of Knowledge, Semantics & Digital Media Technology (EWIMT), London (December 2005).

29 Oyekoya O K and Stentiford F W M: 'An Eye Tracking Interface for Image Search', Eye Tracking Research & Applications Symposium, San Diego, CA (March 2006).

30 Oyekoya O K and Stentiford F W M: 'Perceptual Image Retrieval Using Eye Movements', International Workshop on Intelligent Computing in Pattern Analysis/Synthesis, Xi'an, China (August 2006).

31 'Multimedia Understanding through Semantics, Computation and Learning', EC 6th Framework Programme, FP6-507752 (2005) — http://www.muscle-noe.org/

[redacted]

Oyewole Oyekoya was awarded a scholarship and graduated in June 2001 with upper second class degree in BSc (Hons) Computer Science at Kingston University. He undertook his work placement at Nortel Networks, gaining valuable experience in the telecommunications industry and won company sponsorship to produce an interface for test automation as his final dissertation. He was awarded a CASE studentship with BT and University College London and is now studying for a PhD while based at the UCL Adastral Park Campus His studentship is sponsored at UCL by the Perception and Coding group within the BT Broadband Applications Research Centre.

He was a Faraday Associate sponsored by the Imaging Faraday Partnership and BT. He is now researching eye tracking for image search and retrieval. He was awarded a scholarship from the centre for Scientific Enterprise limited to undertake the commercial evaluation of his PhD research.

[redacted]

Professor Fred Stentiford won a scholarship to study mathematics at St Catharine's College, Cambridge, and obtained a PhD in Pattern Recognition at Southampton University. He first joined the Plessey Company to work on various applications including the recognition of fingerprints and patterns in time varying magnetic fields. He then joined BT and carried out research on optical character recognition and speech recognition. During this time he led a team developing systems employing pattern recognition methods for the machine translation of text and speech. This work led to the world's first demonstration of automatic translation of speech between different languages. He is now a professor of telecommunications based at the UCL Campus at Adastral Park where he leads a team researching multimedia content understanding. He has published over 70 papers and filed more than 20 patents on pattern recognition techniques and is a member of the IET and BCS.