

# An Eye Tracking Interface for Image Search

Oyewole Oyekoya<sup>†</sup> Fred Stentiford<sup>‡</sup>

University College London, Adastral Park, Ipswich United Kingdom IP5 3RE

## Abstract

This paper explores the feasibility of using an eye tracker as an image retrieval interface. A database of image similarity values between 1000 Corel images is used in the study. Results from participants performing image search tasks show that eye tracking data can be used to reach target images quicker than by random selection. The effects of the intrinsic difficulty of finding images and the time allowed for successive selections were also investigated.

**CR Categories:** H.5.2 [Information Interfaces and Presentation]: User Interfaces; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval; H.1.2 [Models and Principles]: User/Machine Systems – Human Information Processing.

**Keywords:** visual search, image retrieval, eye tracking interface..

## 1 Introduction

Images play an increasingly important part in the lives of many people. There is a critical need for automated management, as the flow of digital visual data increases and is transmitted over the network. Retrieval mechanisms must be capable of handling the amount of data efficiently and quickly. Existing systems are capable of retrieving archiving material according to date, time, location, format, file size, etc. However, the ability to retrieve images with semantically similar content from a database is more difficult.

One of the major issues in information searching is the problem associated with initiating a query. Indeed lack of high-quality interfaces for query formulation has been a longstanding barrier to effective image retrieval systems [Venters et al. 1997]. Users find it hard to generate a good query because of initial vague information [Urban et al. 2003] (i.e. “I don’t know what I am looking for but I’ll know when I find it”). Eye tracking presents an adaptive approach that can capture the user’s current needs and tailor the retrieval accordingly. Understanding the movement of the eye over images is an essential component in the research.

---

{<sup>†</sup> o.oyekoya | <sup>‡</sup> f.stentiford}@adastral.ucl.ac.uk

Research in the applications of eye tracking is increasing, as presented in Duchowski’s review [2002] of diagnostic and interactive applications based on offline 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 [Hansen et al. 1995], to vary the screen scrolling speed [Numajiri et al. 2002] and to assist disabled users [Corno et al. 2002]. Dasher [Ward and MacKay 2002] uses a method for text entry that relies purely on gaze direction. 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 [McCarthy et al. 2003]. 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 analyze 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 humans’ perception of image content. More robust methods of interpreting the data are needed. There has been some recent work on document retrieval in which eye tracking data has been used to refine the accuracy of relevance predictions [Puolamäki 2005]. Applying eye tracking to image retrieval requires that new strategies be devised that can use visual and algorithmic data to obtain natural and rapid retrieval of images.

Traditional approaches of image retrieval suffer from two main disadvantages. Firstly there is a real danger that the use of any form of pre-defined feature measurements will be unable to handle unseen material. Image retrieval systems normally rank the relevance between a query image and target images according to a similarity measure based on a set of features. Pre-determined features can take the form of edges, colour, location, texture, and others. Secondly the choice of low-level features is unable to anticipate a user’s high-level perception of image content. This information cannot be obtained by training on typical users because every user possesses a subtly different subjective perception of the world and it is not possible to capture this in a single fixed set of features and associated representations. Thirdly descriptive text does not reflect the capabilities of the human visual memory and does not satisfy users’ expectations. Furthermore the user may change his/her mind and may also be influenced by external factors.

An approach to visual search should be consistent with the known attributes of the human visual system and account should be taken of the perceptual importance of visual material. Recent research in human perception of image content [Itti 2004] suggests the importance of semantic cues for efficient retrieval. Relevance feedback mechanisms [Cox et al. 2000] is often proposed as a technique for overcoming many of the problems faced by fully

## ETRA Symposium, San Diego, March 27-29, 2006

automatic systems by allowing the user to interact with the computer to improve retrieval performance. This reduces the burden on unskilled users to set quantitative pictorial search parameters or to select images (using a mouse) that come closest to meeting their goals. This has prompted research into the viability of eye tracking as a natural input for an image retrieval system. Visual data can be used as input as well as a source of relevance feedback for the interface. Human gaze behaviour may serve as a new source of information that can guide image search and retrieval.

Human eye behaviour is defined by the circumstances in which they arise. The eye is attracted to regions of the scene that convey what is thought at the time to be the most important information for scene interpretation. Initially these regions are pre-attentive in that no recognition takes place, but moments later in the gaze the fixation points depend more upon either our own personal interests and experience or a set task. Humans perceive visual scenes differently. We are presented with visual information when we open our eyes and carry out non-stop interpretation without difficulty. Research in the extraction of information from visual scenes has been explored by Yarbus [1967], Mackworth and Morandi [1967] and Hendersen and Hollingworth [1999]. Mackworth and Morandi [1967] found that fixation density was related to the measure of informativeness for different regions of a picture and that few fixations were made to regions rated as uninformative. The picture was segmented and a separate group of observers were asked to grade the rate of informativeness. Scoring the informativeness of a region provides a good insight into how humans perceive a scene or image. Henderson and Hollingworth [1999] described semantic informativeness as the meaning of an image region and visual informativeness as the structural information. Fixation positions were more influenced by the former compared to the latter. The determination of informativeness and corresponding eye movements are influenced by task demands [Yarbus 1967].

Previous work [Oyekoya and Stentiford 2004] used a visual attention model to score the level of informativeness in images and found that a substantial part of the gaze of the participants during the first two seconds of exposure is directed at informative areas as estimated by the model. Subjects were presented with images with clear regions-of-interest and results showed that these attracted eye gaze on presentation of the images studied. This led credence to the belief that the gaze information obtained from users when presented with a set of images could be useful in driving an image retrieval interface. More recent work [Oyekoya and Stentiford 2005] compared the performance of the eye and the mouse as a source of visual input. Results showed faster target identification for the eye interface than the mouse for identifying a target image on a display.

In this paper, experiments are described that explore the viability of using the eye to drive an image retrieval interface. In a visual search task, users are asked to find a target image in a database and the number of steps to the target image are counted. It is reasonable to believe that users will look at the objects in which they are interested during a search [Oyekoya and Stentiford 2004] and this provides the machine with the necessary information to retrieve a succession of plausible candidate images for the user.

## 2 Methodology

### 2.1 Image Database

1000 images were selected from the Corel image library. Images of 127 kilobytes and 256 x 170 pixel sizes were loaded into the database. The categories included boats, landscapes, vehicles, aircrafts, birds, animals, buildings, athletes, people and flowers. The initial screen (including the position of the target image) is shown in Figure 2. Images were displayed as 229 x 155 pixel sizes in the 4 x 4 grid display.

### 2.2 Similarity Model

Studies in neurobiology and computer vision [Desimone 1998; Itti 2004] are suggesting that human visual attention is enhanced through a process of competing interactions among neurons representing all of the stimuli present in the visual field. The competition results in the selection of a few points of attention and the suppression of irrelevant material. Such a mechanism has been explored [Grigorescu et al 2003] and extended to apply to the comparison of two images in which attention is drawn to those parts that are in common rather than their absence as in the case of saliency detection in a single image [Bamidele et al 2004]. Whereas saliency measures require no memory of data other than the image in question, this similarity measure, termed Cognitive Visual Attention (CVA) makes use of other stored material in order to determine similarity with an unknown image. In [Bamidele and Stentiford 2005], the approach is used to identify clusters within an extremely diverse set of images in the context of the identification of photo locations. The CVA model relies upon the matching of large numbers of pairs of pixel groups taken from patterns A and B under comparison (Figure 3).

The two patterns are compared as follows:

1. Select a small number of pixels  $S_x$  (a fork) in image A. (e.g. 3 pixels shown in Figure 3)
2. Apply this pixel fork to a random position in B ( $S_y$ ), and compare.
3. Increment similarity score for a match.
4. Increment loop count and exit if  $>M$ .
5. Loop to 1.

The CVA similarity score is incremented each time one of the set of pixel sets  $S_x$  matches a set  $S_y$  in pattern B. This means that image pairs A, B which possess large numbers of matching forks will obtain high CVA scores by virtue of the number of such features they possess in common. The CVA algorithm was applied to the 1000 images to pre-compute similarity scores for all pairs of images to obtain a similarity score matrix.

## 3 Experiment Design

### 3.1 Search Task

Images are presented in a 4 by 4 grid with target image presented in the top left corner of the display (Figure 2). The user is asked to search for the target image and on the basis of the gaze behaviour the machine selects the most favourable image. The next set of 15 images are then retrieved from the database and displayed for the next selection. .

### 3.2 Selection of Target Images

Searching on displays produces two main problems. Firstly, if the target is not similar to any of the currently displayed set (e.g. a cluster of visually-different images), then it is very difficult for the user to direct the search away from the displayed set and towards the target [Vinay et al. 2004]. Secondly, the user's selections from successive displays could enter a loop in which the displays lead to each other, and no exit towards the target is possible. In addition, a random selection strategy provided a performance base-line which any more intelligent approach would need to exceed.

An automatic random selection tool, which randomly selects an image from the 15 displayed images, was implemented to investigate the search task. The random selection tool was executed and the number of steps to target recorded with all 1000 images acting as the target image. This was carried out using between 0 and 15 images in the display, retrieved randomly from the database rather than on the basis of the highest similarity scores. Results were consistent, with a typical run displayed in Table 1. As the number of randomly retrieved images was increased, the likelihood of finding the target image in the first 20 displays/steps to target also increased. A histogram plot of the frequency distribution of steps to target for every image in the database revealed the easy-to-find and hard-to-find images and, 4 easy-to-find and 4 hard-to-find target images were picked for the experiment. These are shown in Figure 4.

### 3.3 Criterion for Best Image Selection

The display change is determined by eye selection of an image, using the sum of all fixations of 80ms and above on an image position, up to a fixation threshold. A red rectangle is flashed around an image if a fixation of 80ms is detected. Two fixation thresholds of 400ms and 800ms were employed as a factor in the experiment.

### 3.4 Participants

Thirteen unpaid participants took part in this 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.

### 3.5 Apparatus and Settings

An Eyegaze System [LC Technologies Inc. 2005] was used in the experiments to generate raw gaze-point location data at the camera field rate of 50 Hz (units of 20ms). A clamp with chin rest provided support for chin and forehead in order to minimize the effects of head movements, although the eye tracker does accommodate head movement of up to 1.5 inches (3.8cm). 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" LCD Flat Panel Monitor at a resolution of 1024x768 pixels. The loading of 16 images in the 4 x 4 grid display took an average of 100ms on a Pentium IV 2.4GHz PC with 512MB of RAM. Gaze data collection and measurement of variables were suspended while the system loaded the next display.

The processing of information from the eye tracker is done on a 128MB Intel Pentium III system with a video frame grabber board. The system setup is described in Figure 1.

### 3.6 Experimental Procedure

Four easy to find and four hard to find target images were used (Figure 4). There was one practice run to enable better understanding of the task at hand and to equalise skill levels during the experiment. Participants understood that there will be a continuous change of display until they found the target but did not know what determines the display change. The display included either no or one randomly retrieved image. Participants performed 8 runs, using all image types (easy-to-find and hard-to-find). Four treatment combinations of the two fixation thresholds (400ms and 800ms) and two random-retrieval levels (0 and 1) were applied to each image type. Any sequence effect was minimised by randomly allocating each participant to different sequences of target images. The first four runs were assigned to each image type. There was a 1 minute rest in between runs. The maximum number of steps to target was limited to 26 runs.

| Steps to target | Frequency Distribution              |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----------------|-------------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|                 | Number of randomly-retrieved images |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|                 | 0                                   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  |
| 1-20            | 112                                 | 142 | 153 | 164 | 201 | 189 | 201 | 221 | 220 | 237 | 244 | 257 | 251 | 229 | 270 | 225 |
| 21-40           | 34                                  | 76  | 107 | 111 | 138 | 168 | 166 | 190 | 172 | 159 | 173 | 171 | 190 | 198 | 194 | 215 |
| 41-60           | 21                                  | 75  | 100 | 88  | 99  | 87  | 114 | 119 | 142 | 146 | 129 | 142 | 135 | 134 | 139 | 139 |
| 61-80           | 21                                  | 39  | 64  | 91  | 73  | 79  | 92  | 92  | 92  | 101 | 96  | 112 | 113 | 104 | 96  | 111 |
| 81-100          | 21                                  | 47  | 72  | 61  | 74  | 66  | 71  | 70  | 83  | 79  | 75  | 82  | 77  | 84  | 66  | 77  |
| 101-120         | 14                                  | 43  | 41  | 60  | 62  | 59  | 64  | 50  | 53  | 44  | 58  | 52  | 57  | 69  | 54  | 63  |
| 121-140         | 12                                  | 45  | 33  | 46  | 55  | 67  | 56  | 49  | 49  | 43  | 42  | 50  | 36  | 55  | 37  | 37  |
| 141-160         | 25                                  | 20  | 37  | 37  | 34  | 45  | 30  | 38  | 42  | 39  | 39  | 25  | 29  | 35  | 31  | 32  |
| 161-180         | 5                                   | 30  | 22  | 42  | 41  | 28  | 41  | 34  | 32  | 31  | 33  | 24  | 26  | 21  | 28  | 20  |
| 181-200         | 12                                  | 26  | 29  | 26  | 26  | 22  | 29  | 25  | 15  | 20  | 25  | 18  | 29  | 14  | 15  | 23  |
| Not found       | 723                                 | 457 | 342 | 274 | 197 | 190 | 136 | 112 | 100 | 101 | 86  | 67  | 57  | 57  | 70  | 58  |

Table 1: Results of applying the random selection strategy to the image database (sum of each column = 1000 images)

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

**Table 2: Analysis of Human Eye Behaviour on the Interface (rounded-off mean figures)**

| 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 selection | Easy-to-find | 0                  | 57.7%                        | 20              |
|                  |              | 1                  | 38.5%                        | 16              |
|                  | Hard-to-find | 0                  | 96.2%                        | 25              |
|                  |              | 1                  | 92.3%                        | 26              |

**Table 3: Comparison of Eye and Random Selection (rounded-off mean figures)**

#### 4 Results

Three dependent variables, the number of steps to target, the time to target ( $F_1$ ), and the number of fixations ( $F_2$ ) of 80ms and above were monitored and recorded during the experiment. 24 dependent variables (8 each) were recorded for each participant. The average figures are presented in Table 2.

104 (= 8x24) figures were entered for each dependent variable into repeated measures ANOVA with three factors (image type, fixation threshold and randomly-retrieved).

The results of the 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,  $F(1,12)=0.17$ ,  $p<0.69$ . All two-factor and three-factor interactions were not significant.

Further analysis of the first-order and second-order simple main effects was conducted individually on all levels of the three factors. The image types influenced the steps to target when participants had a set fixation threshold of 400ms,  $F(1,12)=15.41$ ,  $p=0.002$ ; and 800ms,  $F(1,12)=13.39$ ,  $p=0.003$ . The steps to target for easy-to find images were fewer by a significant amount than the hard to find images when the participants experienced a threshold of 400ms (mean difference of 7 steps) and 800ms (9

steps). Again, the image types influenced the steps to target when no randomly retrieved image is included in the display set,  $F(1,12)=18.42$ ,  $p=0.001$ , and when one randomly retrieved image is included,  $F(1,12)=7.53$ ,  $p=0.018$ . The image types also influenced the steps to target when participants experienced a threshold of 400ms with no randomly retrieved image,  $F(1,12)=7.02$ ,  $p=0.021$ , fixation threshold of 800ms with no randomly retrieved image,  $F(1,12)=11.36$ ,  $p=0.006$ , and fixation threshold of 800ms with one randomly retrieved image,  $F(1,12)=5.80$ ,  $p=0.033$ . However, there was no simple effect of image type on the 400ms fixation with one randomly retrieved image condition,  $F(1,12)=2.69$ ,  $p=0.13$ .

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 (40.468s and 125 fixations) than the hard-to-find images (71.331s 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  $F_2(1,12)=0.76$ ,  $p<0.40$ . Image type interacted with the fixation threshold,  $F_1(1,12)=8.04$ ,  $p<0.015$  and  $F_2(1,12)=5.84$ ,  $p<0.032$ , and an analysis of simple main effects indicated a significant difference in time to target and fixation numbers for the fixation thresholds when hard-to-find images were presented,  $F_1(1,12)=20.00$ ,  $p<0.001$  and  $F_2(1,12)=16.25$ ,  $p<0.002$ , but interestingly, no significant difference when easy-to-find images were presented,  $F_1(1,12)=3.62$ ,  $p<0.08$  and  $F_2(1,12)=3.57$ ,  $p<0.08$ .

## ETRA Symposium, San Diego, March 27-29, 2006

The same treatment combinations experienced by all participants were applied to the random selection tool to obtain 104 dependent variables (steps to target). By combining the variables, 208 figures were entered into a mixed design multivariate ANOVA with two observations per cell and three factors (selection mode, image type and randomly-retrieved).

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 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,  $F(2,23)=2.02$ ,  $p<0.16$ .

## 5 Discussion

The participants using the eye tracking interface found the target in fewer steps than the automated random selection strategy and the analysis of 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 random selection.

Easy-to-find target images were found in fewer steps by participants than the hard-to-find images as predicted by the evidence obtained using the random selection tool.

There were more fixations and more time was spent on hard-to-find images than the easy-to-find images. This is consistent with the conclusion of Fitts et al [1950] that complex information leads to longer fixation durations and higher fixation numbers.

The influence of including one randomly retrieved image in each display was investigated. Generally, there was little or no difference in the steps to target, time to target and fixation numbers. Even when compared with the random selection tool, the steps to target did not significantly differ.

There was no significant difference in the time to target and fixation numbers between the threshold levels for the easy-to-find images as opposed to the hard-to-find images. In other words, setting a higher threshold did not significantly differ when either 400ms or 800ms was used for the easy-to-find images, but it did for the hard-to-find images. However, the steps to target did differ for both image types under either of the threshold conditions. A future experiment will be needed to investigate whether the thresholds can be reduced further, at least for the easy-to-find images.

Many did not reach the hard target after 26 successive displays. Future experiment will concentrate on improving the chances of getting to the target using information extracted from the scan path.

## 6 Conclusions

Our experiments have shown that an eye tracking interface

together with pre-computed similarity measures yield a significantly better performance than random selection using the same similarity information. A significant effect on performance was also observed with hard-to-find images. This was not seen with easy-to-find images where with the current database size a random search might be expected to perform well.

An eye controlled image retrieval interface will not only provide a more natural mode of retrieval but also have the ability to anticipate the user's objectives coupled with user relevance feedback, thereby retrieving images extremely rapidly and with a minimum of thought and manual involvement. In future interfaces, eye tracking will not only be used as a rapid and continual information gathering tool for input to improve query formulation but also to build up a visual behavioural pattern using the time series information from the data. The ensuing interface will require a model for matching possible interests between images in the database. Visually similar regions will need to be linked between all regions within the images present in the database. Adaptive algorithms could then be used to improve the model for individual users.

## Acknowledgements

The authors acknowledge the support of BT Research and Venturing, SIRA and the Engineering and Physical Sciences Research Council in this work. The work has been conducted within the framework of the European Commission funded Network of Excellence "Multimedia Understanding through Semantics, Computation and Learning" (MUSCLE) [2005].

## References

- BAMIDELE A. AND STENTIFORD F. W. M 2005. An Attention Based Similarity Measure Used To Identify Image Clusters. In *European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology*, 30 Nov - 1st Dec., London, U.K.
- BAMIDELE A., STENTIFORD F. W. M AND MORPHETT J. 2004. An Attention-Based Approach to Content Based Image Retrieval. In *British Telecommunications Advanced Research Technology Journal on Intelligent Spaces*, Springer Verlag Book edition, (November 2004).
- CORNO F., FARINETTI L. AND SIGNORILE I. 2002. A cost effective solution for eye-gaze assistive technology. In *IEEE Int. Conf. on Multimedia and Expo*, August 26-29, Lausanne.
- COX I.J., MILLER M. L., MINKA T. P., PAPATHOMAS T. V., AND YIANILOS P. N. 2000. The Bayesian image retrieval system, PicHunter: theory, implementation, and Psychophysical experiments. In *IEEE Trans. on Image Processing*, Vol. 9, No 1.
- DESIMONE R. 1998. Visual attention mediated by biased competition in extrastriate visual cortex. In *Phil. Trans. R. Soc. Lond. B*, 353, 1245 – 1255.
- DUCHOWSKI, A. T. 2002. A Breadth-First Survey of Eye Tracking Applications. In *Behaviour Research Methods, Instruments, & Computers (BRMIC)*, 34(4), pp.455-470.
- FITTS, P.M., JONES, R.E., AND MILTON, J.L. 1950. Eye

## ETRA Symposium, San Diego, March 27-29, 2006

- Movement of Aircraft Pilots during Instrument-Landing Approaches. In *Aeronautical Engineering Review* 9, 24-29.
- GRIGORESCU C., PETKOV N., AND WESTENBERG M.A. 2003. Contour detection based on nonclassical receptive field inhibition. In *IEEE Trans. on Image Processing*, 12(7), 729-739.
- HANSEN J.P., ANDERSON A.W., AND P. ROED 1995. Eye gaze control of multimedia systems. In *Symbiosis of Human and Artifact*, Y. Anzai, K. Ogawa, and H. Mori (eds), Vol 20A, Elsevier Science, pp 37-42.
- HENDERSON JOHN M. AND HOLLINGWORTH ANDREW 1999. High-Level Scene Perception. In *Annual Reviews Psychology* 50:243-71.
- ITTI, L. 2004. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Trans. on Image Processing*, 13(10) (2004) 1304-1318.
- LC TECHNOLOGIES INC. 2005. <http://www.eyegaze.com/>
- MACKWORTH, N., AND MORANDI, A. 1967. The gaze selects informative details within pictures. In *Perception and Psychophysics* 2, 547-552.
- MCCARTHY, J, SASSE, M.A. & RIEGELSBERGER, J. 2003. Could I have the menu please? An eye tracking study of design conventions. In *Proceedings of HCI2003*, 8-12 Sep 2003, Bath, UK.
- MULTIMEDIA UNDERSTANDING THROUGH SEMANTICS, COMPUTATION AND LEARNING, NETWORK OF EXCELLENCE. 2005. *EC 6th Framework Programme. FP6-507752*. <http://www.muscle-noe.org/>
- NUMAJIRI T., NAKAMURA A., AND KUNO Y. 2002. Speed browser controlled by eye movements. In *IEEE Int Conf. on Multimedia and Expo*, August 26-29, Lausanne.
- OYEKOYA O. K., STENTIFORD F. W. M. 2004. Exploring Human Eye Behaviour Using a Model of Visual Attention. In *Proceedings of the International Conference on Pattern Recognition*, Cambridge UK, August 2004.
- OYEKOYA O. K., STENTIFORD F. W. M. 2005. A Performance Comparison of Eye Tracking and Mouse Interfaces in a Target Image Identification Task. In *2nd European Workshop on the Integration of Knowledge, Semantics & Digital Media Technolog (EWIMT)*, London, 30th Nov - 1st Dec, 2005.
- PAPATHOMAS T. V., CONWAY T. E., COX I. J., GHOSN J., MILLER M. L., MINKA T. P., AND YIANILOS P. N. 1998. Psychophysical studies of the performance of an image database retrieval system. In *Proc. IS&T/SPIE Conf. Human Vision Electronic Imaging III*, San Jose, CA, pp. 591-602.
- PUOLAMÄKI K., SALOJÄRVI J., SAVIA E., SIMOLA J., KASKI S. 2005. Combining Eye Movements and Collaborative Filtering for Proactive Information Retrieval. In *Proceedings of the 28th ACM Conference on Research and Development in Information Retrieval (SIGIR)*.
- URBAN J., JOSE J.M., VAN RIJSBERGEN C.J. 2003. An adaptive approach towards content-based image retrieval. In *Proc. of the Third International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 119-126.
- VENTERS, C.C., J.P. EAKINS AND R.J. HARTLEY 1997. The user interface and content based image retrieval systems. In *Proc. of the 19th BCS-IRSG Research Colloquium*, Aberdeen, April.
- VISHWA VINAY, INGEMAR J. COX, NATASA MILIC-FRAYLING, KENNETH R. WOOD 2004. Evaluating Relevance Feedback and Display Strategies for Searching on Small Displays. In *Proc. SPIRE*: 131-133.
- WARD D.J. AND MACKAY D.J.C. 2002. Fast hands-free writing by gaze direction. In *Nature* 418 pp 838, Aug. 22.
- YARBUS, A. 1967. *Eye Movements and Vision*. Plenum Press, New York.

# An Eye Tracking Interface for Image Search

Oyekoya and Stentiford

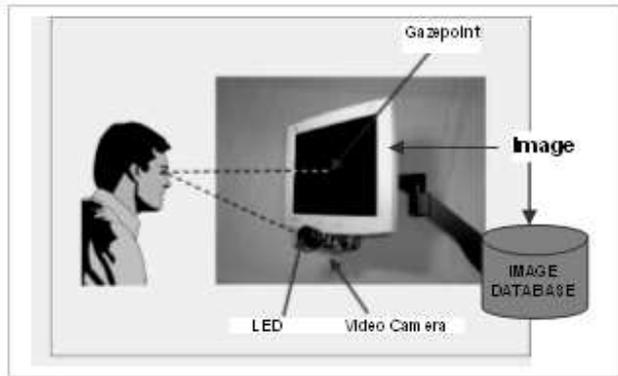


Figure 1: System Architecture



Figure 2: Standard start screens for all participants

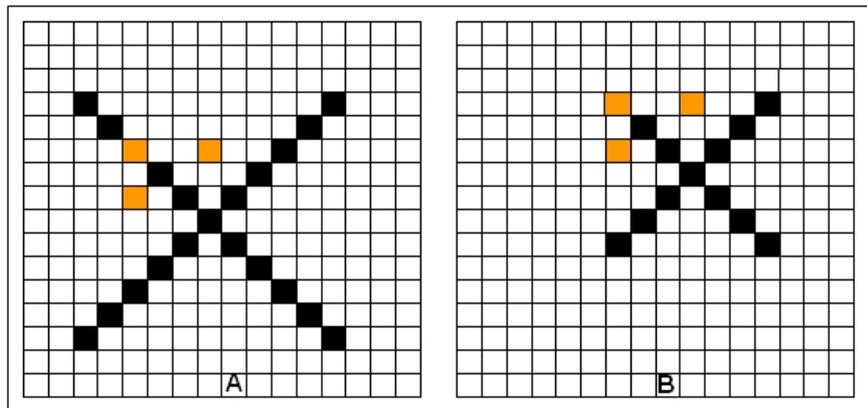


Figure 3: Neighbourhood at location x matching at y

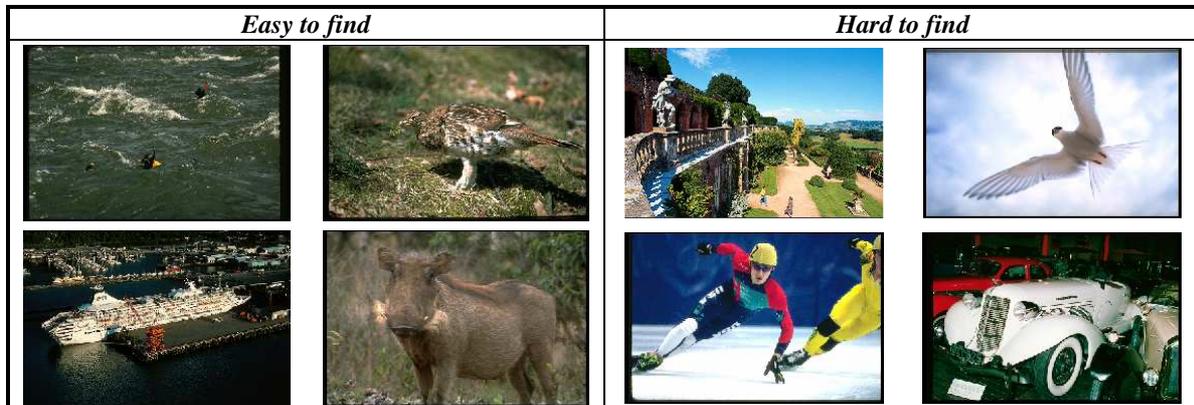


Figure 4: Target Images