Modelling perception using image processing algorithms

Pradipta Biswas, Peter Robinson
Computer Laboratory
15 JJ Thomson Avenue
Cambridge CB3 0FD
University of Cambridge, UK
E-mail: {pb400, pr}@cl.cam.ac.uk

ABSTRACT

User modeling is widely used in HCI but there are very few systematic HCI modelling tools for people with disabilities. We are developing user models to help with the design and evaluation of interfaces for people with a wide range of abilities. We present a perception model that can work for some kinds of visually-impaired users as well as for able-bodied people. The model takes a list of mouse events, a sequence of bitmap images of an interface and locations of different objects in the interface as input, and produces a sequence of eye-movements as output. Our model can predict the visual search time for two different visual search tasks with significant accuracy for both able-bodied and visually-impaired people.

Categories and Subject Descriptors

D.2.2 [Software Engineering]: Design Tools and Techniques – *user interfaces* **I.4.8 [Image Processing and Computer Vision]:** *Scene Analysis*

General Terms

Algorithms, Experimentation, Human Factors, Measurement,

Keywords

Human Computer Interaction, Perception Model, Image Processing.

1. INTRODUCTION

Computer Scientists have studied theories of perception extensively for graphics and, more recently, for Human-Computer Interaction (HCI). A good interface should contain unambiguous control objects (like buttons, menus, icons etc.) that are easily distinguishable from each other and reduce visual search time. In HCI, there are some guidelines for designing good interfaces (like colour selection rules and object arrangement rules [25]). However the guidelines are not always good enough. We take a different approach to compare different interfaces. We have developed a model of human visual perception for interaction with computer. Our model predicts visual search time for two search tasks and also shows the probable visual search path while searching a screen object for able-bodied as well as visually-impaired people. Different interfaces can then be compared using the predictions from the model.

We developed the model by using image processing techniques to identify a set of features that differentiate screen objects. We then calibrated the model to estimate fixation durations and eye movement trajectories. We evaluated the model by comparing its predicted visual search time with actual time for different visual search tasks.

In the next section we present a review of the state-of-the art perception models. In the following sections we discuss the design, calibration and validation of our model. Finally we make a comparative analysis of our model with other approaches and conclude by exploring possibilities for further research.

2. RELATED WORKS

Human vision has been addressed in many ways over the years. The Gestalt psychologists in early 19th century pioneered an interpretation of the processing mechanisms for sensory information [11]. Later the Gestalt principle gave birth to the top-down or constructivist theories of visual perception. According to this theory, the processing of sensory information is governed by our existing knowledge and expectations. On the other hand, bottom-up theorists suggest that perception occurs by automatic and direct processing of stimuli [11]. Considering both approaches, present models of visual perception incorporate both top-down and bottom-up mechanisms [17]. This is also reflected in recent experimental results in neurophysiology [15 & 22].

Knowledge about theories of perception has helped researchers to develop computational models of visual perception. Marr's model of perception is the pioneer in this field [16] and most of the other models follow its organization. In recent years, a plethora of models have been developed (e.g. ACRONYM, PARVO, CAMERA etc. [23]), which have also been implemented in computer systems. The working principles of these models are based on the general framework proposed in the analysis-by-synthesis model of Neisser [17] and also quite similar to the Feature Integration Theory of Triesman [27]. It mainly consists of the following three steps:

Feature extraction: As the name suggests, in this step the image is analysed to extract different features such as colour, edge, shape, curvature etc. This step mimics neural processing in the V1 region of the brain.

Perceptual grouping: The extracted features are grouped together mainly based on different heuristics or rules (e.g. the proximity and containment rule in the CAMERA system, rules of collinearity, parallelism and terminations in the ACRONYM system [23]). Similar types of perceptual grouping occur in V2 and V3 regions of the brain.

Object recognition: The grouped features are compared to known objects and the closest match is chosen as the output.

In these three steps, the first step models the bottom-up theory of attention while the last two steps are guided by top-down theories. All of these models aim to recognize objects from a background picture and some of them have been proved successful at recognizing simple objects (like mechanical instruments). However they have not demonstrated such good performance at recognizing arbitrary objects [23]. These early models do not operate at a detailed neurological level. Itti and Koch [13] present a review of computational models, which try to explain vision at the neurological level. Itti's pure bottom-up model [13] even worked in some natural environments, but most of these models are used to explain the underlying phenomena of vision (mainly the bottom-up theories) rather than prediction. The VDP model [6] uses image processing algorithms to model vision. The model predicts retinal sensitivity for different levels of luminance, contrast etc. Privitera and Stark [21] also used different image processing algorithms to identify points of fixations in natural scenes, however they do not have an explicit model to predict eye movement trajectory.

In the field of Human Computer Interaction, the EPIC [14] and ACT-R [1] cognitive architectures have been used to develop perception models for menu searching and icon searching tasks. Both the EPIC and ACT-R models [12 & 5] are used to explain the results of Nielsen's experiment on searching menu items [18], and found that users search through a menu list in both systematic and random ways. The ACT-R model has also been used to find out the characteristics of a good icon in the context of an icon-searching task [9 & 10]. However the cognitive architectures emphasize modeling human cognition and so the perception and motor modules in these systems are not as well developed as the remainder of the system. The working principles of the perception models in EPIC and ACT-R/PM are simpler than the earlier general-purpose computational models of vision. These models do not use any image processing algorithms [9, 10 & 12]. The features of the target objects are manually fed into the system and they are manipulated by handcrafted rules in a rule-based system. As a result, these models do not scale well to general-purpose interaction tasks. It will be hard to model the basic features and perceptual similarities of complex screen objects using propositional clauses. Modelling of visual impairment is particularly difficult using these models. An object seems blurred in a continuous scale for different degrees of visual acuity loss and this continuous scale is hard to model using propositional clauses in ACT-R or EPIC. Shah et. al. [26] have proposed the use of image processing algorithms in a cognitive model, but they have not published any results about the predictive power of their model yet.

In short, approaches based on image processing have concentrated on predicting points of fixations in complex scenes while researchers in HCI mainly try to predict the eye movement trajectories in simple and controlled tasks. There has been less work on using image processing algorithms to predict fixation durations and combining it with a suitable eye movement strategy in a single model. The EMMA model [24] is an attempt in that direction, but it does not use any image processing algorithm to quantify the perceptual similarities among objects. We have separately calibrated our model for predicting fixation duration based on perceptual similarities of objects and also calibrated it for predicting eye movements. The calibrated model can predict the visual search time for two different visual search tasks with significant accuracy for both able-bodied and visually-impaired people.

3. DESIGN

Our perception model takes a list of mouse events, a sequence of bitmap images of an interface and locations of different objects in the interface as input, and produces a sequence of eye-movements as output. The model is controlled by four free parameters: distance of the user from the screen, foveal angle, parafoveal angle and periphery angle (Figure 1). The default values of these parameters are set according to the EPIC architecture [14].

Our model follows the 'spotlight' metaphor of visual perception. We perceive something on a computer screen by focusing attention at a portion of the screen and then searching for the desired object within that area. If the target object is not found we look at other portions of the screen until the object is found or the whole screen is scanned. Our model simulates this process in three steps.

1. Scanning the screen and decomposing it into primitive features.

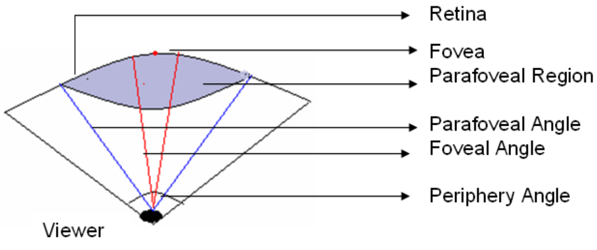


Figure 1. Foveal, parafoveal and peripheral vision

2. Finding the probable points of attention fixation by evaluating the similarity of different regions of the screen to the one containing the target.

3. Deducing a trajectory of eye movement.

The perception model represents a user's area of attention by defining a focus rectangle within a certain portion of the screen. The area of the focus rectangle is calculated from the distance of the user from the screen and the periphery angle (distance X tan(periphery angle /2), Figure 1). If the focus rectangle contains more than one probable target (whose locations are input to the system) then it shrinks in size to investigate each individual item. Similarly in a sparse area of the screen, the focus rectangle increases in size to reduce the number of attention shifts.

The model scans the whole screen by dividing it into several focus rectangles, one of which should contain the actual target. The probable points of attention fixation are calculated by evaluating the similarity of other focus rectangles to the one containing the target. We know which focus rectangle contains the target from the list of mouse events that was input to the system. The similarity is measured by decomposing each focus rectangle into a set of features (colour, edge, shape etc.) and then comparing the values of these features. The focus rectangles are aligned with respect to the objects within them during comparison. Finally, the model shifts attention by combining different eye movement strategies (like Nearest [7, 8], Systematic, Cluster [9, 10] etc.), which are discussed later.

The model can also simulate the effect of visual impairment on interaction by modifying the input bitmap images according to the nature of the impairment (like blurring for visual acuity loss, changing colours for colour blindness). We discussed the modelling of visual impairment in detail in a separate paper [4]. In this paper, we discuss the calibration and validation of the model using the following experiment.

4. EXPERIMENT TO COLLECT EYE TRACKING DATA

In this experiment, we investigated how eyes move across a computer screen while searching for a particular target. We kept the searching task very simple to avoid any cognitive load. The eye gazes of users were tracked by using a Tobii X120 eye-tracker [28].

4.1. Design

We conducted trials with two families of icons. The first consisted of geometric shapes with colours spanning a wide range of hues and luminances (Figure 2). The second consisted of images from the system folder in Microsoft Windows to increase the external validity (Figure 3) of the experiment.

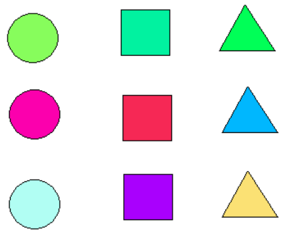


Figure 2 Corpus of Shapes

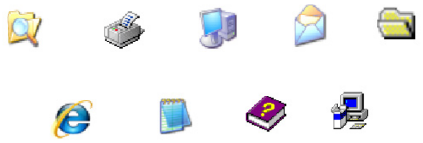


Figure 3. Corpus of Icons

4.2. Participants

We collected data from 8 visually impaired and 10 able bodied participants (Table 1). All were expert computer users and had no problem in using the experimental set up.

Table 1. List of Participants

	Age	Gender	Impairment
C1	22	M	Able-bodied
C2	29	M	
C3	27	M	
C4	30	F	
C5	24	M	
C6	28	M	
C7	29	F	
C8	50	F	
C9	27	M	
C10	25	M	
P1	24	M	Retinopathy
P2	22	M	Nystagmus and acuity loss due to Albinism
P3	22	M	Myopia (-3.5 Dioptre)
P4	50	F	Colour blindness - Protanopia
P5	24	F	Myopia (-4.5 Dioptre)
P6	24	F	Myopia (-5.5 Dioptre)
P7	27	M	Colour blindness - Protanopia
P8	22	M	Colour blindness - Protanopia

4.3. Material

We used a 1024 × 768 LCD colour display driven by a 1.7 GHz Pentium 4 PC running the Microsoft Windows XP operating system. We also used a standard computer Mouse (Microsoft IntelliMouse® Optical Mouse) for clicking on the target and a Tobii X120 Eye Tracker for tracking eye gaze pattern, which has an accuracy of 0.5° of visual angle. The Tobii studio software was used to extract the points of fixation. We used the default fixation filter (Tobii fixation filter) and fixation radius (minimum distance to separate two fixations) of 35 pixels.

4.4. Process

The experimental task consisted of shape searching and icon searching tasks. The task was as follows

1. A particular target (shape or icon) was shown.
2. A set of 18 candidates was shown.
3. Participants were asked to click on the candidate(s), which are same as the target.
4. The number of candidates similar to the target was randomly chosen between 1 and 8 to simulate both serial and parallel searching effects [27], the other candidates were distractors.
5. The candidates were separated by 150 pixels horizontally and by 200 pixels vertically.
6. Each participant did five shape searching and five icon searching tasks.

4.5. Calibration for predicting fixation duration

Initially we measured the drift of the eye tracker for each participant. The drift was smaller than half the separation between the candidates, so we could classify most of the fixations around the candidates. We calibrated the model to predict fixation duration by following two steps.

Step 1: Calculation of image processing coefficients and relating them to the fixation duration

We calculated the colour histogram [19] and shape context coefficients [2, 3] between the targets and distractors, and measured their correlation with the fixation durations (Table 1). The image processing coefficients correlate significantly with the fixation duration, though the significance is not indicative of their actual predictive power, as the number of data points is large. However, the colour histogram algorithm in YUV space is moderately correlated (0.51) with the fixation duration (Figure 4).

We then used an SVM and a cross-validation test to identify the best feature set for predicting fixation duration for each participant as well as for all participants. We found that the Shape Context Similarity coefficient and the Colour Histogram coefficient in YUV space work best for all participants taken together. The combination also performs well enough (within the 5% limit of the best classifier) for individual participants. The classifier takes the Shape Context Similarity coefficient and Colour Histogram coefficient in YUV space of a target as input and predicts the fixation duration on it as output.

Table 1. Correlation between fixation duration and image processing algorithms

Image Statistics	Colour Histogram (YUV)	Colour Histogram (RGB)	Shape Context	Edge Similarity
Spearman's Rho	0.507	0.444	0.383	0.363

**All are significant at 0.01 level

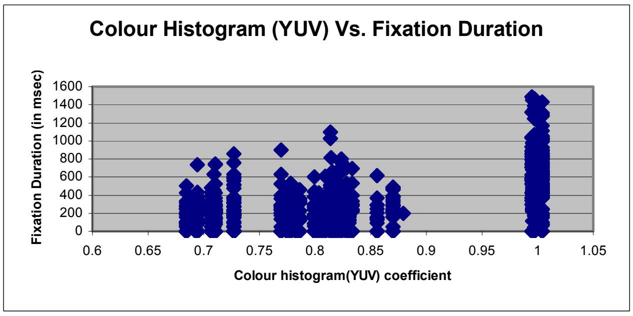


Figure 4. Relating colour histogram coefficients with fixation duration

Step 2: Number of fixations

We found in the eye tracking data that users often fixed attention more than once on targets or distractors. We investigated the number of fixations with respect to the fixation durations (Figures 5 and 6). We assumed that in case of more than one attention fixation, the recognition took place during the fixation with the largest duration. Figure 6 shows the total number of fixations with respect to the maximum fixation duration for all able-bodied users and each visually-impaired user.

We found that visually impaired people fixed eye gaze a greater number of times than their able bodied counterparts. Participant P2 (who has nystagmus) has many fixations of duration less than 100 msec and only two fixations having duration more than 400 msec.

It can be seen as the fixation duration increases, the number of fixations also decreases (Figures 5 and 6). This can be explained by the fact that when the fixation duration is higher, the users can recognize the target and do not need more long fixations on it. The number of fixations is smaller when the fixation duration is less than 100 msec, probably these are fixations where the distractors are very different from the targets and users quickly realize that they are not intended target. In our model, we predict the maximum fixation duration using the image processing coefficients (as discussed in the previous section) and then decide the number of fixations based on the value of that duration.

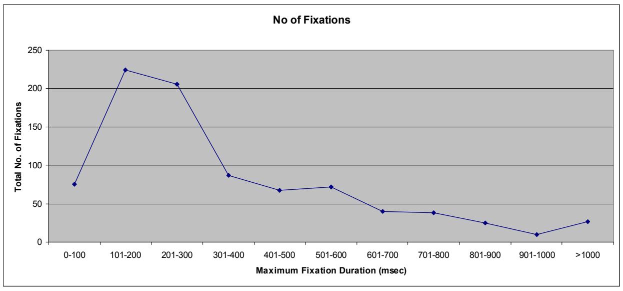


Figure 5. Total no. of fixations w.r.t. fixation duration

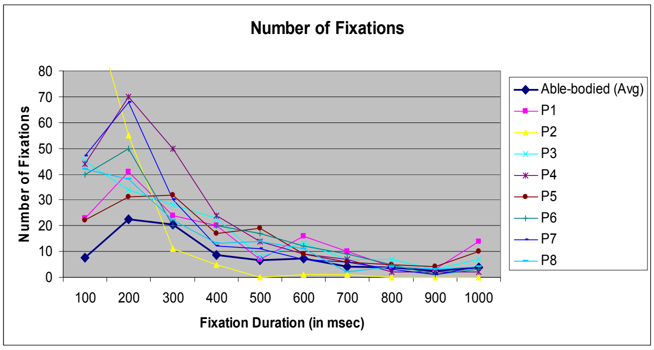


Figure 6. Number of fixations w.r.t. fixation duration

5.6. Calibration for predicting eye movement patterns

We investigated different strategies to explain and predict the actual eye movement trajectory. We rearranged the points of fixation given by the eye tracker following different eye-movement strategies and then compared the rearrangements with the actual sequences (which signify the actual trajectory).

We used the average Levenshtein distance between actual and predicted eye fixation sequences to compare different eye movement strategies. We converted each sequence of points of fixation into a string of characters by dividing the screen into 36 regions and replacing a point of fixation by a character according to its position in the screen [21]. The Levenshtein distance measures the minimum number of operations needed to transform one string into the other, where an operation is an insertion, deletion, or substitution of a single character. We considered the following eye movement strategies,

Nearest strategy [9 and 10]: At each instant, the model shifts attention to the nearest probable point of attention fixation from the current position.

Systematic Strategy: Eyes move systematically from left to right and top to bottom.

Random Strategy: Attention randomly shifts to any probable point of fixation.

Cluster Strategy: The probable points of attention fixation are clustered according to their spatial position and attention shifts to the centre of one of these clusters. This strategy reflects the fact that a saccade tends to land at the centre of gravity of a set of possible targets [7, 8 & 20], which is particularly noticeable in eye tracking studies on reading tasks.

Cluster Nearest (CN): The points of fixations are clustered and the first saccade launches at the centre of the biggest cluster (highest number of points of fixation). Then the strategy switches to the Nearest strategy.

Figures 7 and 8 show the average Levenshtein distance for different eye movement strategies for able-bodied and visually-impaired participants respectively.

The best strategy varies across participants. However one of the Cluster, Nearest and Cluster Nearest (CN) strategies comes as best for each participant individually. We did not find any difference in the eye movement pat-terns of able-bodied and visually impaired users. If we consider all participants together, the Cluster Nearest strategy is the best. It is also significantly better than the random strategy (Figure 9, paired T-test, t = 3.895, p<0.0005), which indicates that it actually captures the pattern of eye movement in most of the cases.

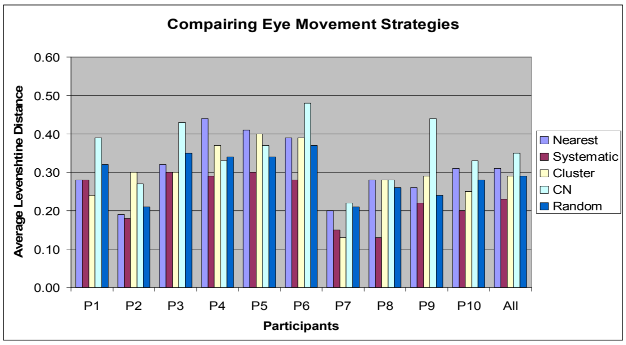


Figure 7. Average Levenshtein Distance for different eye movement strategies for able bodied users

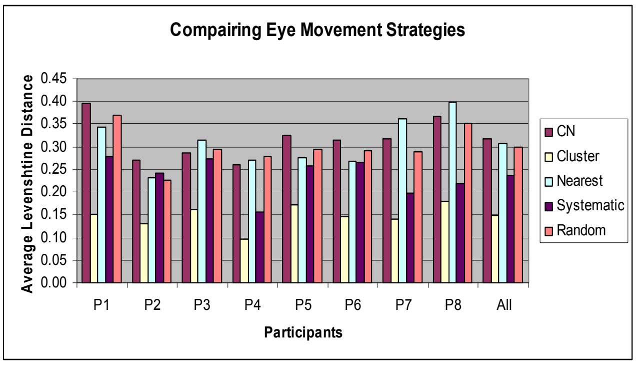


Figure 8. Average Levenshtein Distance for different eye movement strategies for visually impaired users

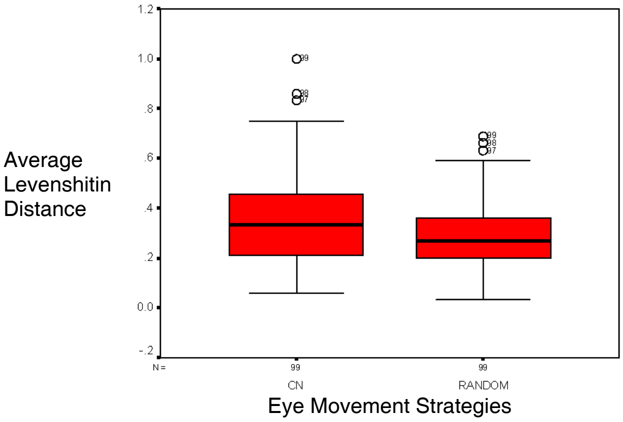


Figure 9. Comparing the best strategy against the Random strategy

5. VALIDATION

Initially we have used a 10-fold cross-validation test on the classifiers to predict fixation durations. In this test we randomly select 90% of the data for training and test the prediction on the remaining 10%. The process is repeated 10 times and the prediction error is averaged. It can be seen that the prediction error is less than or equal to 40% for 12 out of 18 participants and 40% taking all participants together (Figure 10).

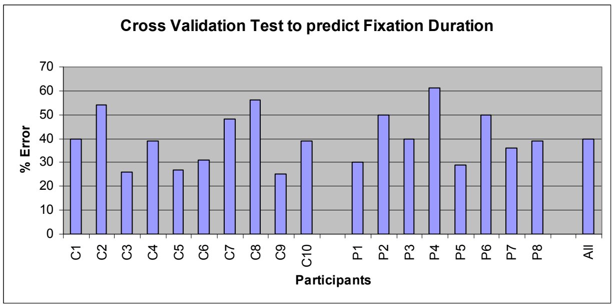


Figure 10. Cross validation test on the classifiers

Then, we have used our model to predict the total fixation time (summation of all fixations, which is nearly same as the visual search time) for each individual search task by each participant. Table 2 shows the correlation coefficient between actual and predicted time for each participant. Figure 11 shows a scatter plot of the actual and predicted times taking all able-bodied participants together and Figure 12 shows the scatter plot for each visually-impaired participant.

Table 2. Correlation between actual and predicted total fixation time

Participants	Correlation
C1	0.740*
C2	0.788**
C3	0.784**
C4	0.455
C5	0.441
C6	0.735*
C7	0.530
C8	-0.309
C9	0.910**
C10	0.655*
P1	0.854**
P2	0.449
P3	0.625
P4	0.666*
P5	0.843**
P6	0.761**
P7	0.728**
P8	0.527

** $p< 0.01$
* $p< 0.05$

For able-bodied participants, the predicted time significantly correlates with the actual for 6 participants (each undertook 10 search tasks), correlates moderately for 3 participants and did not work for one participant (participant C8). For visually impaired participants, the predicted time significantly correlates with the actual for 5 participants (each undertook 10 search tasks), correlates moderately for 3 participants. We are currently working to improve the accuracy further.

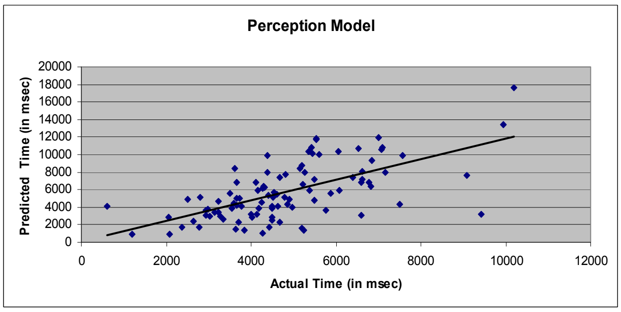


Figure 11. Scatter plot of actual and predicted time for able-bodied users

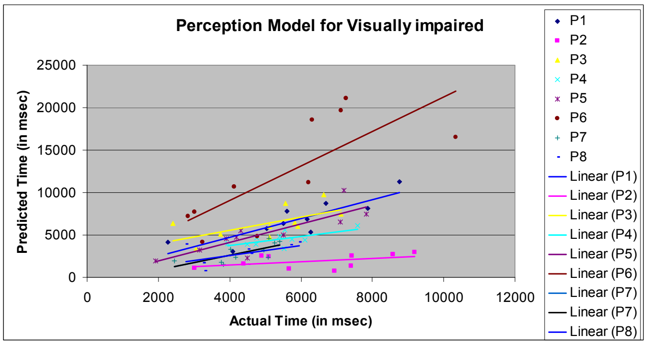


Figure 12. Scatter plot of actual and predicted time for visually-impaired users

We also validated the model using a Leave-1-out validation test. In this process we tested the model for each participant by training the classifiers using the data from the other participants. Figure 13 shows the scatter plot of actual and predicted time and Figure 14 shows the histogram of percent error. The predicted and actual time correlates significantly (ρ = 0.5, p<0.01) while the average error in prediction is about 40%.

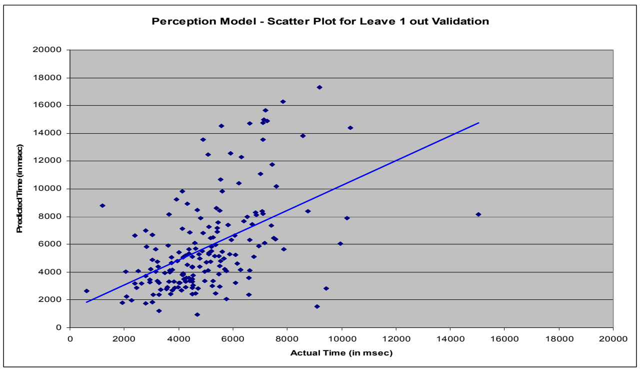


Figure 13. Scatter plot of predicted and actual time

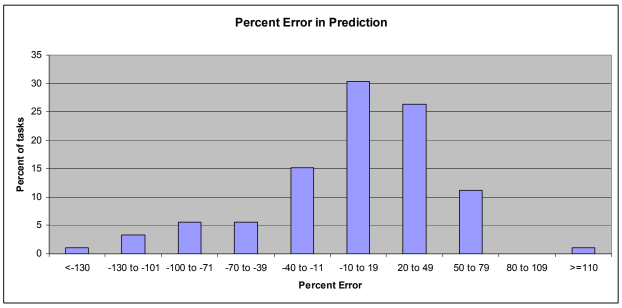


Figure 14. Percent error in prediction

Then we validated the model by taking data from some new participants (Table 3). We used a single classifier for all of them which was trained by our previous data set. We did not change the value of any parameter of the model for any participant. Table 3 shows the correlation coefficients between actual and predicted time for each participant. Figure 15 shows a scatter plot of the actual and predicted times for each participant. It can be seen our prediction significantly correlate with actual for 6 out of 7 participants.

Table 4 shows the actual and predicted visual search paths for some sample tasks. The prediction is similar though not exactly same. Our model successfully detected most of the points of fixation. In the second picture of Table 3, we have only one target, which pops out from the background. Our model successfully captures this parallel searching effect while the serial searching is also captured in the other cases. In the last figure we show the prediction for a protanope (a type of colour-blindness) participant and so the right hand figure is different from the left hand one as we simulate the effect of protanopia on the input image.

Table 3. New Participants

Participants	Age	Gender	Correlation	Impairment
V1	29	F	0.64*	None
V2	29	M	0.89**	None
V3	25	F	0.7*	None
V4	25	F	0.72*	Myopia -4.75/-4.5
V5	25	F	0.69*	Myopia -3.5
V6	27	F	0.44	Myopia -8/-7.5
V7	26	M	0.7*	None

*p<0.05
**p<0.01

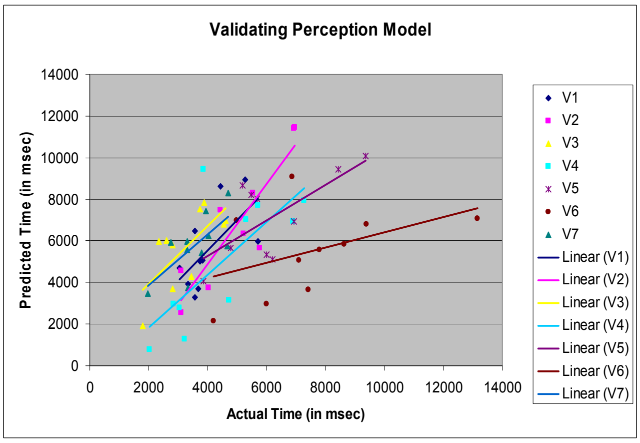


Figure 15. Scatter plot of actual and predicted time for new users

6. DISCUSSION

The eye-tracking data shows that the eye movement patterns are different for different participants. The performance of the eye tracker (drift, fixation identification etc.) also differs across participants.

We found that the visual search time is greater for visually-impaired users than for able-bodied users. However, the eye movement strategies of visually impaired users are not different from their able-bodied counterparts. This is due to the fact that the V4 region in the brain controls the visual scanning and our visually-impaired participants did not have any brain injury and so the V4 region worked the same as the able-bodied users. However visually-impaired users had a greater number of attention fixations which made the search time longer. Additionally the difference between the numbers of fixations for able-bodied and visually impaired users is more prominent for shorter duration (less than 400 msec) fixations. Perhaps this means visually impaired users need many short duration fixations to confirm the recognition of target. From an interface designers' point of view, these results indicates that the clarity and distinctiveness of targets are more important than the arrangement of the targets in a screen. Since the eye-movement patterns are almost same for all users, the arrangement of the targets need not be different to cater visually-impaired users. However clarity and distinctiveness of targets will reduce the visual search time by reducing recognition time and number of fixations as well.

Regarding our model, we tried to keep it as general as possible by using the same feature set (Shape Context Similarity coefficient and Colour Histogram coefficient in YUV space) to predict fixation duration for all participants. Additionally we also used the same eye move#ent strategy (Cluster Nearest) for all participants. The result demonstrates that

- The model is robust and scalable.
- The accuracy can be further increased by personalizing it for each individual user.

The experimental task consisted of searching for both basic shapes and real life icons. We found that the

Table 4. Actual and predicted visual search path

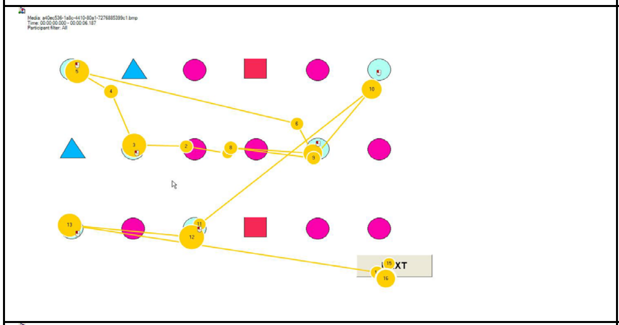
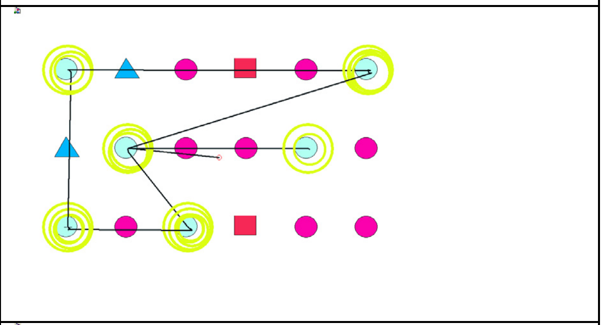
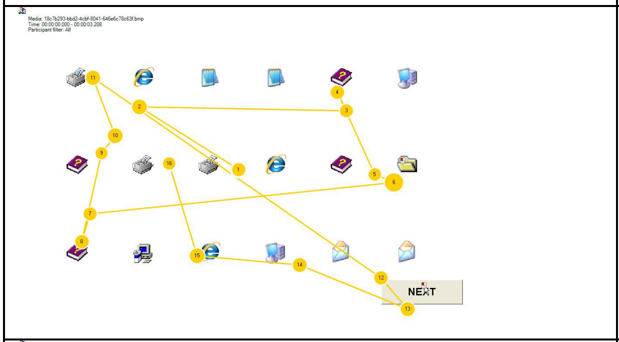
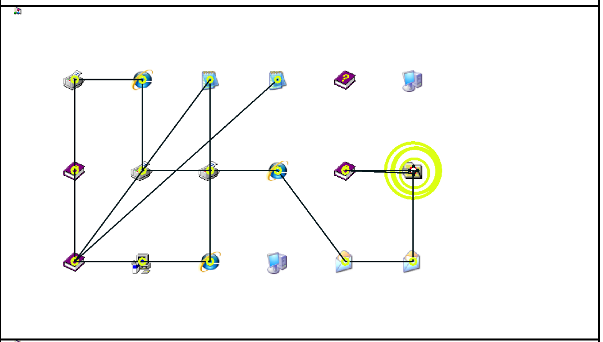
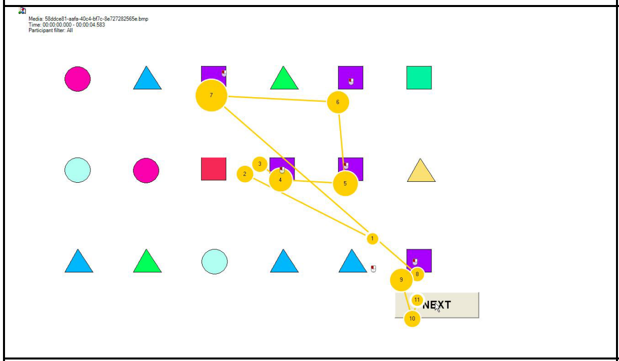
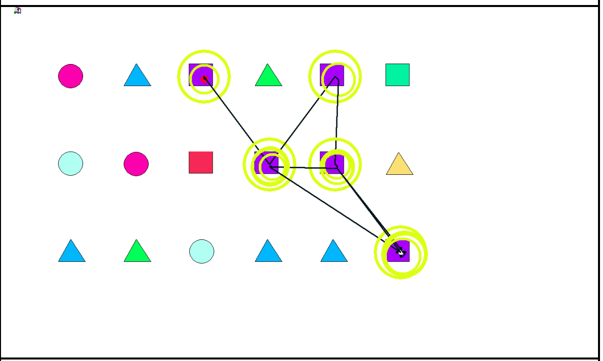
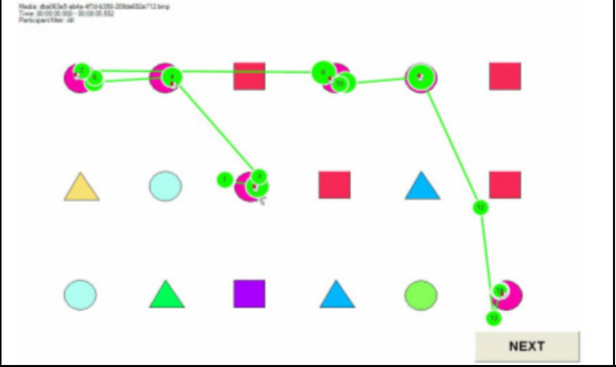
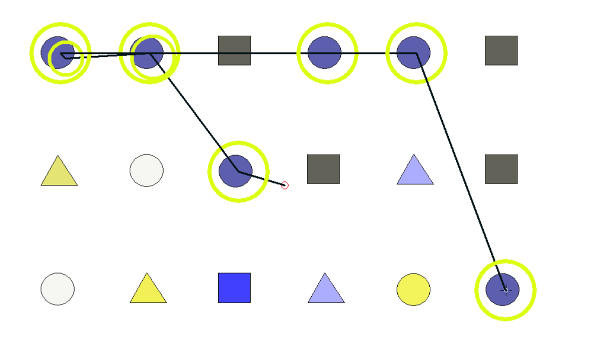
Actual Eye Gaze Pattern	Predicted Eye Gaze Pattern
	
	
	
	

Table 5. Comparative analysis of our model

	ACT-R/PM or EPIC models	Our Model	Advantages of our model
Storing Stimuli	Propositional Clauses	Spatial Array	Easy to use and Scalable
Extracting Features	Manually	Automatically using Image Processing algorithms	
Matching Features	Rules with binary outcome	Image processing algorithms that give the minimum squared error	More accurate
Modelling top down knowledge	Not relevant as applied to very specific domain.	Considers the type of target (e.g. button, icon, combo box etc.).	More detailed and practical
Shifting Attention	Systematic/ Random and Nearest strategy	Clustering/ Nearest /Random strategy	Not worse than previous, probably more accurate

fixation duration does not depend on the type of the target (icon/shape), hence, the model does not need to be tuned for a particular task and works for both types of search task. Table 5 presents a comparative analysis of our model with the ACT-R/PM and EPIC models. Our model seems to be more accurate, scalable and easier to use than the existing models.

However, in real life situations the model fails to take account of the domain knowledge of users. This knowledge can be either application specific or application independent. There is no way to simulate application specific domain knowledge without knowing the application beforehand. However there are certain types of domain knowledge that are application independent and apply for almost all applications. For example, the appearance of a pop-up window immediately shifts attention in real life, however the model still looks for probable targets in the other parts of the screen. Similarly, when the target is a text box, users focus attention on the corresponding labels rather than other text boxes, which we do not yet model. There is also scope to model perceptual learning. For that purpose, we could incorporate a factor like the frequency factor of EMMA model [24] or consider some high level features like the caption of a widget, handle of the application etc. to remember the utility of a location for a certain application. These issues did not arise in most previous work since they considered very specific and simple domains.

7. CONCLUSION

In this work, we have developed a systematic model of visual perception which works for people with a wide range of abilities. We have used image processing algorithms to quantify the perceptual similarities among objects and predict the fixation duration based on that. We also calibrated our model by considering different eye movement strategies. Our model intended to be used by software engineers to design software interfaces. So we tried to make the model easy to use and comprehend. As a result it is not so detailed and accurate to explain the results of any psychological experiment on visual perception. However, it is accurate enough to select the best interface among a pool of interfaces based on the visual search time. Additionally, it can be tuned to capture the individual differences among users and to give accurate prediction for any user.

ACKNOWLEDGEMENTS

We would like to thank the Gates Cambridge Trust for funding this work. We like to thank the participants from Cambridge to take part in our experiments. We are grateful to Dr. H. M. Shah (Shah & Shah), Prof. Gary Rubin (UCL) and Prof. John Mollon (Univ. of Cambridge) for their useful suggestions regarding visual impairment simulation. We also like to thank Dr. Alan Blackwell of University of Cambridge and Dr. T. Metin Sezgin for their help in developing the model.

REFERENCES

[1] Anderson, J. R., & Lebiere, C., The Atomic Components of Thought. Hillsdale, NJ: Erlbaum, 1998

[2] Belongie S., Malik J., & Puzicha J., Shape Matching & Object Recognition Using Shape Contexts, IEEE Transactions on Pattern Analysis & Machine Intelligence 24 (24): 509-521, 2002

[3] Belongie S., Malik J., and Puzicha J. "Shape Context: A new descriptor for shape matching and object recognition". NIPS 2000.

[4] Biswas P. and Robinson P., Modelling user interfaces for special needs, Pradipta Biswas, Peter Robinson, Accessible Design in the Digital World (ADDW) 2008 Available from: http://www.cl.cam.ac.uk/~pb400/Papers/pbiswas_ADDW08.pdf Accessed on: 12/12/08

[5] Byrne M. D., ACT-R/PM & Menu Selection: Applying A Cognitive Architecture To HCI, International Journal of Human Computer Studies,vol. 55, 2001

[6] Daly S. 1993. The Visible Differences Predictor: An algorithm for the assessment of image fidelity. In Digital Images and Human Vision, A. B. Watson, Ed. MIT Press, Cambridge, MA, 179–206, 1993

[7] Findlay J. M., Programming of Stimulus-Elicited Saccadic Eye Movements. In K. Rayner (Ed.), Eye Movements and Visual Cognition: Scene Perception and Reading, New York, Springer Verlag (Springer series in Neuropsychology) 8-30, 1992

[8] Findlay J. M., Saccade Target Selection during Visual Search, Vision Research, 37 (5), 617-631, 1997

[9] Fleetwood , M. F. and Byrne, M. D., 2006. Modeling the Visual Search of Displays: A Revised ACT-R Model of Icon Search Based on Eye-Tracking Data, Human-Computer Interaction, Vol. 21, No. 2, 153-197, 2006

[10] Fleetwood, M. F. & Byrne, M. D. Modeling icon search in ACT-R/PM.Cognitive Systems Research, Vol. 3 (1), 25-33,2002

[11] Hampson P. & Moris P., Understanding Cognition, Blackwell Publishers Ltd., Oxford, UK, 1996

[12] Hornof, A. J. & Kieras, D. E., Cognitive Modeling Reveals Menu Search Is Both Random & Systematic. In Proc. of the ACM/SIGCHI Conference on Human Factors in Computing Systems, 107-115, 1997

[13] Itti L. & Koch C., Computational Modelling of Visual Attention, Nature Reviews, Neuroscience, Vol. 2, 1-10, March 2001.

[14] Kieras, D. & Meyer, D.E.. An Overview of The EPIC Architecture For Cognition & Performance With Application To Human-Computer Interaction, Human-Computer Interaction, vol. 14, 391-438, 1990

[15] Luck S. J. et. al., Neural Mechanisms of Spatial Selective Attention In Areas V1, V2, & V4 of Macaque Visual Cortex, Journal of Neurophysiology, vol. 77, 24-42, 1997

[16] Marr, D. C., Visual Information Processing: the structure & creation of visual representations. Philosophical Transactions of the Royal Society of London B, 290, 199-218, Jul 8, 1980

[17] Neisser, U., Cognition & Reality, San Francisco, Freeman, 1976

[18] Nilsen E. L., Perceptual-motor Control in Human-Computer Interaction (Technical Report No. 37), Ann Arbor, MI: The Cognitive Science & Machine Intelligence Laboratory, the Univ. of Michigan, 1992

[19] Nixon M. & Aguado A., Feature Extraction & Image Processing, Elsevier, Oxford, First Ed., 2002

[20] O'Regan K. J., Optimal Viewing position in words and the Strategy-Tactics Theory of Eye Movements in Reading, In K. Rayner (Ed.), Eye Movements and Visual Cognition: Scene Perception and Reading, New York, Springer Verlag (Springer series in Neuropsychology) 333-355, 1992

[21] Privitera C. M. and Stark L. W., Algorithms for defining Visual Regions-of-Interests: Comparison with Eye Fixations. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 22(9), 970-982, 2000

[22] Reynolds J. H. & Desimone R., The Role of Neural Mechanisms of Attention In Solving The Binding Problem, Neuron 24: 19-29, 111-145, 1999

[23] Rosandich, R. G., Intelligent Visual Inspection using artificial neural networks, Chapman & Hall, London, First Edition, 1997

[24] Salvucci D. D., An integrated model of eye movements & visual encoding, Cognitive Systems Research, January, 2001

[25] Shneiderman B., Designing the User Interface: Strategies for Effective Human--computer Interaction, Addison-Wesley, 1992

[26] Shah K. et. al., Connecting a Cognitive Model to Dynamic Gaming Environments: Architectural & Image Processing Issues, In Proc. of the 5th Intl. Conf. on Cognitive Modeling,189-194, 2003

[27] Treisman A. and Gelade G., A Feature Integration Theory of Attention, Cognitive Psychology, 12, 97-136, 1980

[28] Tobii Eye Tracker, Available online: http://www.imotionsglobal.com/Tobii+X120+Eye-Tracker.344.aspx Accessed on: 12/12/08