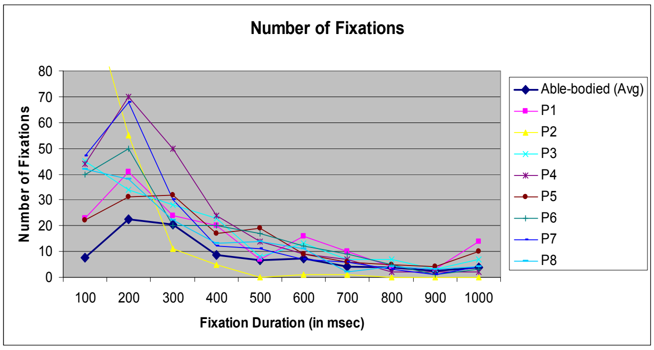



Figure 6. Number of fixations w.r.t. fixation duration

### 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.

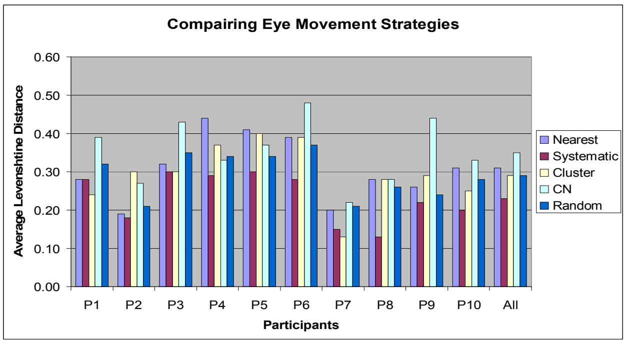



Figure 7. Average Levenshtein Distance for different eye movement strategies for able bodied users

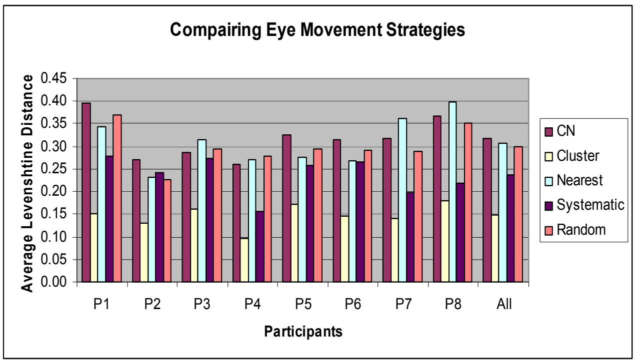



Figure 8. Average Levenshtein Distance for different eye movement strategies for visually impaired users

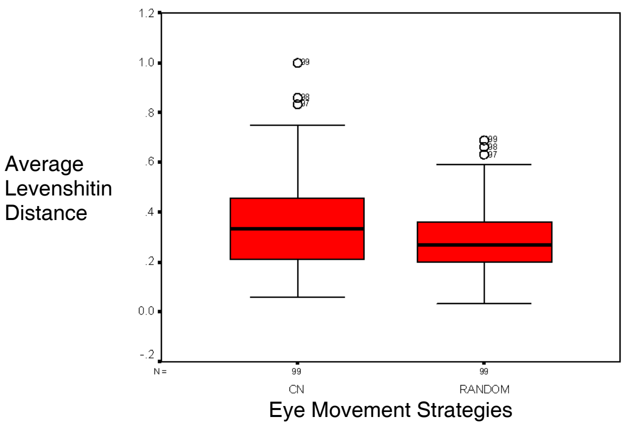



Figure 9. Comparing the best strategy against the Random strategy

## 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).

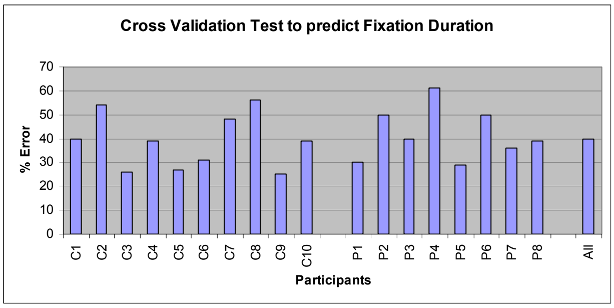



Figure 10. Cross validation test on the classifiers

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.

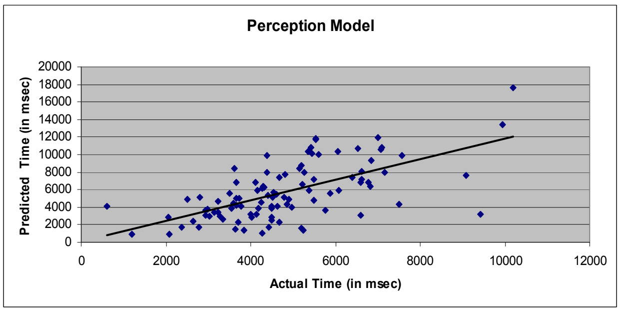



Figure 11. Scatter plot of actual and predicted time for able-bodied users

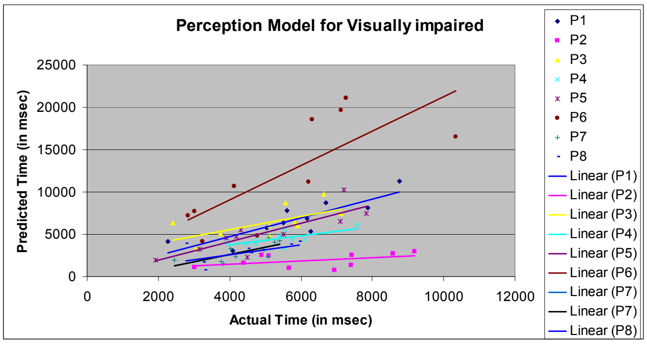



Figure 12. Scatter plot of actual and predicted time for visually-impaired users

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 (□ = 0.5, $p<0.01$) while the average error in prediction is about 40%.

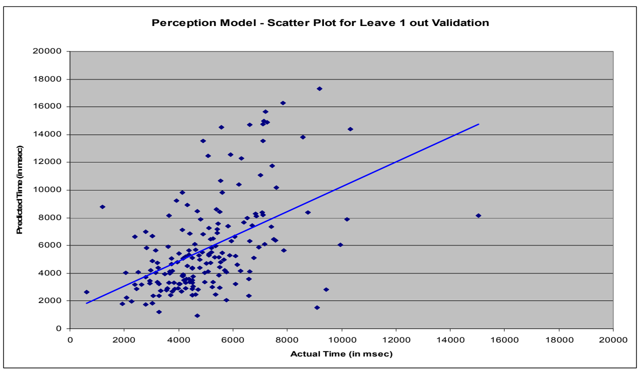



Figure 13. Scatter plot of predicted and actual time

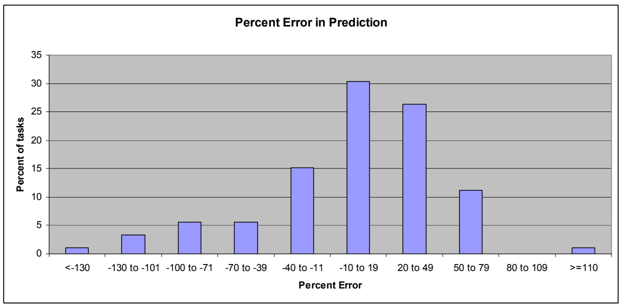



Figure 14. Percent error in prediction

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

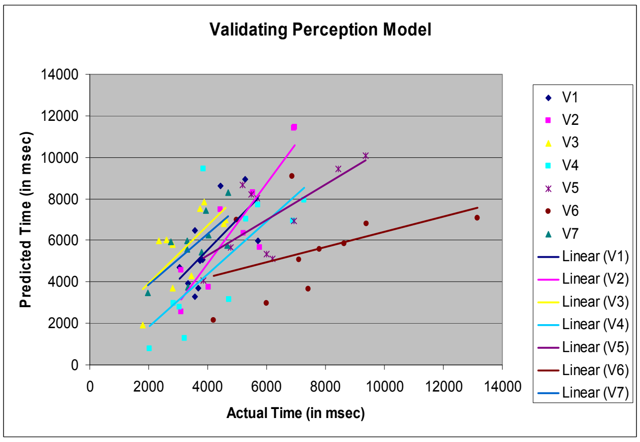



Figure 15. Scatter plot of actual and predicted time for new users

## 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

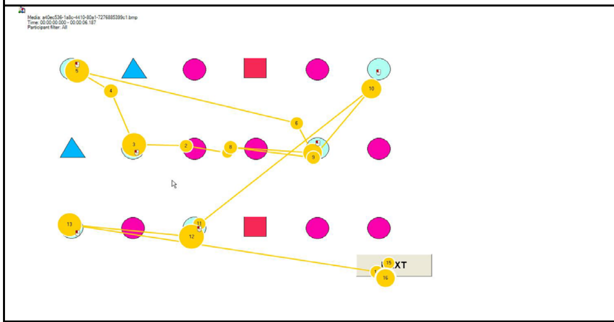
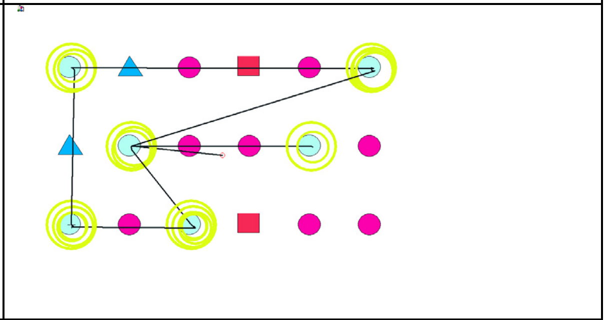
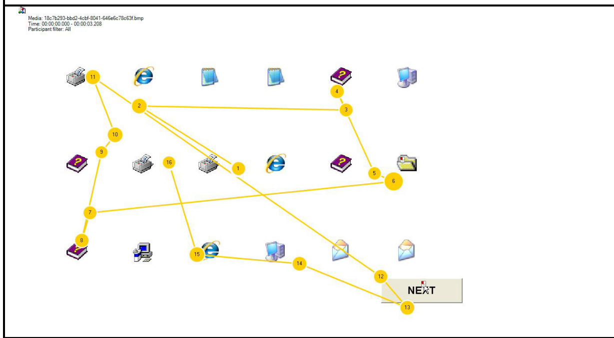
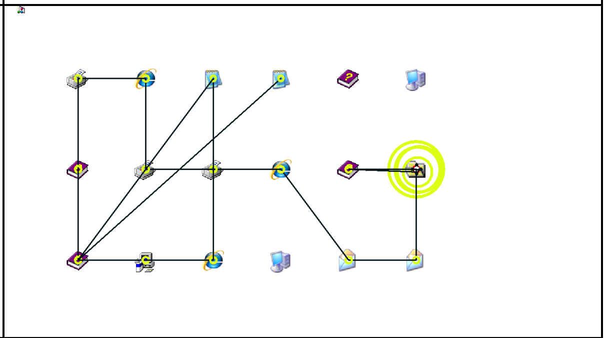
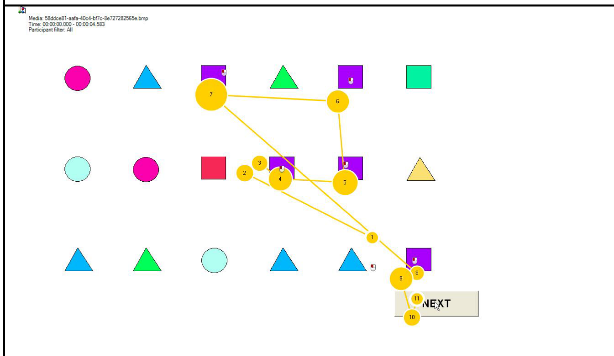
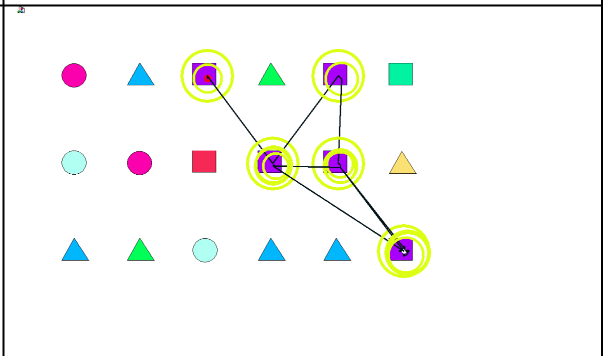
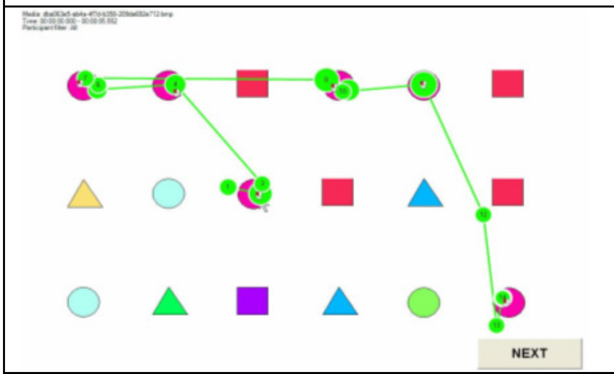
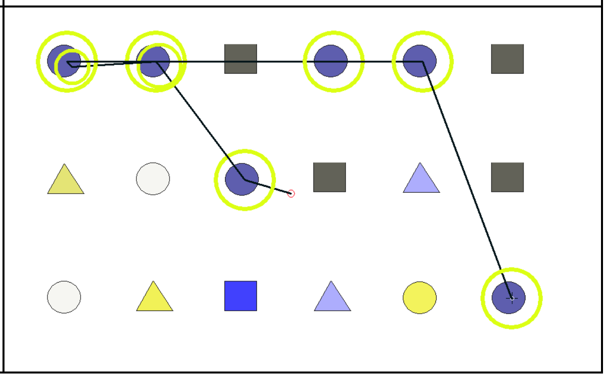
| Actual Eye Gaze Pattern | Predicted Eye Gaze Pattern |
|---|---|
|  |  |
|   |  |
|   |  |
|   |  |

**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