Towards a Flexible, Reusable Model for Predicting Eye Movements During Visual Search of Text

Tim Halverson thalvers@cs.uoregon.edu
Anthony J. Hornof hornof@cs.uoregon.edu
Department of Computer and Information Science
University of Oregon

Abstract

Eye movements produced during visual search can provide valuable information about people's cognitive processes. This research investigates, with detailed eye movement data analysis and computational cognitive modeling, the perceptual, strategic, and oculomotor processes people use to visually search. A cognitive model is evolved in a principled manner based on eye movement data, past modeling efforts, and recent psychological literature. We start with a base level model with a random strategy and develop it into a robust and reusable model with a flexible strategy that, while currently focused on the visual search of text, could work with a wide range of visual stimuli. Re-usable, parsimonious, local strategies interact with perceptual-motor constraints to predict the bulk of the eye movement data, including aspects of the data that appear to require task-specific global strategies in addition to fixation-to-fixation local strategies.

Visual Search Strategies

The visual search strategies people employ have a substantial effect on the time it takes people to find a target in a visual layout.

Visual search strategies can be:

- Local – motivated by properties of the individual objects being searched. For example, Shen, Reingold, and Pomplun(2003) found that people tend to shift their visual search strategy very quickly based on which visual feature is most informative for a given layout.
- Global – motivated by actual or expected properties of the layout. For example, Hornof (2005) found that people employ different visual search strategies based on the presence or absence of a reliable visual menu hierarchy.

Computational Cognitive Modeling

One way to better understand what visual search strategies people use, and why they use them, is through computational cognitive modeling.

Most computational models of visual search account for one or two of the perceptual, strategic, or oculomotor processes involved in visual search, but not all three. Ideally, a model of visual search would explain some aspect of each process involved in visual search.

The purpose of the present research is to understand and predict how people integrate perceptual, strategic and motor processes in visual search by developing a computational cognitive model that explains detailed eye movement of visual search with sufficiently few task-specific constraints.

EPIC

A series of computational cognitive models described in this study were built using the EPIC (Executive Process Interactive Control) cognitive architecture (Kieras & Meyer, 1997).

EPIC captures human perceptual, cognitive, and motor processing constraints in a computational framework that is used to build cognitive models.
Figure 1 shows a representation of EPIC.

The EPIC Cognitive Architecture

Figure 1. An overview of the EPIC cognitive architecture by Kieras and Meyer (1997). On the left, the simulated task environment, including the simulated input and output devices, and the flow of data among the devices. On the right, the various simulated sensory and motor organs, processors (ovals), memories (rectangles), and the flow of information among the various components.

The Task and Previous Modeling

This research builds directly on previous research of menu search (Hornof, 2004).

The task studied in the current research is the visual search of structured, menu-like layouts without a visual hierarchy. Figure 2 shows a sample layout from the experiment.

Sixteen participants searched four different screen layouts for a precued target object. Each layout contained one, two, four, or six groups. Each group contained five objects. The groups always appeared at the same physical locations on the screen. One-group layouts used group A. Two-group layouts used groups A and B. Four-group layouts used groups A through D.

Previous modeling accounted for the reaction time and a fair number of eye movement measures (see Hornof & Halverson, 2003).

However, the previous modeling resulted in models that were fairly brittle as the models were somewhat tuned to aspects of the visual layout used.

Figure 2. A 6-group layout. The precue, in the top left, would disappear when the layout appeared. The gray text did not appear during the experiment.

Evolving the Cognitive Strategy

The motivation for this research is the need for a computational model that is flexible enough to predict performance on a variety of menu-like visual layouts, and that can explicitly account for a wider range of eye movement measures than previous models.

This work shows some steps in the principled evolution of a model of visual search.

The model is gradually improved based on "low-level" eye movement data (for example, fixation duration and saccade distances).

Basic visual search research or previous computational modeling motivated each change.

Step 1: Start with the baseline model

The Step 1 model is the result of finding the common elements of two previous models of visual search: saccade destinations were selected at random and without replacement.

In addition:
Saccades were initiated after the fixated objects were identified.
EPIC's oculomotor feature preparation time was changed to zero ms.

The Step 1 model predicted only one observed eye movement metric quite well, namely mean fixation duration. Figure 3 shows the predicted and observed fixation durations by layout size.

Figure 3. Fixation duration observed (solid line) and predicted by the Step 1 model (dashed line). AAE = 7.8%.

Step 2: Refine the saccade destinations

Step 2 worked to improve the prediction of saccade destinations.

Two metrics were used to measure saccade destinations: mean saccade distance and mean fixations per group.

When searching, people are more likely to saccade to objects that are relatively nearby rather than objects across the layout. The selection of saccade destinations is largely based on proximity to the center of fixation (see, for example, Motter & Belky, 1998).

Direct visual inspection of hundreds of individual eye movements made by participants revealed once participants had finished dwelling on a group, they tended not to revisit that group until the remainder of the layout had been searched (this was true 94% of the time).

The Step 2 model was modified to select saccade destinations based on EPIC's visual property eccentricity as follows:

(a) After each saccade, the eccentricity property is updated based on the new eye position.
(b) The eccentricity property is scaled by a normally distributed fluctuation factor (M = 1; SD = 0.3). This scaling factor is sampled for each object after each saccade.
(c) Objects whose text has not been identified and that were in unvisited groups are marked as potential candidates for the saccade destination.
(d) The candidate object with the lowest eccentricity property, after the scaling factor is applied, is selected as the next saccade destination.

The Step 2 model predicted the saccade distance and mean fixations per group well. Figure 4 and 5 show the predictions of the Step 1 and Step 2 models, and observed results.

Figure 4. Saccade distance observed (circle), predicted by the Step 1 model (squares), and predicted by the Step 2 model (triangles). Step 1 AAE = 112%, Step 2 AAE = 5.8%.

Figure 5. Fixations per group observed (circles), predicted by the Step 1 model (squares), and predicted by the Step 2 model (triangles). Step 1 AAE = 42%, Step 2 AAE = 4.6%.

Step 3: Account for whole-task performance

One goal of this research is to produce a model that accounts for multiple eye movement measures at multiple scales.

The number of fixations per trial. as a function of the number of groups in the layout, was the first metric used to evaluate whole-task performance.

It was observed that the participants sometimes looked at or near the target but continued to search. This suggests that the participants may be failing to recognize the target occasionally.

Previous research suggests that people occasionally fail to recognize fixated objects (Halverson & Hornof, 2004; Peterson, et al., 2001).

The model was modified to include a text recoding failure rate, which represents the probability that the text property of a fixated visual object will be unknown. A value of 9% provided the best fit for the number of fixations per trial shown in Figure 6.

Figure 6. Fixations per trial observed (circles), predicted by the Step 2 model (squares), and predicted by the Step 3 model (triangles). Step 2 AAE = 14.3%, Step 3 AAE = 4.2%.

Step 3A: Increase visual working memory decay

An interaction of the text recoding failure rate, visual memory persistence, and small layouts cause the model to terminate search after a couple of saccades without finding the target. This behavior was not observed with the participants. A variety of solutions were pursued, but only one was consistent with recent literature and did not worsen eye movement predictions.

Woodman, Vogel, and Luck (2001) showed that when visual working memory (VWM) is occupied, visual search remains efficient. One interpretation of these findings is that VWM decays quickly for goal-irrelevant information, like non-targets.

The model was modified by setting the perceptual store property retaining-time parameter to 50 ms (500 ms default). With this setting, the text of objects from one fixation decayed from EPIC's visual perceptual store before the text of objects from the following fixation entered the visual perceptual store.

Step 4: Add a global strategy

A model of visual search should not only account for the frequency, duration, and distance of saccades, but also the scanpaths that people use.

The frequencies of the three most commonly observed scanpaths were compared to the frequencies with which the model followed the same scanpaths. Figure 7 shows this comparison.

While the Step 3 model predicted the two most common scanpaths quite well, it under predicted the third most common scanpath.

A global component was added to the model's strategy such that the model could develop a global "preference" for scanning horizontally or vertically.

The addition of a global strategy improved the model predictions for the third most commonly observed scanpath, but resulted in an over prediction of the two most common scanpaths. In addition, as shown in Figures 8 and 9, the global strategy slightly improved the model predictions of the number of fixations per trial as a function of the target position.

The improvements made by adding the global strategy are subtle, but they add to the fidelity of the model. However, the addition of the global strategy does not improve the quantitative fit of the model substantially and the addition may be considered overfitting as the additional production rules introduce additional free parameters.

Observed:	30%	18%	11%
Step 3:	30%	21%	1%
Step 4:	36%	27%	6%

Figure 7. The most commonly observed scanpaths in six-group layouts and how often each path was taken by the participants (observed) and the models (Step 3 and 4).

Figure 8. Fixations per trial observed (solid lines) and predicted by the Step 3 model (dashed lines). AAE = 8.1%.

Figure 9. Fixations per trial observed (solid lines) and predicted by the Step 4 model (dashed lines). AAE = 6.5%.

Conclusions

A flexible and reusable model of visual search was developed that accounts for a wide variety of search data.

The final model explains the observed:
(a) saccade distances
(b) number of fixations to each group in a layout
(c) total number of fixations in a trial
(d) number of fixations to find an object based on the object's location in the layout
(e) fixation duration
(f) scanpaths that people used

The strategy is not tuned to the visual layout of the task. This is a strength of the model.

The current research, and the similar findings in past research, suggests that future modeling of menu-like search tasks should:
(a) select relevant saccade destinations based on noisy eccentricity
(b) use a text recoding failure rate of around 9%

While it seems clear from the scanpath data that some sort of global strategy is being employed, adding a global strategy to the model added complexity (free parameters) with little gain in accuracy (i.e. this may be overfitting).

The integration of recent, relevant psychological phenomena (Motter & Belky, 1998; Woodman, Vogel, & Luck, 2001) benefits the continued integrative development of computational models and advances in basic psychological research, and thus for Cognitive Science in general.

Future Work

Work is in progress to investigate the effects of grouping on visual search.

Specifically, we are investigating the effects of common region and potential interactions with semantic consistency of grouped words in common regions.

This future work will be used to improve and extend the models developed in the research presented here.

Acknowledgments

This research was supported by the Office of Naval Research and the National Science Foundation.

References

Hornof, A. J. (2004). Cognitive Strategies for the Visual Search of Hierarchical Computer Displays. Human-Computer Interaction, 19(3), 183-223.

Hornof, A. J., & Halverson, T. (2003). Cognitive strategies and eye movements for searching hierarchical computer displays. Paper presented at the Proceedings of the Conference on Human Factors in Computing Systems, Ft. Lauderdale, FL.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. Human-Computer Interaction, 12(4), 391-438.

Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. Vision Research, 38(12), 1905-1815.

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. Psychological Science, 12(4), 287-292.

Shen, J., Reingold, E. M., & Pomplun, M. (2003). Guidance of eye movements during conjunctive visual search: The distractor-ratio effect. Canadian Journal of Experimental Psychology, 57(2), 76-96.