Evaluating Variable Resolution Displays with Visual Search: Task Performance and Eye Movements

Derrick Parkhurst
The Department of Psychology and
The Zanvyl Krieger Mind/Brain Institute

Eugenio Culurciello
The Department of Electrical and Computer Engineering and
The Zanvyl Krieger Mind/Brain Institute

Ernst Niebur
The Department of Neuroscience and
The Zanvyl Krieger Mind/Brain Institute

The Johns Hopkins University, Baltimore, Maryland
{*derrick.parkhurst* | *euge* | *niebur*}@jhu.edu

Abstract

Gaze-contingent variable resolution display techniques allocate computational resources for image generation preferentially to the area around the center of gaze where visual sensitivity to detail is the greatest. Although these techniques are computationally efficient, their behavioral consequences with realistic tasks and materials are not well understood. The behavior of human observers performing visual search of natural scenes using gaze-contingent variable resolution displays is examined. A two-region display was used where a high-resolution region was centered on the instantaneous center of gaze, and the surrounding region was presented in a lower resolution. The radius of the central high-resolution region was varied from 1 to 15 degrees while the total amount of computational resources required to generate the visual display was kept constant. Measures of reaction time, accuracy, and fixation duration suggest that task performance is comparable to that seen for uniform resolution displays when the central region size is approximately 5 degrees.

CR Categories: I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction Techniques;

Keywords: Variable Resolution Displays, Visual Search, Virtual Reality, Eye Movements

1 Introduction

Vision is the dominant modality for the acquisition of perceptual information in humans and the quality of most visual displays is determined by the available spatial and temporal resolution. The level of detail that can be rendered in real time is essentially limited by the available processing power (e.g. in virtual reality applications) and communication bandwidth (e.g. in Internet image transmission application). In light of these restrictions, it is important to allocate resources efficiently. Presenting a uniform level of visual detail across the whole display wastes resources since the human visual system does not process all information at the same spatial resolution, but rather focuses processing near the center of gaze. This aspect of the human visual system can be exploited to minimize the resource requirements by using gaze-contingent variable resolution displays that render a high degree of visual detail only around the center of gaze. Given that the visual system is a highly nonlinear adaptive system, it is important that the behavioral consequences of such manipulations be understood thoroughly. In this research, we examine the behavioral consequences of strategies adopted by human viewers when the distribution of visual detail is linked in real time to the center of gaze.

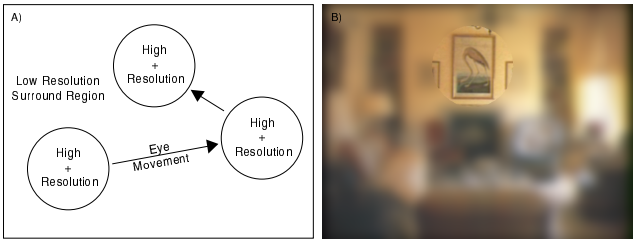

Figure 1: Variable resolution display. A) The region of high resolution tracks the observer's point of gaze in real-time. The remainder of the image is presented in a lower resolution. B) An example variable resolution display used in the experiment.

The approach we take to evaluate these displays is guided by three principles. First, variable resolution displays can explicitly take advantage of quantitative measures of visual system sensitivity by presenting only as much visual detail at a given eccentricity as can be processed at that eccentricity. Presenting more detail, as is done with traditional uniform resolution displays wastes computational resources. The obvious trade-off is that the location of the center of gaze has to be determined in real time. Although variable resolution display techniques need not utilize visual system

[1] Presented at:
The Eye Tracking Research and Applications Symposium
November 6-8, 2000
Palm Beach Gardens, Fl, USA
Pages 105-109

sensitivity to determine display parameters, we use quantitative behavioral measures of visual sensitivity to guide the way in which visual detail is distributed in a scene (for a similar approach see reference [2]).

Second, a simple two-region approximation (see figure 1) is used in lieu of matching resolution to visual system sensitivity everywhere in a display because doing so would incur a substantial computational cost in generating such complex displays. The two-region approximation consists of a central high-resolution region centered on the viewer's point of gaze surrounded by a low-resolution region. Although the resolutions of these regions can be based on visual system sensitivity, this still leaves undetermined the size of the central region. In this study, we examine the different behavioral consequences associated with variable resolution displays that use a range of central region sizes. Intuitively, having a large central region of high resolution would always be advantageous compared to a smaller region of high resolution, everything else being equal. Such a comparison is unfair because the display with the larger central region would require more computational resources. To make the comparison fair and allow generalization of our results to applications, each condition will be required to use the same total amount of computational resources to generate the display. By requiring the use of constant computation resources for each condition, the price paid for a large high-resolution central region is a very low resolution surround and the benefit of using a smaller high-resolution region is that more resources can be dedicated to maintain high resolution in the surround. In an applied setting, the primary goal is to improve the perceptual quality of the display given the resources of the system which may already be specified because the system has to work with the available hardware.

Finally, for variable resolution techniques to be widely adopted, it is important that the behavioral consequences of using variable resolution displays are well-understood even when the displays are generated with ordinary computer hardware and eye tracking equipment. Much of the work examining variable resolution displays has utilized expensive high-end hardware (e.g., see reference [1]). In this study we utilize a standard 400mhz Pentium-based computer and a low-end 60hz ISCAN video-based pupil tracking system with a combined retail price of approximately $10,000 (August 2000).

In the following experiment, a model virtual reality environment is used as a testbed to study the behavioral consequences of variable resolution display techniques. The environment is simplified in the sense that we present participants with static scenes of home interiors and that there is no possibility for interaction between the viewer and the environment. A visual search task was chosen, consisting in the search for targets in a limited variety of sizes and easily identifiable shapes. The visual search targets, medium sized bowls, were on the one hand sufficiently varied in our image database to avoid pop-out effects. On the other hand, they were sufficiently stereotypic that a target could be identified without ambiguity. The task required participants to actively search the image by making a series of eye movements. It should be realized that in choosing this task, we are purposely examining a situation where the results are *not* obvious. For instance, the targets we selected typically subtended 0.5°, and therefore a higher resolution at the point of gaze might be advantageous to identifying those targets, once foveated. On the other hand, in a visual search task where peripheral information could be used to guide eye movements, reallocating scene detail towards the center of gaze may actually be disadvantageous. By studying a task in which performance could potentially be degraded by a variable resolution display scheme this technique can be stringently evalutated.

2 Experimental Design

We use a two-region variable resolution display with a range of different resolution weighting schemes. A circular high-resolution region of radius r tracks and is centered on the viewer's point of gaze. The remaining area surrounding the central region is presented in a lower resolution. Across the different experimental conditions, the display parameter that is explicitly varied is the radius of the central region. The radius takes on values from $r = 1.25^\circ$ to $r = 15^\circ$. The resolution of the central region and the resolution of the surrounding region are subsequently determined by a set of constraints. The first constraint requires that all the resolution weighting schemes use an equal amount of computational resources. That is to say, the total computational effort required to generate the visual display is equal for each of the schemes. The way in which this constraint is implemented is described below in Section 2.1. The second constraint requires that no computational resources are wasted by presenting more visual detail (i.e. a higher resolution) at a given eccentricity than can be processed by the human visual system at that eccentricity. For our two-region display, this is accomplished by maintaining a resolution across the entire central region which is no greater than that which can be resolved at the border of the central area at radius r. The resolution of the surrounding region is always maintained lower than that of the central region and is a function of the amount of resources that remain after the central region has been painted. The way in which this is accomplished is described below in Section 2.3.

2.1 Constant Computational Resources

We start from the premise that a fixed amount C of computational resources is available and that these resources are divided into two parts. One part (C_C) is devoted to generate the central region around the center of gaze, and the remaining part (C_S) is used to generate the surrounding region in the periphery:

$$C = C_C + C_S. \tag{1}$$

A simple measure of the computational cost C is the number of elementary features to be painted in every frame, computed as the area s of the displayed region multiplied by the square of its linear resolution R:

$$C = s \cdot R^2. \tag{2}$$

In this study, we define the linear resolution as the highest spatial frequency displayed which is a simple but realistic measure of the computational cost for bitmapped raster operations. For 3D graphic engines, replacing R^2 with a measure in terms of polygons/area would yield a closer approximation of the actual computational cost and our methods can be applied to this measure immediately. We note that equation 2 has the advantage of being independent of hardware details.

2.2 Sensitivity of the Human Visual System

There are numerous measures for the sensitivity of the visual system, defined in anatomical and physiological terms as well as through various functionally defined psychophysical measures (e.g. detection or discrimination of sinusoidal gratings or of letters of the alphabet). We selected the well-studied spatial frequency transfer function of the visual system. Virsu and Rovamo [6] determined the 50% detection threshold at which sinusoidal gratings extending 5° could be detected for a range of eccentricities and spatial frequencies. Although we are interested in using this measure to guide the parameters of different resolution weighting schemes, these data do not necessarily generalize to naturalistic viewing conditions used in typical virtual reality environments. Therefore, we conservatively

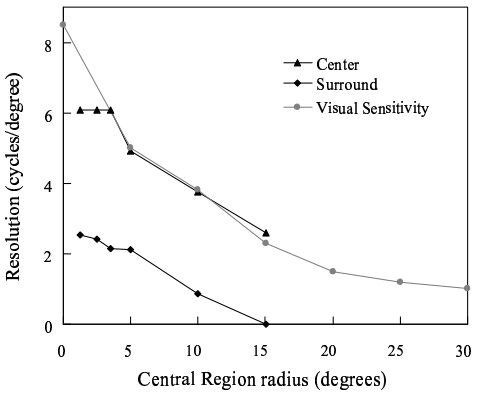

Figure 2: The derived estimate of visual sensitivity $\nu(x)$ (circles), the spatial frequency cut-off for the center high resolution region (triangles), and the spatial frequency cut-off for the surrounding low resolution region (diamonds) are shown as a function of the central region radius r.

adapt this measure of visual sensitivity by computing the "optimal" spatial frequency at a given eccentricity, defined as the frequency with the lowest contrast threshold [6]. We then increase the threshold contrast at this frequency by 3dB and determine the highest spatial frequency ν that can just be resolved at this contrast. The limiting spatial frequency $\nu(x)$ was then taken as an estimate of visual system sensitivity and used as the cut-off frequency for image presentation at eccentricity x. Although the details of this procedure are not critical, it assures that $\nu(x)$ is higher than that given by the threshold frequency at eccentricity x, resulting in a generous estimate of visual sensitivity. The threshold data obtained from this procedure are plotted in figure 2 (circles). For practical purposes, we approximate the visual sensitivity function by linear interpolation between the extracted data points.

2.3 Resolution Weighting Schemes

To determine the resolution weighting schemes the following procedure was used. First, the resolution of the central region ν_c was set to the visual sensitivity at the border between the central and peripheral regions $\nu(r)$. Exceptions were made for the smallest radii, $r = 1.25^\circ$ and $r = 2.5^\circ$ where the resolution of the central region was maintained at the highest resolution available in the original digitized images (6.09 cycles/degree) since any resources gained by painting this small area in lower resolution would be negligible and not lead to any visible improvement in the representation of the periphery. This method assures that no computational resources were wasted by displaying too much visual detail in the central region. The resolution of the surrounding region ν_s was determined by the constant computational resource constraint described by equations 1 and 2;

$$\nu_s = \sqrt{\frac{C - (s \cdot \nu_c^2)}{(S - s)}} \tag{3}$$

where S represents the total area of the display (30°x22.4°) and $s = \pi r^2$ represents the area of the central region. The total amount of computational resources C, which determines the overall level of scene detail, was chosen to be low enough that a significant visual difference in resolution between the central and surround regions was obtained in each condition (otherwise, results would be trivial). For the largest central region $r = 15^\circ$, which encompasses most of the display, all resources were allocated to generate the highest resolution possible in the center, and no resources were allocated to the surround which was therefore represented as a uniform area with correct average luminance and hue. From equations 1 and 2 the resolution in this central region ν_c is given by

$$\nu_c = \sqrt{C/s} \tag{4}$$

and results in a resolution only slightly higher than indicated by the visual sensitivity function. As can be seen in figure 2, the resolutions ν_c and ν_s of each weighting scheme for the most part remain below the derived visual sensitivity estimate ν. Therefore it was rare that more visual detail was being presented than the visual system could process.

3 Experimental Methods

Five Johns Hopkins students (3 female) were paid for participation in the experiment. All participants had normal or corrected-to-normal vision and all were naive with respect to the purpose of the study. Participants were seated a normal viewing distance (58cm) in front of a standard 17 inch computer screen that was used for stimulus presentation. The screen subtended 30.0° of visual angle horizontally and 22.4° vertically. Stimuli were preprocessed and displayed in variable resolution by a computer that also recorded the responses of the participants. Stimuli were 100 color images of size 640×480 pixels showing photographs of home interiors that were scanned in from interior design catalogs. In order to generate images or parts of images with a desired resolution, low-pass filtering was performed digitally by convolving the image with a Gaussian filter, yielding an attenuation of at least 70dB at and above the cut-off frequency.

3.1 Procedure

Participants were instructed to search the displayed images for the presence of medium-sized bowls. To familiarize participants with the targets, several examples of potential targets and non-targets (e.g., vases) were shown to each participant in a set of example images. Target bowls subtended, on average, 0.5 degrees of visual angle. All of the images contained at least one target.

To begin each trial the participant was required to fixate a cross in the center of the screen and press a mouse button. The image was subsequently presented on the screen until the participant responded, or until 20 seconds had elapsed. Participants were instructed to find a bowl, look at it, and respond immediately by clicking a mouse button. Note that the eye position, not the mouse cursor position, was used to indicate the location of the target. It was repeatedly stressed to the participants that accuracy was the most important quality of the response expected of them.

Over the course of the experiment, each participant was presented with seven versions of each image. One of the versions was presented at uniform full resolution. For the other six versions, the image was presented in low-resolution in the instantaneous visual periphery of the participant and in high resolution around the center of gaze. Each participant completed a total of 700 trials, broken into 14 blocks of 50 images each. All images were shown in random order with the exception that the any given image could not be repeated for ten trials. In order to familiarize participants with the experimental setup and eye tracker, 16 practice trials with feedback were conducted; half with and half without eye tracking. At the beginning of each block the eye tracker was recalibrated. The calibration phase consisted of a series of 9 fixation crosses that the participants were required to sequentially fixate. At the end of each block, an eye tracking error measurement was taken by having the participants fixate six randomly positioned crosses.

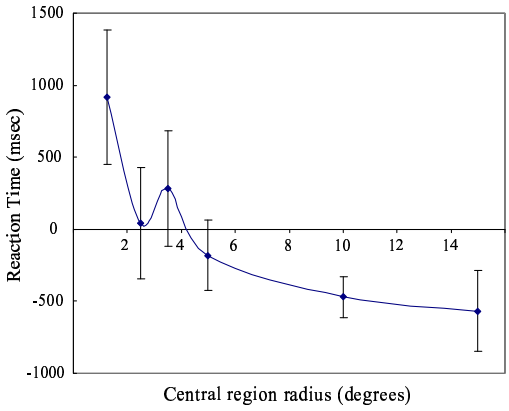

Figure 3: Reaction times are normalized to the uniformly high-resolution baseline condition and averaged across participants. Values greater than zero indicate reaction times slower than baseline and values less than zero indicate reaction times faster than baseline.

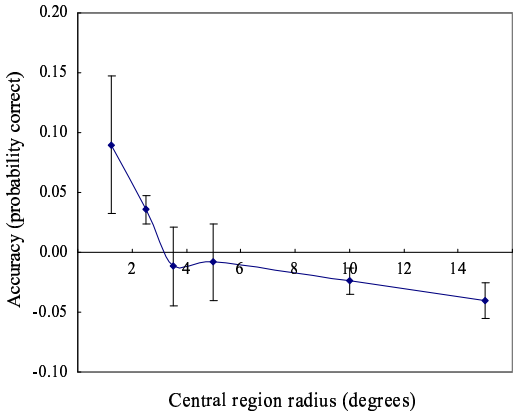

Figure 4: Accuracy is normalized to the uniformly high-resolution baseline condition and averaged across participants. Values greater than zero indicate an accuracy higher than baseline and values less than zero indicate an accuracy lower than baseline.

3.2 Eye Tracking

An ISCAN model RK-416 eye tracker was used to monitor eye position. This model is a real time digital image processor that tracks the center of the participant's pupil and measures its size from an infrared video image of the participant's eye. The unit automatically computes the position of the pupil over the two-dimensional matrix of the eye imaging camera. Pupil coordinates and diameter are computed at a rate of 60Hz, the same rate at which the adaptive video display was updated in the experiment. A custom-built head restraint and chin rest was employed to minimize the effects of head movements. A bi-cubic nonlinear interpolation (cubic in both horizontal and vertical dimensions) between a grid of nine calibration points was used to calibrate the eye tracker [5]. This procedure helped to minimize errors from non-linearities due to infrared source reflections. Additionally, the calibration was adjusted using a procedure where an eye sample from the fixation point at the beginning of each trial shifted the original interpolation. Full recalibration and adjustment of the eye tracker was intermittently required during a block of trials in the case of excessive head movements.

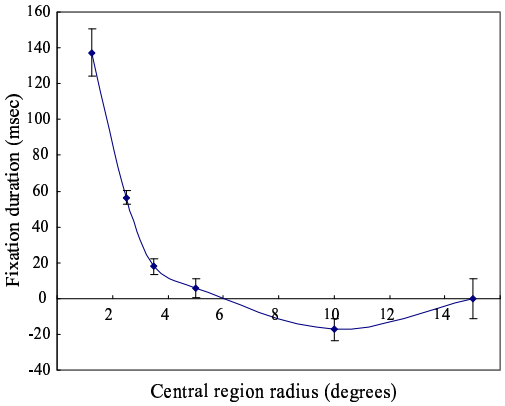

Figure 5: Fixation durations are normalized to the uniformly high-resolution baseline condition and averaged across participants. Values greater than zero indicate longer fixations than baseline and values less than zero indicate shorter fixations than baseline.

4 Results

For ease of interpretation, all dependent measures have been normalized to the baseline, defined as the uniformly full-resolution condition, on a participant-by-participant basis before averaging across participants. That is to say that the values obtained for each dependent measure at each radius r had the value obtained in the baseline condition subtracted. The resulting measure thus indicates how a particular condition differs from the baseline condition. Positive numbers indicate that the observed value was greater than in the baseline condition, and negative numbers indicate that the observed value was smaller than in the baseline condition. All normalized data are presented in the figures as group averages with the error bars representing plus and minus one standard error of the mean.

4.1 Task Performance

Figure 3 shows mean reaction times normalized relative to the baseline. Presumably due to the fact that accuracy was stressed throughout the experiment, overall reaction times were slow (mean=5043ms). Reaction times were slowest for small central regions and fastest for large central regions. A one-way repeated measures analysis of variance on reaction times was performed with central region size as the relevant factor. A significant main effect was observed ($F(5, 20) = 8.92, p < .001$). Note that reaction times for central region sizes in the range of 2.5 to 5 degrees are approximately normal (i.e. not different from the full-resolution baseline condition).

Task accuracy is shown in Figure 4 as the probability of obtaining correct responses normalized to the baseline condition. A response was scored correct if at the time the participant responded, their center of gaze fell within 1 degree of a target. A one-way repeated measures analysis of variance on accuracy was performed with central region size as the relevant factor. A significant main effect was observed ($F(5, 20) = 3.28, p < .05$). Note that accuracy for central region sizes in the range of 3.5 to 5 degrees is approximately normal.

4.2 Eye Movements

Figure 5 shows the mean duration of individual fixations in a trial. A strong decrease of the fixation duration is seen with increasing central region size, with the exception of the central region size of 15 degrees. A one-way repeated measures analysis of variance on the fixation durations was performed with central region size as the relevant factor. A significant main effect was observed ($F(5, 20) = 104.36, p < .001$). Note that fixation durations at 5 and 15 degrees are approximately normal.

Errors in the eye tracker calibration tended to accumulate over the course of a block. Therefore, the error measurements (measuring the distance between the actual gaze position and the recorded gaze position) made at the end of each block probably represent the worst-case error. The average end-of-block error across all participants was 1.6 degrees.

5 Conclusions

The goal of this study was to examine the behavioral consequences of using gaze-contingent variable resolution displays. The primary finding is that reaction time and accuracy co-vary as a function of the central region size. This result is a clear indicator of a strategic speed/accuracy tradeoff [4] where participants favor speed in some conditions and accuracy in others. By examining the reaction time results in Figure 3 alongside the accuracy results in Figure 4 the similarity in functional shape can be seen. For small central region sizes, slow reaction times are accompanied by high accuracy. Conversely, for large central region sizes, fast reaction times are accompanied by low accuracy. In tasks, experimental or otherwise, participants make a decision about which of these two factors to favor. Often experimental instructions stress that participants emphasize one or the other, but experimental manipulations such as monentary payoff or stimulus frequency manipulations can also serve to bias a participant in one or the other direction [3].

Considering the present results, it is clear that one or more of the display parameters (central region size, central resolution, or peripheral resolution) caused participants to shift their bias from favoring accuracy at small central region sizes to favoring speed at larger sizes. Although it may be tempting to attribute these effects to resolution differences alone, for central region sizes 1.25, 2.5, and 3.5 degrees the resolution stays relatively constant, yet reaction times vary by approximately 1 second and accuracy varies by approximately 10 percent (see figures 3 and 4). One might also suspect that the probability of detecting a peripheral target on any given fixation as a function of central region size might account for the results. This would predict either decreasing reaction times or increasing accuracy with central region size but not decreasing reaction time *and* decreasing accuracy as is observed. Most likely, a combination of these factors influenced participants to use different strategies. Practically, it is important to note that for both reaction times and accuracy, central region sizes of 3.5 degrees and 5 degrees were not different from that observed in the full-resolution baseline condition.

A secondary finding was that fixation duration varies as a function of central region size. For small central region sizes, participants tend to spend more time examining each fixation than under normal viewing conditions. For large central regions, fixation durations tend to be closer to normal. In agreement with reaction time and accuracy, fixation duration is approximately normal with a central region size of 5 degrees. For central region sizes less than 5 degrees, the substantial increase in fixation durations may have been due to the limited accuracy of the eye tracker. On fixations where the central region became mis-aligned with the point of gaze and the lower resolution of the surround was actually presented at the point of gaze, an increased fixation duration may have been required to determine if that location contained a target. We suspect that the effects of eye tracking inaccuracies (worst-case error of 1.6°) disappear with a central region radius of 5 degrees or more.

Overall, these results indicate that approximately normal visual search behavior can be obtained for task performance and eye movement measures when observers use gaze-contingent variable resolution displays with a central region size of approximately 5 degrees. Therefore, we conclude that variable resolution displays can save computational resources without significant behavioral consequences.

6 Acknowledgements

We thank Tim Horiuchi for helpful suggestions, Klinton Law for aid in the programming and setup of the eye tracking equipment and Bartlett Mel for providing the original set of images. This research was supported by an NSF CAREER grant and an Alfred P. Sloan Fellowship to EN as well as a NIH-NEI visual neuroscience training fellowship to DP.

References

[1] R. Danforth, A. T. Duchowski, R. Geist, and E. McAliley. A platform fore gaze-contingent virtual environments. In A. Butz, A. Kruger, and P. Olivier, editors, *Smart Graphics Symposium*, pages 66–70. AIII, Menlo Park, CA, 2000.

[2] A. T. Duchowski. Acuity-matching resolution degredation through wavelet coefficient scaling. *IEEE Transactions on Image Processing*, 9(7):1437–1440, 2000.

[3] D. M. Green and J. A. Swets. *Signal Detection Theory and Psychophysics.* Wiley, New York, N.Y., 1966.

[4] A.V. Reed. Speed-accuracy trade-off in recognition memory. *Science*, pages 574–576, 1973.

[5] D. M. Stampe. Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, and Computers*, 25(2):137–142, 1993.

[6] V. Virsu and J. Rovamo. Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research*, 37(3):475–494, 1979.