

PERCEPTUAL EFFECTS OF A GAZE-CONTINGENT MULTI-RESOLUTION DISPLAY BASED ON A MODEL OF VISUAL SENSITIVITY?

Lester C. Loschky
George W. McConkie

University of Illinois at Urbana-Champaign
405 N. Mathews
Urbana, IL, USA

Jian Yang

Michael E. Miller

Eastman Kodak Company
1700 Dewey Avenue
Rochester, NY, USA

ABSTRACT

Many interactive single-user image display applications have prohibitively large bandwidth requirements. However, bandwidth can be greatly reduced by using gaze-contingent multi-resolution displays (GCMRDs) that put high-resolution only at the center of vision based on eye position. A study is described in which photographic GCMRD images were filtered as a function of contrast, spatial frequency, and retinal eccentricity, on the basis of a model of visual sensitivity. This model has previously only been tested using sinusoidal grating patches. The current study measured viewers' image quality judgments and their eye movement parameters, and found that photographic images filtered at a level predicted to be at or below perceptual threshold produced results statistically indistinguishable from that of a full high-resolution display.

INTRODUCTION

Many interactive computer applications currently under development require large displays with high-resolution and fast update rates. For example, aircraft simulators often require high enough resolution to identify aircraft from several miles away, and large enough displays to provide over 200° of visual angle, which is necessary for certain aerial maneuvers. Video teleconferencing and telemedicine generally use much smaller displays, but require high-resolution and extremely fast update rates to avoid communication and manual manipulation errors. Remote piloting applications require high-resolution to identify terrain, large displays to provide accurate motion perception, and fast update rates. Unfortunately, in these and similar applications, such image processing demands outstrip the system's computational power and bandwidth limitations and are therefore impossible to achieve. Because many of these applications are meant for single users, a solution to this bottleneck is to reduce these requirements by utilizing gaze-contingent multi-resolution displays (GCMRDs) in which

high-resolution imagery is presented only at the location to which the gaze is directed, with lower resolution elsewhere. Such displays benefit from the variable resolution of the human visual system by varying display resolution as a function of where the viewer is looking. Doing so can result in considerable computation and bit rate savings (Niu, 1995, reported in Loschky & McConkie, 2000). Furthermore, one should be able to optimize such a system such that bandwidth savings are maximized while perceptual difficulties are minimized if the GCMRD is most closely tuned to the relevant parameters of the human visual system.

One of the best-known findings in visual science is the fact that primate visual resolution quickly degrades as one moves from the fovea into the visual periphery. This drop-off in resolution as a function of retinal eccentricity can be measured in many ways, e.g., in terms of Snellen, Landolt, or Vernier acuity. But the theoretically most complete measure of visual performance employs sinusoidal gratings. This is because any visual stimulus can be decomposed into a discrete set of sine waves and there is good evidence that the human visual system is roughly tuned to spatial resolution channels (Graham, 1992). These channels appear to be based on the structure of the retina, and perhaps also the lateral geniculate nucleus and primary visual cortex. Specifically, in the retina, there is a steep drop-off in the density of cones with distance from the center of the fovea, and the ratio of cones to retinal ganglion cells is greater in the visual periphery. Thus, spatial sampling becomes ever more coarse with distance from the center of vision.

A number of studies have investigated the relationship between sensitivity to visual contrast as a function of both spatial frequency and retinal eccentricity (Cannon, 1985; Pointer & Hess, 1989; Robson & Graham, 1981), using grating stimuli in static acuity tasks. These have shown that contrast sensitivity exponentially decreases to zero for higher spatial frequencies as retinal eccentricity increases. Thus, the highest resolvable spatial frequency of a grating decreases non-linearly and rapidly with retinal eccentricity.

This aspect of human visual sensitivity has been modeled by many authors (e.g., Geisler & Perry, 1998; Peli, Yang, & Goldstein, 1991), with one such model producing the contrast thresholds shown in Figure 1. Such findings have been well replicated using simple grating stimuli, acuity measures, and static viewing conditions. However, only a few studies have shown that such findings actually scale up to more natural stimuli, such as photographs, measures, such as image quality judgments or eye movement parameters, and viewing conditions, such as GCMRDs (see Loschky & McConkie, 2000 for review).

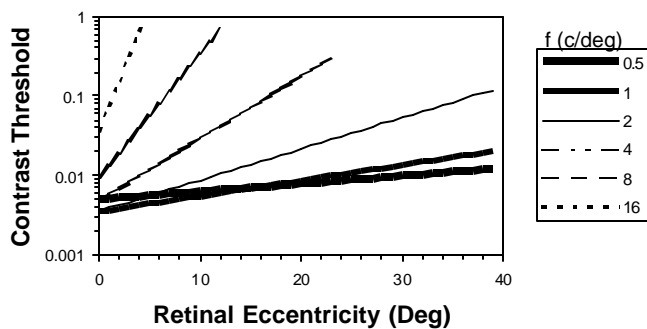


Figure 1. Contrast threshold versus retinal eccentricity. The visual stimuli were grating patches with spatial frequencies ranging from 0.5 to 16 cpd.

Previous studies by Loschky and McConkie have used a GCMRD to investigate spatial vision. In one study (Loschky & McConkie, 2000, exp 4), the authors had participants view digitized photographs through a bi-resolution wavelet filtered GCMRD while searching for objects or trying to memorize details of the picture. They then measured the effects of level of filtering and size of the high-resolution inset on visual search times and eye movement parameters. For saccade lengths, they found strong effects for level of filtering and high-resolution inset size as well as an interaction between these two factors. Specifically, the eyes tended to move shorter distances either as high spatial frequencies became more attenuated, or as those attenuated high spatial frequencies were brought closer to the fovea. Conversely, as the level of high spatial frequency attenuation decreased, or the attenuated high spatial frequencies were moved into the visual periphery, saccade lengths increased until there was essentially no difference between the bi-resolution GCMRD and a full high-resolution display. Similar effects were found for both fixation durations and search times, though the effect for high-resolution inset size was greater than that of level of filtering for these dependent variables.

Several questions nevertheless remain. First, while the wavelet filtering method used in the above studies provides a convenient environment for implementing GCMRDs, the precise functional relationships between spatial frequency and retinal eccentricity and the dependent measures were somewhat unclear. Wavelet filters attenuate wide bands of

spatial frequencies, but do not eliminate any frequencies altogether. Thus, a method of image filtering, which eliminates specified high spatial frequencies as a function of retinal eccentricity would allow a more detailed analysis of the effects of filtering on vision. Second, the various effects on the dependent measures mentioned above are assumed to have been due to the factors of level of filtering and high-resolution inset size. However, due to the fact that the GCMRDs were bi-resolution, it is possible that an artifact due to the discontinuity between the high and lower resolution areas in the image caused the effects. Thus, it is desirable to use a method of image filtering that produces smoothly degraded multi-resolution displays. If the various effects found earlier with the bi-resolution displays are also found with smoothly degraded multi-resolution displays, this would eliminate the discontinuity-based explanation of those effects.

Based on the above questions, we decided to replicate the previous study using a multi-resolution display in which the image filtering algorithm smoothly degraded the image, and in which the filtering eliminated specified spatial frequencies at specified retinal eccentricities based on the results of previous spatial vision experiments using grating stimuli. In addition to implicit measures of perception based on eye movement parameters (fixation durations and saccade lengths) that we had used in our previous studies of the psychophysics of GCMRDs, we added an explicit measure of the effects of image degradation, subjective image quality judgments.

METHOD

Subjects. There were 6 paid participants, 3 female, all of whom were university students, ages 19-22, and had at least 20/20 near acuity.

Stimuli. The stimuli used in the experiment were monochrome digitized photographs of various indoor and outdoor scenes from the Corel Image Database (all of which are uncopyrighted). All were relatively complex and rich in detail. The images measured 768 x 512 pixels and at 26.55" (67.44 cm) viewing distance were 18° x 12° visual angle.

To evaluate a multi-resolution display, it was important to create and display a foveated image for every possible eye fixation location in a given picture. Because the system we developed cannot generate foveated images in real time, it was necessary to pre-compute and store an image for each possible eye fixation location. These images were then retrieved and displayed to simulate a real-time system. In order to minimize image generation time, all multi-resolutional images were pre-computed. For this experiment, three hundred thirty multi-resolution versions of each picture were created within a 22 x 15 grid of locations on the 18° x 12° screen. Having these 330 versions of each picture

simultaneously stored in video memory allowed the simulation of a real-time system that could generate a multi-resolution image in which the center of each high-resolution area was separated from its closest neighbors by approximately 0.8° horizontally and vertically.

The images were filtered by an algorithm developed at Eastman Kodak Company. The algorithm is a modified version of the Geisler and Perry (1998) foveated multi-resolution pyramid. In the Geisler and Perry approach, different levels of the pyramid are circularly truncated based on the estimated cut-off spatial frequency of the visual system at different eccentricities. The reconstructed image from the zone-limited pyramid contains fine structure at the point to which the gaze is directed, and becomes more blurred towards the peripheral retina. In the modified method, one computes a contrast threshold map based on the peak frequency of the pyramid level, and uses the contrast threshold map to threshold the image content. Whenever the image contrast of a specific frequency band is below the corresponding visual contrast threshold, the image contrast is thresholded. All other image content is kept unchanged. In this way the foveated zones are not circular, but have somewhat irregular boundaries. For this reason, this method avoids the hard boundaries that are inherent in the Geisler and Perry approach.

In the study, we used 5 levels of filtering, one of which, level 2, was predicted to match the human visual system's location-specific threshold distribution, one predicted to be below threshold, and three levels predicted to be above threshold. The curves representing these 5 filter cut-off levels are represented in Figure 2. An additional, full high-resolution control condition was included as a baseline comparison. Example images in various conditions are shown in Figures 3-5.



Figure 3. Full high-resolution control Image.



Figure 4. Image with filtering level 2 ($a = 3.11$ deg.)



Figure 5. Image with filtering level 5 ($a = 0.39$ deg.)

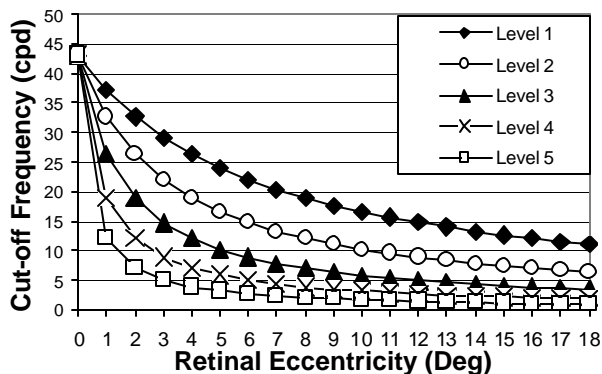


Figure 2. Cut-off spatial frequency as a function of retinal eccentricity and filtering level for the five filtering conditions in the study. Contrast is set to 100%.

We describe the degradation level by defining the cut-off spatial frequency of the filter employed by the algorithm at different eccentricities. The cut-off frequency is defined here using the corresponding frequency that produces a threshold contrast of 100% at a particular eccentricity E . The cutoff frequency of the filters employed for image manipulation can be described using the equation:

$$f_c = 43.1 * a / (a + E)$$

where f_c is the cut-off frequency, E refers to the retinal eccentricity, and a is a parameter with the value of 6.22, 3.11, 1.55, 0.78, and 0.39 deg, for filtering levels 1-5 respectively (see Figure 2). In the study, there were 3 sets of six images each. In each set, the 6 conditions were counterbalanced across the 6 images in pseudorandom order across participants, such that 6 participants filled all cells in the design.

Apparatus. Eye movements were monitored using a Dual Purkinje Generation 5 eyetracker, sampling at a rate of 1,000 Hz (i.e., once per millisecond), with a spatial resolution of approximately $1/16^\circ$ visual angle. The eyetracking signal was parsed into saccades and fixations on-line, using a velocity-based algorithm. In GCMRD trials, when a saccade was detected, the program looked for the end of the saccade (defined as the peak of the 'overshoot' of the eye movement; see Deubel & Bridgeman, 1995), and within 5 milliseconds (ms) initiated the replacement of the currently displayed multi-resolution image with a new one centered approximately at the new fixation point. All images were stored on a ViewGraphics model VS6000 video memory and display controller. The multi-resolution version of the image whose center of high-resolution corresponded to the grid point closest to the current eye position was then displayed on a ViewSonic P815 monitor. The monitor operated with a 60 Hz refresh rate, though the display of a new image could be initiated at any point in the refresh cycle.

PROCEDURES

In the experiment, participants were given two tasks. The primary task was to remember the contents of the pictures in order to prepare for a recognition memory test. The memory task had a learning phase and a test phase. In the learning phase, participants viewed each picture in a set for 20 seconds. In the test phase, they viewed the same set of images again for 20 seconds each, but some of the pictures were changed. Changes included having an object moved, added or deleted, two objects' positions switched, or the entire image mirror reversed (a horizontal flip). After seeing each image for a second time, participants were queried as to whether it was exactly the same or in any way different from the first presentation. Unknown to the participants, the sole purpose of the memory task was to provide them with motivation to carefully examine each image for a total of 40 seconds and, in doing so, to make numerous eye movements. Immediately following the offset of the image, and prior to the memory query during the test phase, participants carried out their secondary task. This was to rate the quality of each image on a Likert scale of 1-9, with 1 = worst, and 9 = best. This task provided an explicit measure of the perceptibility of the degree of filtering of an image. Prior to carrying out these tasks in the actual experiment, participants first read detailed written instructions explaining the task, and then carried out both

tasks on a practice set of 6 images. During practice, participants received feedback from the experimenter as to their performance and asked questions about the tasks. After practice, participants took a 40-minute break and then received three blocks of trials with 8 images per block, 2 of which were changed during the test phase. Data from these change trials were not included in the analyses. Participants were given a short break between blocks and the eyetracker was also recalibrated before beginning each block.

RESULTS

Precursors. Participants took the memory task seriously, performing quite accurately, detecting 68% of the changed images, and with only 16% false alarms. Prior to analyzing the quality judgment data, we eliminated trials in which there were eyetracking failures (1 second or more total loss of lock time in a 20 second trial). This eliminated artifacts due to the GCMRD not following the participants' eyes. This resulted in data with a mean loss of lock time of 314 ms per trial with the modal loss of lock being 0 ms (43% of trials). There were no empty cells in the design, though several cells for individual subjects had only a single observation, and across subjects there was a range of 17-21 observations per filtering condition. Prior to analyzing the eye movement data, we eliminated all blinks, all loss of lock, and the longest 1% of fixation durations (≥ 1010 ms). This resulted in a range of 1968-2103 observations per filtering condition. All statistical tests of significance were done at the .05 level.

Quality judgments. While the quality judgment for the least degraded condition (level 1) did not differ significantly from the control condition (in fact, it was rated slightly higher), there was a main effect such that greater filtering levels led to significantly lower ratings ($F(5, 25) = 41.15, p = .000$), as can be seen in Figure 6. Both the linear and quadratic trends were significant ($F(1, 5) = 140.46, p = .000$; $F(1, 5) = 29.83, p = .003$ respectively), reflecting a steeper drop-off of ratings with increased filtering level. Furthermore, repeated contrasts indicated that only those filtering levels predicted to be above threshold produced ratings reliably different from the full high-resolution control condition ($F(1, 5) = 8.15, p = .036$; $F(1, 5) = 39.96, p = .001$; $F(1, 5) = 78.62, p = .000$, for levels 3-5 respectively).

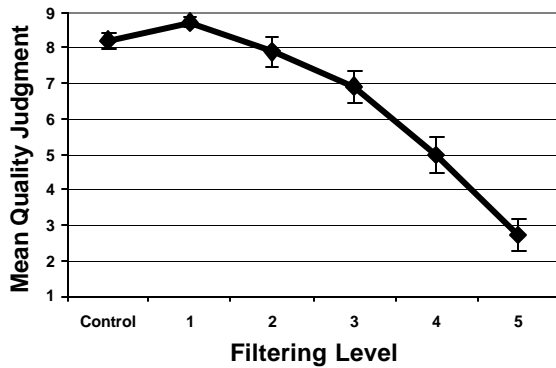


Figure 6. Mean quality judgment as a function of filtering level. Error bars represent ± 1 standard error.

Eye movement parameters. Fixation durations showed a main effect due to level of filtering ($F(2.42, 12.10)$ (Huynh-Feldt)) = 4.75, $p = .025$): as level of filtering increased, fixations became longer (see Figure 7). The data are well fit by the following power function: $y = a \cdot x^b$, where $a = 293.13840$, and $b = -0.40621$. This function has an r value of -0.93 . Thus, it appears that as filtering is increased exponentially (iteratively halving the eccentricity to halve the cut-off frequency), fixation durations increase as a power function. Importantly, repeated contrasts showed that there were no significant differences between the full high-resolution control condition and the conditions predicted to be at sub-threshold or threshold levels of filtering (levels 1 and 2). Only conditions having levels of filtering that were predicted to be above threshold produced statistically reliably longer mean fixations ($F(1, 5) = 7.052$, $p = .045$; $F(1, 5) = 18.134$, $p = .008$; $F(1, 5) = 8.232$, $p = .035$, for levels 3-5 respectively).

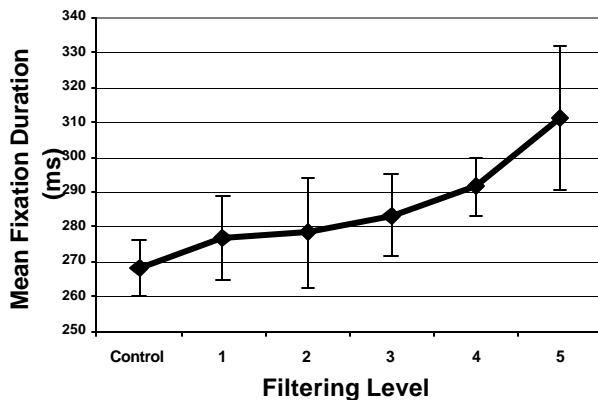


Figure 7. Mean fixation duration as a function of filtering level. Error bars represent ± 1 standard error.

As evidenced in Figure 8, similar analyses of mean saccade lengths as a function of level of filtering found a highly reliable main effect and a significant linear trend such that as the level of filtering in GCMRDs increased, saccade lengths decreased ($F(4.12, 20.62)$ (Huynh-Feldt)) = 6.59, $p = .001$; $F(1, 5) = 12.64$, $p = .016$). However, in this case the best fitting polynomial trend was a quadratic equation ($F(1, 5) =$

29.54, $p = .003$), because mean saccade lengths initially became slightly longer for filtering levels 2 and 3 and then sharply dropped for filtering levels 4 and 5. However, the only filtering condition that produced reliably different saccade lengths from the full high-resolution control condition was level 5, the highest level of filtering ($F(1, 5) = 16.76$, $p = .009$).

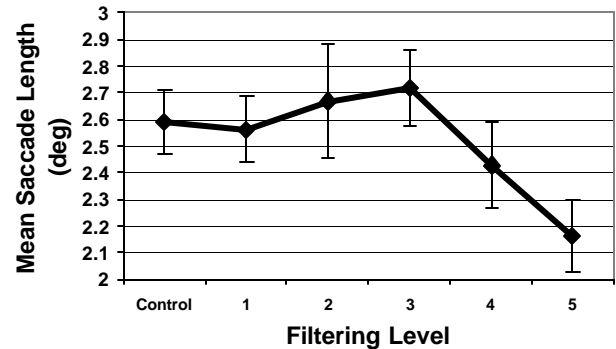


Figure 8. Mean saccade length as a function of filtering level. Error bars represent ± 1 standard error.

DISCUSSION

The above results suggest that findings from studies using simple sine wave grating stimuli, and measuring contrast sensitivity thresholds during enforced fixation do, in fact, scale up quite well to more naturalistic stimuli, tasks, and measures. Based on previous studies of contrast sensitivity, we predicted that image degradation in a GCMRD at filtering level 1 would be sub-threshold and at filtering level 2 approximately at threshold, thus producing quality judgments and eye movement results equivalent to or only slightly different from the full high-resolution control condition. Both of these predictions were borne out by our results. Likewise, we predicted that filtering levels 3, 4, and 5 would produce incrementally above-threshold image degradation that would be reflected in quality judgments and eye movements different from the control condition. Again, these predictions are strongly supported by the results.

We also predicted, based on previous results from a GCMRD with wavelet-filtered bi-resolution images, that saccade lengths would decrease as the level of filtering increased. Our explanation for this previous finding was that excessive filtering reduces the salience of potential eye movement targets in the filtered regions of the visual periphery. However, a competing explanation would be that the edge formed between the high-resolution and low-resolution areas would attract attention and thus eye movements, and this could result in the shorter mean saccade lengths found earlier. The finding of shorter mean saccade lengths in this study, in which the image is continuously filtered (i.e., with no discrete boundaries between levels) casts doubt on that hypothesis, and lends support to the decreased saliency hypothesis. However, the current results also suggest another filtering-related effect on saccade lengths may be at

work as well. That is, there was a slight trend for saccade lengths to increase in the conditions in which filtering was at predicted threshold and just above threshold levels. If this finding is not spurious, it suggests that a just barely noticeable level of image degradation may actually draw the eyes. This could be due to a tendency for the eyes to be drawn to objects that cannot be perfectly resolved at the fovea. However, as the level of resolution drops off further, those items become so degraded as to be less attractive as potential eye movement targets.

CONCLUSION

Previous studies (Niu, 1995, results reported in Loschky & McConkie, 2000) have suggested that gaze-contingent multi-resolution displays can produce large bandwidth savings in single-user applications. Furthermore, a large body of psychophysical research on spatial vision suggests that it should be possible to match the contrast sensitivity of the human visual system with image filters that take into account the contrast, spatial frequency, and retinal eccentricity of areas in a gaze-contingent multi-resolution image (e.g., Geisler & Perry, 1998; Peli, Yang, & Goldstein, 1991). The current study provides support for these predictions. Both explicit and implicit measures of perception, from subjective image quality judgments to eye fixation durations and saccade lengths, produced results consistent with predictions drawn from previous studies using grating patch stimuli. Only those levels of filtering that were predicted to be above perceptual threshold were statistically different from a full high-resolution display. Those filtering conditions predicted to be at or below threshold were virtually indistinguishable from a full high-resolution display.

We have argued above that gaze-contingent multi-resolution displays (GCMRDs) can considerably reduce requirements for computation during image reconstruction and for image transmission bit rate. The current study shows that, given a system with sufficiently precise eye tracking, a fast image update rate, and using the image filtering method we tested, one can produce a GCMRD without any negative impact or even sense of reduced image quality on the part of the human observer.

REFERENCES?

- Cannon, M. W. (1985). Perceived contrast in the fovea and periphery. *Journal of the Optical Society of America*, 2(10), 1760-1768.
- Deubel, H., & Bridgeman, B. (1995). Fourth Purkinje image signals reveal eye-lens deviations and retinal image

distortions during saccades. *Vision Research*, 35(4), 529-538.

- Geisler, W. S., & Perry, J. S. (1998). A real-time foveated multi-resolution system for low-bandwidth video communication. In B. Rogowitz & T. Pappas (Eds.), *Human Vision and Electronic Imaging* (Vol. 3299, pp. 294-305).
- Graham, N. (1992). Breaking the visual stimulus into parts. *Current Directions in Psychological Science*, 1(2), 55-61.
- Loschky, L. C., & McConkie, G. W. (2000). User Performance With Gaze Contingent Multiresolutional Displays. In A. T. Duchowski (Ed.), *Eye Tracking Research & Applications Symposium 2000* (pp. 97-103). New York, NY: ACM.
- Niu, E. L.-C. (1995). *Gaze-based video compression using wavelets*. Unpublished Master of Science, University of Illinois at Urbana-Champaign, Urbana, IL.
- Peli, E., Yang, J., & Goldstein, R. B. (1991). Image invariance with changes in size: The role of peripheral contrast thresholds. *Journal of the Optical Society of America*, 8(11), 1762-1774.
- Pointer, J. S., & Hess, R. F. (1989). The contrast sensitivity gradient across the human visual field: With emphasis on the low spatial frequency range. *Vision Research*, 29(9), 1133-1151.
- Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, 21(3), 409-418.