

Subjective Evaluation of Retinal-Dependent Image Degradations

Jian Yang, Terry Coia, and Michael E. Miller
Eastman Kodak Company
Rochester, New York

Abstract

To understand the contributions and the requirements of image resolution at different eccentricities to perceived image quality, we measured subjective image quality for degraded multiresolution images. A vision model equation was used to estimate contrast threshold as a function of retinal eccentricity. Two high-resolution images were processed using a vision-performance-based algorithm. Eight such degraded images were computed for each original scene by scaling the model-determined performance. During this experiment, the participants were asked to fixate on a target. The images were displayed for 250 ms in each trial to prevent saccadic eye movements. Five observers participated in the experiment and were instructed to rate the image quality on a ratio scale. Each image presentation was repeated 20 times in a random sequence. The average response over the five observers showed that the perceptible image degradation level was slightly higher than predicted by the model equation. This is consistent with our expectations as the visibility of the degradation should be masked by the background image. This experiment demonstrates that the vision model, which is based on visual sensitivity to simple grating patches at different eccentricities and spatial frequencies, provides a useful tool to predict the perceptible image degradation in real complex scenes as a function of retinal eccentricity.

Introduction

In situations where the field of view of the video screen needs to be large, such as in immersive entertainment, the required bandwidth for image communication can become exceedingly high. As the end-receiver is a human with a bandwidth-limited visual system, there is a potential to reduce the required communication bandwidth by filtering the image to eliminate information from the image that cannot be seen by the human visual system. If this filtering is performed prior to a bottleneck in the image communications system, this filtering can reduce the required bandwidth.

When motion imaging is concerned, the time interval in which an image frame is displayed is so short that saccadic eye movements are not possible. In this case, the human fovea is not able to scrutinize more than one location of a single frame; most parts of the image frame are seen by the

peripheral retina. As human vision is spatially inhomogeneous, with the greatest spatial and chromatic resolution at the fovea when viewing images under photopic luminance levels, the peripheral retina is less sensitive to fine image structure. Therefore, one could monitor the eye fixation location and utilize the inhomogeneous property to reduce the information content of the portion of the image that is displayed in the periphery, i.e. the so-called gaze contingent display technique.^{1,7}

To make this gaze contingent system work appropriately, there are several requirements: (1) one must understand human visual performance at different eccentricities; (2) image processing tools must be created that can appropriately implement human visual functions; (3) an eye tracking device must be able to reliably determine gaze position; and (4) fast real-time image processing must be available. In this paper, we are only concerned with the first two issues. Our primary goal is to evaluate the perceived image quality for these non-uniformly degraded images.

Peripheral Visual Performance

Psychophysically, one can measure the highest spatial frequency that can be resolved at a particular retinal location with patched gratings (e.g., Refs 8-9). For a more complete characterization, investigators usually measure the contrast thresholds for detecting the patched gratings of different spatial frequencies at different eccentricity (e.g., Refs 10-12). Based on the study of Peli et al.,¹³ the contrast threshold for detecting a patched grating of spatial frequency f at an eccentricity r can be described as

$$C_c(r, f) = C_c(0, f) \exp(k f r), \quad (1)$$

where $C_c(0, f)$ is the contrast threshold at the fovea, and k is a parameter. In their formulation of this equation, Peli et al.¹³ fit this equation to data from six previous experiments that had determined the contrast threshold of monochromatic gratings. These fits demonstrated that the k value ranged from 0.030 to 0.057. Based on Eq. 1, the contrast threshold increases rapidly with eccentricity at high spatial frequencies. These relationships indicate that the high spatial frequency information can be only retrievable by the central parts of the retina.

Equation 1 provides a relative contrast threshold at different eccentricities, when the fovea threshold is normalized to 1. This equation does not address the actual

thresholds as the fovea thresholds $C_t(0, f)$ are not specified. Geisler and Perry¹⁵ developed a similar model, with a specification at the fovea. The model was written in a different format, but it can be expressed by Eq. 1, with the fovea contrast threshold expressed as

$$C_t(0, f) = \exp(\alpha f) / S_0 \quad (2)$$

where S_0 is a constant with empirically obtained values of about 64 to 75, and α is the spatial frequency decay constant.

The fovea contrast threshold described by Eq. 2 increases monotonically with spatial frequency and it provides a nice fit to particular sets of psychophysical data. However, this equation is not adequate for a general prediction of contrast thresholds at low spatial frequencies. It is well known that contrast threshold dips at an intermediate spatial frequency of about 1 to 4 cpd, and increases at both low and high spatial frequencies.¹⁴ Yang et al.,¹⁵ developed a model to capture such a behavior, with a simplified version of the equation expressed as

$$C_t(0, f) = [N + \eta \sigma^2 / (f^2 + \sigma^2)] \exp(\alpha f) \quad (3)$$

where, N , η , σ , and α are parameters. The calculated contrast threshold versus eccentricity at six different spatial frequencies according to Eqs. 1 and 3, and a nominal set of parameter values is shown in Fig. 1.

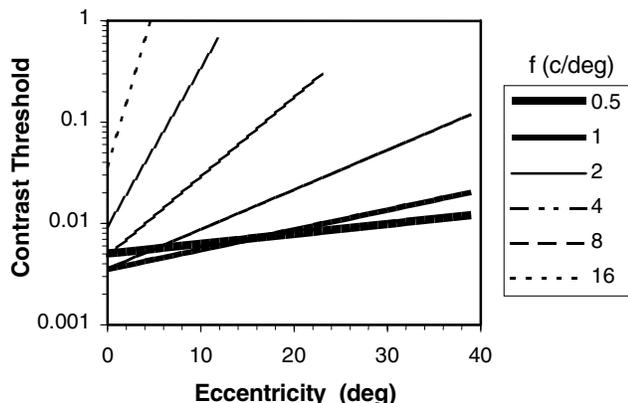


Figure 1. Contrast threshold versus eccentricity. The visual stimuli were grating patches with spatial frequency ranging from 0.5 to 16 cpd. The thresholds were calculated based on Eq. 3, with a nominal set of parameter values.

Before proceeding, it is helpful to point out some potential limitations in the model equation of the contrast threshold. The threshold equations were based on pure target detection on a uniform field. When a real image is used, the detection of the image degradation is often accomplished against a complex image background. In this case, due to the effects of visual masking, the threshold contrast could increase. Furthermore, the used model parameters were obtained under certain experimental conditions. The conditions may not be exactly the same in

other experiments, and the parameter values might change accordingly. For example, one needs to choose the field size in the measurements of the visual sensitivity at different eccentricities and different spatial frequencies. However, it is difficult to determine the appropriate target size for a generalization of the results to spatially extended patterns. For example, the difference between Eqs. 2 and 3 could be explained by the effect of field size used in the experiments. Because of the existence of these uncertainties in the model, we can only use the model to roughly estimate the contrast threshold at different eccentricities. For a more accurate description, we will rely on experimental results using spatially extended patterns.

The threshold equations can be used to estimate the degradation levels that can be just detected. To provide a range of image degradation in the study, we varied the contrast threshold Eq. 1 by transforming the eccentricity r with a shifting parameter S_1 and a scaling parameter S_2 :

$$r = S_2 (r - S_1) \quad (4)$$

When $S_1 = 0$ and $S_2 = 1$, it returns to the original form, however, scaling of these parameters allows the aggressiveness of the model to be altered and the perceived degradation be evaluated.

Processing Multiresolution Images

Whereas one purpose of creating foveated images is to transmit the least information from the original image while minimizing the perceived image degradation, it is important to understand the impact of this type of filtering on perceived image quality. After having a model of visual performance at different eccentricities, the next question is how can one process images to take advantage of the inhomogeneity of the human visual system to allow the resulting images to be matched to the information processing capacity of the human visual system.

Many methods have been proposed to achieve this goal. Some of the work simply separates the image into two resolution zones in space (e.g., Refs 1 and 7). In this method, the images displayed in a circular zone centered at the fixation position have a high spatial resolution, and the images in the surrounds only keep lower spatial frequency information. Although the two-zone method is simple and straightforward for the processing of foveated images, it does not take full advantage of the inhomogeneity of the human visual system, the sensitivity of which changes smoothly as a function of retinal eccentricity.

To optimize the foveated images, the changes in image quality with eccentricity should be matched to the smooth variation of human visual performance. Along this direction, Geri and Zeevi¹⁶ used a variable resolution scheme to process images. In their approach, the point-spread function of the human vision was modeled as a Gaussian function, with its standard deviation increasing with eccentricity. In a different approach, Kortum and Geisler² developed an algorithm to sample the original images based on human vision knowledge. Kortum and

Geisler² calculated the required sampling interval based on a cortical magnification factor. The code values within a sampling block are assigned the mean value of the block, i.e., the so-called SuperPixel.

To maintain a consistency in the description of human visual performance provided earlier, we applied the threshold Eqs. 1 and 3 to determine the required sampling interval. The cut-off spatial frequency f_c at an eccentricity r can be defined by setting the contrast threshold C_t to 1, which gives $f_c = - \{ \ln[N + \eta \sigma^2 / (f_c^2 + \sigma^2)] \} / (\alpha + k r)$, and it can be further approximated to

$$f_c = - \ln(N) / (\alpha + k r), \quad (5)$$

when the cut-off frequency f_c is much higher than σ , which is often the case. Any frequency components that are higher than the cut-off frequency are not useful to visual perception, and can be discarded. In order to carry the frequency components up to the cut-off frequency f_c , the sampling interval should not be larger than

$$\Delta x = 1 / (2f_c) = - 0.5 (\alpha + k r) / \ln(N). \quad (6)$$

Geisler and Perry⁵ further developed a foveated multiresolution pyramid to segregate an image into different spatial frequency bands. In the Geisler and Perry approach, different levels of the pyramid were 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 center of the fixation, and it gets more blurred towards the peripheral retina.

Experimental Method

The purpose of the experiment is to evaluate the image quality of the processed multiresolution images, and to understand whether the vision model, which is based on visual sensitivity to simple grating patches at different eccentricities and spatial frequencies, is useful to predict the perceptible image degradation in real complex scenes as a function of retinal eccentricity. We only consider the SuperPixel algorithm here.

Viewing Environment

The experiments were run on a Power Mac G3 computer with a 17" monitor (Mitsubishi Diamond Pro 87TXM). We developed the experimental software in Matlab using the extensions provided by the high-level Psychophysics Toolbox¹⁷ and low-level VideoToolbox¹⁸ to control video sequences on the monitor. The screen resolution was 1280 by 1024 with a frame refresh rate of 75 Hz. The output luminance on the screen was proportional to the input code value as delivered by using a linear lookup table. The maximal luminance of the screen was 89 cd/m². The viewing field extended 36 by 28.8 deg at a viewing distance of 45 cm, which gives 1.7 min of arc per pixel. The display was viewed in a darkened room.

Image Preparation

The original images, referred to as TAXI and DISNEY and are shown in Fig. 2. The images were stored in linear RGB code value with 8 bits for each color channel. We applied the SuperPixel algorithm to process the images. For each image, we produced eight images of different levels of degradation manipulated by the parameters S_1 and S_2 .

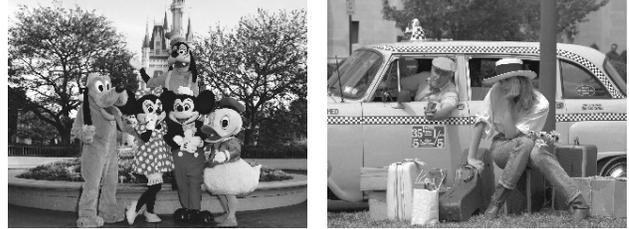


Figure 2. The original image DISNEY and TAXI

To probe a larger peripheral retina, the fixation target was located at the pixel and line position of (100, 100) in the upper-left corner of the screen. The extent of the image degradation is indicated by two parameters (S_1 , S_2), and was chosen to be in a range from less than the threshold level to well above threshold as determined by some pilot tests.

Table 1. The parameters S_1 and S_2 for the eight degraded image using SuperPixel algorithm.

	I1	I2	I3	I4	I5	I6	I7	I8
S_1	0	0	-3.7	0	-1.5	3	0.98	3.33
S_2	0.5	1	0.92	2	1.94	4.2	4	6.2

There were eight degradation levels for the SuperPixel images. The eight sets of parameters (denoted from I1 to I8) are shown in Table 1.

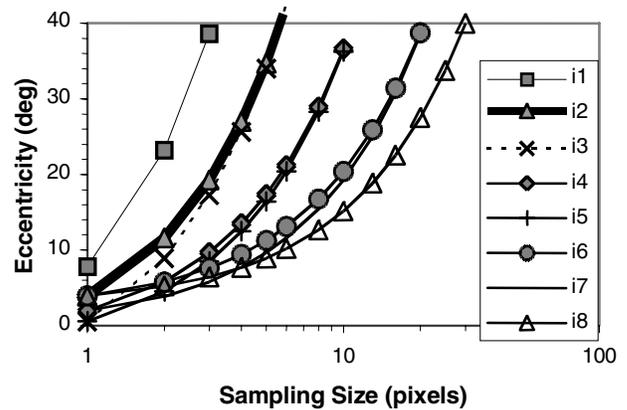


Figure 3. The relationships between required sampling size and eccentricity calculated based on Eqs. 4 and 6 with a nominal set of parameter values, and the S_1 and S_2 values supplied in Table 2. The heavy line represents the condition determined by the threshold equation, that is, with the scaling parameters of 0 and 1.

Based on Eqs. 4 and 6, one can obtain the relationship between the sampling interval and the eccentricity for each intended image (I1 to I8). As the video display resolution is limited by its pixel size, the possible simulated sampling intervals are the integer multiples of this pixel width, which is 1.7 min of arc in the current experimental setting. Figure 3 shows the relationship between the maximal eccentricity and the required sampling size in number of pixels for the processed images. The heavy line represents the condition determined by the threshold equation, that is, with the scaling parameters of 0 and 1. Examples of the degraded image of an arbitrary degradation level are shown in Fig. 4.



Figure 4. An example of the degraded images of an arbitrary degradation level using SuperPixel algorithm. The crosses indicate the fixation position.

Psychophysical Procedure

Each observer took part in two separate runs, each with a different scene, in a single session. In each run, there were a total of nine images (one original and eight different degradation levels). Each image was repeated ten times, presented in a blocked pseudo-random sequence. In each block, the nine images were presented in a pseudo-random order, but there were no repeats of the same images within a single block. In each trial, the image was displayed for 250 ms, where the global image contrast was modulated following the first-half period of a 2 Hz sine wave. Between trials, the screen was a uniform field with the same luminance and color as the mean of the images, with a fixation target at screen location (100, 100). The procedure is similar to the one used by Peli et al.¹⁹ The observers were instructed to estimate image quality by using a ratio scaling method. Observers were asked to assign a scale of 100 to the original image that was displayed on the screen before the start of each run. At the start of each trial, the observer fixated on the cross and pressed the space key on the keyboard, allowing the image to be displayed for 250 ms. After image presentation, the observer verbally indicated an estimate of the quality magnitude, by comparing the perceived image quality of the present trial with the previous one.

Observers

Results from five observers, including one of the authors, with normal or corrected vision of at least 20/20 are reported in this paper.

Experimental Results

The major task here is to find a degradation level that is just perceptible to human observers. The perceptible image degradation is reflected in the subjective image quality scale in reference to the original non-degraded image, (i. e., image I0). Figure 5 shows the subjective image quality scale at different image degradation levels. The data points are geometric means over 20 repetitions for each observer. It seems that individual observers assigned the quality number with a consistent bias on an absolute scale, that is, they continuously assigned values that were biased towards values below 100 even for non-degraded images that were included as catch trials. Nevertheless, this difference is not critical to the interpretation of the data when the relative image quality in reference to the original image, i. e., I0, is concerned. In this sense, all the individual results are compatible, with a similar tendency in the perceived image quality when the degradation level is changed. The dark heavy lines in each panel show the geometric means over the five observers. They are much smoother than the curves obtained for individual observers.

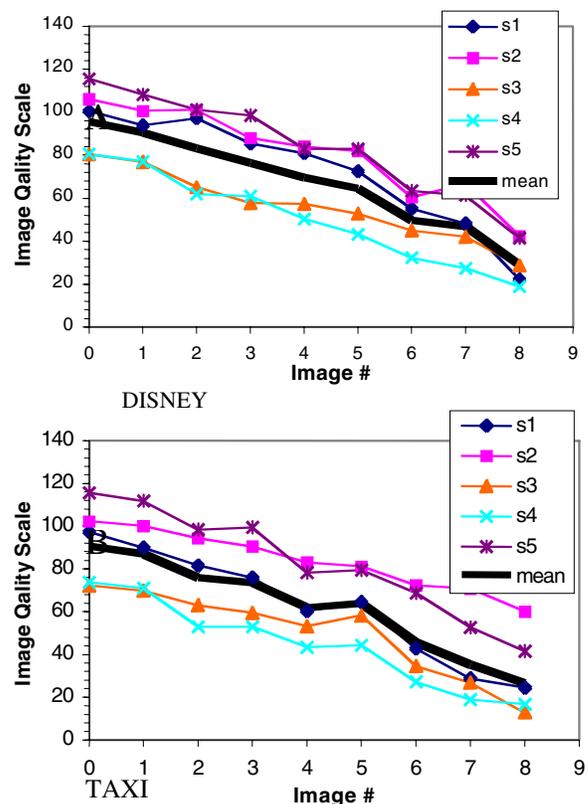


Figure 5. Subjective ratio scales versus degradation level (images #0 to #8) for the five observers, with (A) the scene DISNEY, and (B) the scene TAXI. The data points are geometric means over 20 repetitions for each observer. The dark heavy lines in each panel are the geometric mean over the five observers.

By looking at Fig. 5, it seems one can conclude that the perceived quality of image #1 is hardly different from the original image, for either DISNEY or TAXI. The obvious perceptual difference starts with image #2. At this degradation level, the image processing was based on the actual vision model (see Eq. 2) with the parameters S_1 being 0 and S_2 being 1. Further, a two-factor ANOVA test (degradation levels and different sessions of all the observers) showed that there was no significant difference in rating images #0 and #1 ($p > 0.1$). On the other hand, there was a significant difference in rating images #0 and #2 ($p < 0.001$).

Comparisons

We use the mean curves (i. e., the two dark heavy lines) to compare the subjective image quality between the two scenes. The curves were re-scaled to reflect the same score of 100 for image #0. Figure 6A shows relative subjective image quality. From Fig. 6A, one can see that the effective image degradation is not sensitive to a particular scene.

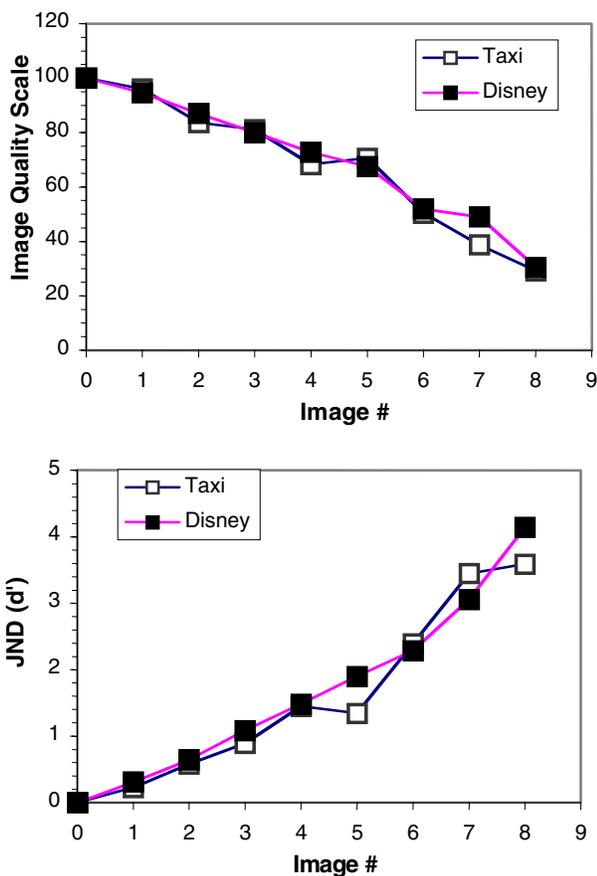


Figure 6. Comparing subjective image quality among the original scenes with a relative ratio scale (upper panel) and with the measurements of the just noticeable difference in the unit of d' (lower panel).

Ratio Scales and JNDs

The ratio scale provides a nice description of the monotonic relationship between the physical difference in the stimuli and the perceived subjective response for the concerned attribute. However, it is difficult to relate the exact scale values to underlying sensory or higher level processes. These numbers may not tell anything beyond this particular experiment. For example, there is no objective method to determine the threshold degradation level from the curves shown in Fig. 6A. To solve this problem, we used a method to convert the ratio scales to JNDs using d' prime. The results are shown in Fig. 6B, where the difference of 1 d' prime corresponds to 76% correct responses in a discrimination task when the experimental paradigm is two-alternative forced choice (see Macmillan and Creelman20). Based on the JND values shown in Fig. 6B, it is more informative to say that the threshold level corresponds to degradation level 3, at which the JND values are about 1 d' prime. The corresponding visual resolution at different eccentricities is shown by the dashed curve in Fig. 3.

Conclusion

The vision model based on visual sensitivity to simple grating patches at different eccentricities and spatial frequencies provides a useful tool to predict the perceptible image degradation in real complex scenes. The experimentally obtained threshold degradation level is somewhat higher than the vision model would predict. However, this is consistent with our expectation since it was expected that visual sensitivity is lower when the background consists of complex patterns, as in the real scenes used in the current experiment.

References

1. C. H. Anderson and C. R. Carlson, *U.S. Patent, 4692806*, (1987).
2. P. T. Kortum and W. S. Geisler, Human Vision and Electronic Imaging, *Proceedings of the SPIE*, **2657**, pg. 350 (1996).
3. P. M. J. Van Diepen, *Spatial Vis.*, **10**, pg. 335 (1997).
4. E. Chang and C. K. Yap, *ACM Symposium on Computational Geometry*, **13**, pg. 397 (1997).
5. W. S. Geisler and J. S. Perry, *Human Vision and Electronic Imaging III, Proceedings of the SPIE*, **3299**, pg. 294 (1998).
6. T. Duchowski, *IEEE Trans. Image processing.*, **9**, pg. 1437 (2000).
7. L. C. Loschky and G. W. McConkie, *ACM Eye Tracking Research & Applications Symposium 2000*, pg. 97 (2000).
8. S. J. Anderson, K. T. Mullen, and R. F. Hess, *J. Physiology (Lond)*, **442**, pg. 47 (1991).
9. L. N. Thibos, D. L. Still, and A. Bradley, *Vis. Res.*, **36**, pg. 249 (1996).
10. J. G. Robson and N. Graham, *Vis. Res.*, **21**, pg. 409 (1981).
11. M. W. Cannon, Jr. *J. Opt. Soc. Am. A*, **2**, pg. 1760 (1985).
12. J. S. Pointer and R. F. Hess, *Vis. Res.*, **29**, pg. 1133 (1989).

13. E. Peli, J. Yang, and R. Goldstein, *J. Opt. Soc. Am. A*, **8**, pg. 1762 (1991).
14. F. L. Van Nes and M. A. Bouman *J. Opt. Soc. Am.*, **57**, pg. 401 (1967).
15. J. Yang, X. Qi, and W. Makous, *Vis. Res.*, **35**, pg. 1965 (1995).
16. G. A. Geri and Y. Y. Zeevi, *J. Opt. Soc. Am. A*, **12**, pg. 2367 (1995).
17. D. H. Brainard, *Spatial Vis.*, **10**, pg. 443 (1997).
18. D. G. Pelli, *Spatial Vis.*, **10**, pg. 437 (1997).
19. E. Peli, J. Yang, R. Goldstein, and A. Reeves *J. Opt. Soc. Am. A*, **8**, pg. 1352 (1991).
20. N. A. Macmillan and C. D. Creelman, *Detection Theory: A User's Guide*. Cambridge University Press. (1990).

Biography

Jian Yang received a BS degree in physics from Fudan University in 1982, an MS degree in optics from the Shanghai Institute of Optics and Fine Mechanics in 1984, and a PhD degree in experimental psychology from Northeastern University in 1991. He previously worked at

the University of Rochester and the University of Houston. He joined Eastman Kodak Company in 1998. His current research is to apply human vision knowledge to image and information processing.

Terry Coia has been employed by Eastman Kodak Company for over 20 years. She has provided technical expertise to a variety of groups within Kodak, including CD&U-Human Factors, Performance Technology and Quality Assurance. Key projects include involvement in the commercialization of the ImageSource series of copier/printers, optical lens development and Image Evaluation Program project management.

Michael E. Miller received BS and MS degrees in Industrial and Systems Engineering from Ohio University and a PhD in Industrial and Systems Engineering from Virginia Tech. He has been employed as a human factors engineer at IBM and joined Eastman Kodak Company as a human factors engineer in 1994. He currently leads a group of engineers and psychologists interested in applying knowledge of human perception and cognition to imaging system design.