

# Estimating the Visibility of 2D Noise and Banding-Like Artifacts, Using a Human Vision Model

*Jian Yang and Sharon Field  
Eastman Kodak Company  
Rochester, New York*

## Abstract

Hard copy and soft copy pictures often contain artifacts, such as noise, banding, and streaking, which can reduce the perceived image quality. It is valuable to have a quantitative model of human visual performance to estimate the visibility of the combined artifact. In this paper, we present a vision model to estimate the visibility of luminance-only artifacts superimposed on a flat field. In this model, the image processing in the human visual system is simulated with four components: (1) a front-end, low-pass filter to simulate the function of the optics of the eyes and the sampling mosaic in the retina; (2) a compressive, nonlinear operation in the space domain to simulate light adaptation in the retina; (3) a compressive, nonlinear operation in the spatial frequency domain to simulate frequency masking at a cortical level; and (4) a local, linear summation, followed by a Minkowski summation, to derive the perceptible visual signal strength. Psychophysical data with soft copy stimuli were collected to test the model. There were 7 test patterns of different artifact types. Each of the patterns was manipulated and tested at 2 mean luminance levels, 3 contrast levels, and 2 viewing distances, totaling 12 conditions. The artifact visibility of each stimulus was described by the detectability index ( $d'$ ). The current model is reasonably accurate to fit the data with a residual RMS error of 0.29 and accounts for 98% of the variance in the measured  $d'$  values.

## Introduction

Photographic prints usually contain undesirable artifacts; such as banding, streaking, mottle, coalescence, dithering, and random noise, etc. If the artifacts are visible to human observers, they can reduce the perceived image quality of the pictures. Therefore, a method of how to estimate and predict the visibility of the artifacts is of value to the imaging and photographic industry.

Usually, one can use either subjective methods or objective, computational models to assess the visibility of any artifacts. Subjective methods are straightforward and

have been widely used in the image quality evaluation. Many subjective studies have been reported in the literature for evaluating banding visibility.<sup>1,2</sup> The disadvantage associated with subjective methods, however, is that it can be very time consuming for the human observers and, hence, very costly.

Often a large amount of observations are necessary to deliver a set of reliable data about the artifact visibility. In contrast, computational models are more efficient in providing repeatable estimates of the visibility for specified artifacts.

The difficulty with the computational approach is the development of a model that closely matches human visual performance. Artifact patterns can have different structures, and this difference can lead to different visibility. Furthermore, the visibility of the artifact is a function of many variables, such as contrast, mean luminance level, background image, and viewing distance.

There are several types of computational models. In some well-defined engineering problems, the artifact patterns contain limited variations. In this case, one could develop computational algorithms to capture the relationship between the artifact variables and the corresponding visibility.<sup>3,4</sup> The disadvantage of this type of model is the lack of generalizability; this type of model often cannot be extended to applications where the conditions are not exactly the same as those used to construct the model. Sometimes, it is also often important to determine the sources of printing defects. This requires information on the types of image artifacts that may be generated by any imaging system. Along this direction, some models such as Kane et al.<sup>5</sup> and Donohue et al.,<sup>6</sup> concentrated on how to decompose the patterns into different types of artifacts or image structures. The major concern of this paper is to estimate the overall visibility of luminance-only artifacts that are superimposed on a flat field by using a human vision model.

Vision modeling is an active research area; human vision knowledge has been widely implemented in computational models.<sup>7-9</sup> Nevertheless, the existing models have been tested against very limited psychophysical data. The purpose of the current research is to test and refine the

Yang model<sup>9</sup> and to make the model estimation consistent to our new experimental data with regard to artifact visibility. In the experimental test, visual stimuli, i. e., the artifact patterns, included seven different spatial structures (see Fig. 1 for two samples of the seven patterns). We also manipulated contrast, mean luminance level, and viewing distance to cover a broad range of viewing conditions.

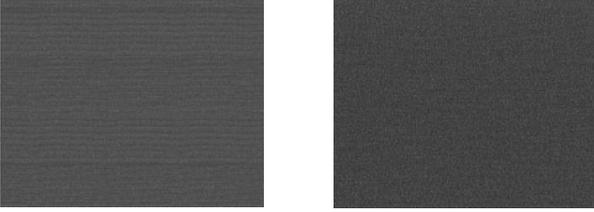


Figure 1. Examples of two artifact patterns that are superimposed on flat fields of different colors. The pattern shown on the left panel is dominated by banding-like artifacts on a green field, and the pattern shown on the right panel is dominated by 2D noise on a purple field.

### Implementation of Human Visual Image Processing

Human vision models have been developed to simulate the underlying mechanisms of the visual system and thus to provide an estimate of visual performance for detecting and discriminating visual targets. The most commonly used components in a vision model are a front-end, band-pass spatial frequency filter and Gabor-like frequency channels.<sup>10-12</sup> The band-pass filter is to simulate the shape of contrast sensitivity functions (CSFs) on an assumption that the shape originates from the retina. Yang and Makous<sup>13</sup> suggest that the drop of CSF at low spatial frequency is caused by the effect of frequency masking from a zero frequency component. Along this thought, the front-end frequency filter is likely to result in a low-pass shape and, consequently, Yang<sup>9</sup> developed an algorithm to simulate visual image processing with three functional components: (1) a front-end, low-pass filter to simulate the function of the optics of the eyes and the sampling mosaic in the retina; (2) a compressive nonlinear operation in the space domain to simulate light adaptation in the retina; and (3) a compressive, nonlinear operation in the spatial frequency domain to simulate frequency masking at a cortical level as schematized in Fig. 2 and explained in the following sections.

#### Front-End, Low-Pass Filtering

The optical lens of the eyes, and the sampling mosaic in the retina, has the role of attenuating high spatial frequency components in an image. As a result, the combined front-end, low-pass filter is approximated here by:

$$MTF(f) = \exp\{-(\alpha_0 + \delta / \sqrt{L_0})f\}, \quad (1)$$

where  $L_0$  is the mean luminance of the visual pattern,  $f$  the spatial frequency,  $\alpha_0$  a parameter specifying the rate of attenuation, and it can be varied to fit individual difference and viewing conditions.  $\delta$  is a correction term to show the effect of the mean luminance on the low-pass filter and was set to 0.013.

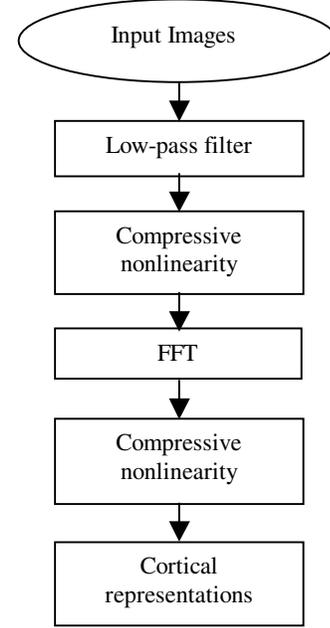


Figure 2. The schematized framework of visual image processing.

#### Retinal Compressive Processing

This step is to simulate light adaptation in the retina to show that the visual system is adaptively reducing sensitivity to incoming signals when the retina is exposed to high-intensity signals. Light adaptation effects are spatially localized.<sup>14</sup> Here we assume that the adaptation pools are constrained by an aperture window, which is approximated here by a Gaussian function:

$$W_g(x, y) = \frac{1}{2\pi r_g^2} \exp\left(-\frac{x^2 + y^2}{2r_g^2}\right), \quad (2)$$

where  $r_g$  is the standard deviation of the aperture and was set to 1 arc min. The pooled adaptation signal  $I_g$  is the convolution of the low-passed input image  $I_c$  with the window  $W_g$ . The retinal image representation  $I_R$  is the output of a compressive nonlinearity, as assumed here a modified Naka-Rushton equation:

$$I_R = \frac{I_c^{n_1}}{I_g^{n_1} + I_1^{n_1}}, \quad (3)$$

where  $n_1$  and  $I_1$  are the parameters that represent the exponent and the semi-saturation constant of the Naka-Rushton equation, respectively.

### Cortical Compressive Processing

To capture the frequency- and orientation-specific nonlinearity of the cortical cells, the image  $I_R$  is converted from a spatial domain representation to a spatial frequency domain representation via a Fourier transform to  $T(f_x, f_y)$ . Here  $f_x$  and  $f_y$  are spatial frequencies in  $x$  and  $y$  directions, respectively. Cortical nonlinear property is usually described by frequency masking. In the current model, we assume  $T_m(f_x, f_y)$ , the signal strength in the masking pool, being the convolution of the absolute signal amplitude  $|T(f_x, f_y)|$ , and an exponential window function:

$$W_c(f_x, f_y) = \exp[-(f_x^2 + f_y^2)^{0.5}/\sigma], \quad (4)$$

where  $\sigma$  correlates with the extent of the frequency spreading. After applying the same form of compressive nonlinearity as in the retina, the cortical signal is expressed as:

$$T_c(f_x, f_y) = \text{sign}[T(f_x, f_y)] \frac{|T(f_x, f_y)|^{n_2}}{T_m(f_x, f_y)^{n_2} + I_2^{n_2}}, \quad (5)$$

where  $n_2$  and  $I_2$  are parameters that represent the exponent and the semi-saturation constant of the Naka-Rushton equation for the cortical, nonlinear compression, respectively.

This model was able to simulate the change of CSF with mean luminance level after adding a decision stage that is simply a Minkowski summation of the cortical signal  $T_c$  over different frequency components.<sup>9</sup> As the energy of sinusoidal gratings is localized in the frequency domain (the so-called narrow-band stimuli), the summation rule is not particularly critical to the model results. Conversely, the spectra of current visual patterns contain a wide distribution in the frequency domain. The summation rule can be critical for delivering final signal strength in the decision-making stage. Thus, we have compared different summation rules to compute the final signal strength.

### Signal Summation in the Decision Stage

Physiological results with single cells showed that the frequency response curves of cortical neurons are quite broad; the response bandwidth covers a range of approximately 1 to 2 octaves.<sup>15</sup> Input signals within this range are linearly summed together to derive the neuronal response. Accordingly, we assume there is a local linear summation stage. For computational convenience, the linear summation window is approximated by a square window with the side length of  $w$  in the 2D spatial frequency domain, as suggested by Makous and Yang.<sup>16</sup>

Following this linear summation stage, Makous and Yang<sup>16</sup> add a Minkowski rule to sum up the absolute values of each channel activity. This summation is similar to the approach taken by investigators such as Watson and Solomon<sup>12</sup> and Watson.<sup>17</sup> Under this operation, every input data point has the same weight in computing the model output response.

Minkowski summation is a simple form for evaluating the overall signal strength. It is hard to claim this is the only or the best approach to compute the output visual response. In the current implementation, we explored a few variations of the summation. In one implementation, the output signal is the weighted sum of the absolute channel responses by a Gaussian window with a standard deviation of  $\sigma_w$ , where the input signals that are located at a far distance in the 2D frequency domain have very small weights to the output signal. Furthermore, the final visual response is given by the maximal value of the output signals.

### Other Model Refinements

It was reported that the amount of spatial frequency spread in the visual processing reduces as the field size increases, following an empirical formula:<sup>18</sup>

$$\sigma = 0.042 + k/D, \quad (6)$$

where  $D$  is the field size in unit of arc degrees and  $k$  is a constant with a value of 0.64.<sup>18</sup> In the current experiment, the field size was manipulated by changing the viewing distance, while keeping the physical display area constant. We inserted Eq. (6) to Eq. (4) to reflect the effect of the field size on the frequency spreading and, thus, the masking pool size.

The second refinement is on the expression of the compressive, nonlinear functions. Nonlinear functions have been commonly used in vision models, such as the divi-sive, nonlinear normalization process used by Heeger,<sup>19,20</sup> Foley,<sup>21</sup> Itti and Koch.<sup>22</sup> In those models, the exponent in the numerator is set to be different from the exponent in the denominator. In the current model refinements, we followed this approach by replacing  $n_1$  with  $n_{n1}$  in the numerator and  $n_{d1}$  in the denominator in Eq. (3), and replacing  $n_2$  with  $n_{n2}$  in the numerator and  $n_{d2}$  in the denominator in Eq. (5). After this change, the model is more flexible and is able to fit a variety of experimental data.

### Psychophysical Data

Psychophysical data with softcopy stimuli were collected to test the model. There were 7 test patterns, each with a different artifact type (see Fig. 1 for examples). Each of the patterns was further manipulated and tested at 2 mean luminance levels, 3 contrast levels, and 2 viewing distances, totaling 12 conditions.

### Methods

The experiments were run on a Power Mac G4 computer with a 24" monitor (Sony Trinitron Color Graphic Display). The monitor has 8 bits of gray level for each of the R, G, and B channels, and the luminance of the displayed image was linearized to the image code value after applying a lookup table. Experimental software was developed in MATLAB using the extensions provided by the high-level Psychophysics Toolbox.<sup>23</sup> The screen resolution was set to 1280 x 1024 pixels.

**Table 1. Characteristics of the 7 Visual Patterns**

Pattern #	Luminance (cd/m <sup>2</sup> )	CIE_x	CIE_y	RMS contrast
P1	37.22	0.282	0.401	0.0502
P2	58.88	0.257	0.301	0.0218
P3	27.44	0.258	0.262	0.0657
P4	30.52	0.283	0.292	0.0832
P5	27.88	0.272	0.248	0.0638
P6	52.18	0.375	0.291	0.0187
P7	30.40	0.278	0.282	0.1567

Visual stimuli were created from 7 scanned patches that were printed on inkjet printers. The original patterns have different colors, luminance levels, and artifact types that look like banding or 2D noise. As the current model does not deal with color detection, we kept only luminance artifacts in the stimuli. The artifact contrast maps were scaled equally over the R, G, and B channels. Table 1 shows mean luminance level, CIE x and y coordinates, and RMS contrast of the 7 patterns used in the experiment. For each original pattern, 8 different versions of the images were created by using a luminance multiplier of 1 or 0.1 and a contrast multiplier of 1, 0.5, 0.25, or 0.

Ten observers with normal or corrected to normal vision participated in the experiment. Observers were instructed to give ratio scores on the visibility of the artifact patterns. Two anchor points were added in the rating procedure: 0 for uniform fields and 5 when the pattern was slightly detectable. Any positive number could be used to rate the pattern visibility. Furthermore, the observers were shown a reference image, i.e., a sinusoidal grating of 10% contrast at the start of the experiment. The reference image was told to have a rating of about 50.

The experiment was run in a darkened room. A given session of the experiment included 2 separated blocks of trials, with the luminance multiplier being either 1 or 0.1. Within each block, there were 28 different stimuli (i.e., 7 patterns x 4 contrast levels) that were displayed in pseudo-random sequences. Furthermore, each stimulus was repeated 10 times within a block. Each observer took part in 2 sessions of the experiment on different days. In the first session, the viewing distance was 167 cm and the viewing field size with 1280 x 1024 pixels was 12.5 x 10.0 arc degrees. In the second session, the viewing distance was 334 cm and the field size was 6.4 x 5.1 arc degrees.

## Results

For each observer, there were 10 visibility-rating scores for each visual stimulus at any fixed viewing distance, contrast, and mean luminance. These raw scores were used to calculate the  $d'$  values, that are equivalent to just noticeable difference, of the experimental conditions based on the overlap of the score histograms.

Figure 3 shows the detectability of seeing the 7 patterns with the contrast multiplier of 1 (diamonds), 0.5 (squares), and 0.25 (triangles), at a viewing distance of 167 cm. The top panel was obtained with the mean luminances that are specified in Table 1, i. e., with a luminance multiplier of 1. The bottom panel was with a luminance multiplier of 0.1, i. e., the mean luminance of a given pattern was 10% of the same pattern in the top panel. The detectability has been adjusted after subtracting an offset  $d'$  value. The offset was the  $d'$  values obtained with the zero contrast stimulus that was run, otherwise, in the same experimental conditions.

The results shown in Fig. 4 were obtained with a far viewing distance of 334 cm. Furthermore, pattern #2 was not tested in this session in the experiment. Other conditions are exactly the same as those for Fig. 3.

The smaller the detectability, the less visible the artifacts on the soft copy display were to the observers. The experimental results show that the visibilities of the seven patterns are different and, for a given pattern, the visibility index is a monotonic function of the contrast. The primary aim of this paper is to estimate the visibility value when a visual stimulus is known. The curves are the model fits to be discussed later.

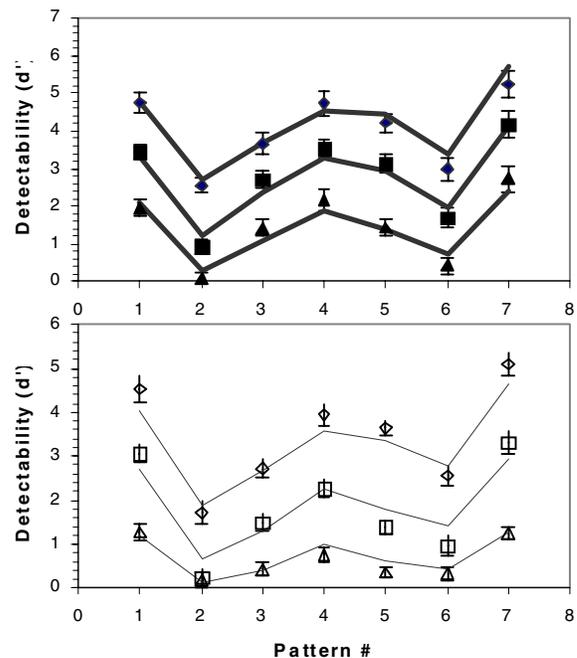


Figure 3. The detectability of 7 patterns with the contrast multiplier of 1 (diamonds), 0.5 (squares), and 0.25 (triangles,) at a viewing distance of **167** cm. The mean luminance of the patterns was either that shown in Table 1 (top panel, or reduced to 10% of the Table values (bottom panel). Error bars stand for  $\pm 1$  standard error of the data over 10 observers.

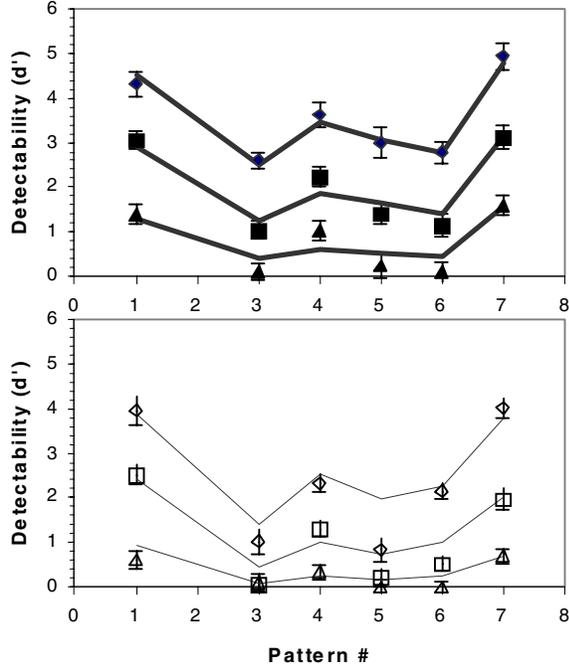


Figure 4. The detectability of 6 patterns with the contrast multiplier of 1 (diamonds), 0.5 (squares), and 0.25 (triangles,) at a viewing distance of 334 cm. The mean luminance of the patterns was either that shown in Table 1 (top panel, or reduced to 10% of the Table values (bottom panel). Error bars stand for  $\pm 1$  standard error of the data over 10 observers.

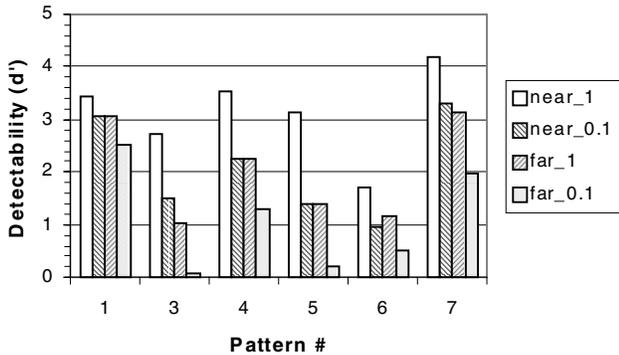


Figure 5. A comparison of detectability over viewing distance and mean luminance level with 6 artifact patterns at the contrast multiplier of 0.5. The viewing distance was either 167 cm (near) or 334 cm (far), and the luminance multiplier was either 1 or 0.1.

The effects of the pattern and contrast differences on the visibility are easy to see from Figs. 3 and 4. However, these figures do not clearly show the visibility differences caused by viewing distance and mean luminance level. Thus, we plotted Fig. 5 to compare the effects of viewing distance and mean luminance level directly at a median contrast multiplier of 0.5. In the coded caption, “near”

refers to a viewing distance of 167 cm, and “far” refers to the distance of 334 cm. The number following the dash is the luminance multiplier, which is 1 or 0.1 times the pattern luminance. As the plot shows, the detectability decreases as viewing distance increases or as mean luminance decreases. These trends are consistent with the CSF behavior at median and high spatial frequencies, where CSF reduces as spatial frequency (or viewing distance) increases and as luminance decreases.<sup>24</sup> The results at contrast multiplier 1 and 0.25 show a similar trend, and they are not plotted here.

## Model Fits

Some procedures for fitting the model results with the experimental data need to be specified here. The framework of the vision model is described in the model section. To estimate the visual response to a visual stimulus, the first step is to pre-process the input digital image in order to capture its luminance distribution in units of  $\text{cd}/\text{m}^2$ . Furthermore, it is known that soft copy displays usually contain veiling glare.<sup>25</sup> The estimated veiling glare in our display is about 2% of the mean luminance of a displayed image, and this amount is added to the images. Next, the luminance-coded image is entered to the model algorithm to derive its cortical representation. Finally, the signal entered to the decision-making stage is a frequency difference map:

$$S_d(f_x, f_y) = T_1(f_x, f_y) - T_0(f_x, f_y), \quad (7)$$

where  $T_1$  and  $T_0$  are the cortical representations of the stimulus and a reference image that are determined by Eq. (5). In the current case, the reference image is the uniform background field.

To fit the model with the data, one also needs to find an appropriate conversion of the visual response to visual detectability and to obtain the model parameter values. Based on a hypothesis that the perceptual, linear detectable difference is proportional to the logarithmic transform of the perceptual response magnitude, we have a relationship

$$d' = \log(\rho R_c + 1), \quad (8)$$

where  $\rho$  is a constant, and  $R_c$  is the model-based visual response from the decision stage. Furthermore, we used a MATLAB routine *fminsearch* to search for the appropriate parameter values for an optimal fit of all the experimental data simultaneously.

The curves shown in Figs. 3 and 4 are the obtained model fits to the experimental data, with a residual RMS error of 0.29, and accounts for 98% of the variance in the experimental data.

We also explored two additional decision summation rules for the fit to the experimental data as shown in Figs. 3 and 4. The simplest approach is to apply Mikowski summation to the absolute value of the frequency difference map as determined by Eq. (7). The resulting optimal fit has a residual RMS error of 0.34, which is higher than the RMS error for the fit shown above. The

optimal exponent for the summation is 1.22. Another approach is to sum the frequency difference map linearly together within square windows. In this case, the absolute value of the signal is further summed with a Gaussian weighting function. The final, visual response is given by the maximal value of the output signals. After implementing this decision summation rule, the obtained error of the fit was 0.31, which has an error level that is in the middle of the other error produced by the other two methods.

## Conclusions

1. The experimental results show that artifact visibility increases as contrast or mean luminance increases, and it decreases as viewing distance increases. These changes are consistent with our intuition.
2. The modified vision model is able to describe the artifact visibility at both threshold level and suprathreshold levels for a variety of viewing conditions, including different artifact patterns, illumination and contrast levels, and viewing distances. The residual RMS error of the best-model fit to the total 78 experimental data was 0.29, in terms of the detectability index  $d'$ .
3. We compared three summation rules in the decision making stage: a) Minkowski summation; b) linear summation within square windows followed by Minkowski summation; and c) linear summation within square windows followed by a weighted Gaussian summation. The amount of residual errors in fitting current experimental data are similar for the three methods, with the case b) method providing the smallest residual RMS error of 0.29.
4. Some of the model parameter values are quite different from those obtained in the previous fits with other experimental results.<sup>9</sup> This is an area that requires more attention in further model comparison.

## Acknowledgments

The authors thank Dr. Michael Miller for helpful comments on an early version of the paper. We thank the observers who participated in the experiment.

## References

1. C. Cui, D. Cao, and S. Love, *IS&T 2001 PICS Proc.*, pg. 84 (2001).
2. C. Cui, L.G. Hassebrook, C. Guan, and S. Love, *IS&T 2002 PICS Proc.*, pg. 99 (2002).
3. D. R. Rasmussen, E. N. Dalal, and K. M. Hoffman, *IS&T 2001 PICS Proc.*, pg. 90 (2001).
4. P. J. Kane, T. F. Bouk, P. D. Burns, and A. D. Thompson, *IS&T 2000 PICS Proc.*, pg. 79 (2000).
5. B. W. Keelan, *Handbook of image quality*, New York, NY, Marcel Dekker (2002).
6. K. D. Donohue, M. V. Venkatesh, and C. Cui, *IS&T 2002 PICS Proc.*, pg. 42 (2002).
7. X. Feng, J. Speigle, and A. Morimoto, *IS&T 2002 PICS Proc.*, pg. 5 (2002).
8. M. Lian, Q. Yu, and D. W. Couwenhoven, *IS&T 2002 PICS Proc.*, pg. 11 (2002).
9. J. Yang, *SPIE Proc.*, **4662**, pg. 84 (2002).
10. E. Peli, *Optom. Vis. Sci.*, **69**, pg. 15 (1992).
11. S. Daly, *SPIE Proc.*, **1666**, pg. 2 (1992).
12. A.B. Watson and J.A. Solomon, *J. Opt. Soc. Am. A*, **14**, pg. 2378 (1997).
13. J. Yang and W. Makous, *Vis. Res.*, **34**, pg. 2569 (1994).
14. M. M. Hayhoe and M. V. Smith, *Vis. Res.*, **29**, pg. 457 (1989).
15. R. L. De Valois and K. K. De Valois, *Spatial Vision*, New York, Oxford University Press, (1988).
16. W. Makous and J. Yang, OSA Annual Meeting, pg. 52 (2000).
17. A. B. Watson, *Optics Express*, **6**, pg. 12 (2000).
18. J. Yang and W. Makous, *Vis. Res.*, **37**, pg. 1917 (1997).
19. D. J. Heeger, *Visual Neurosci.*, **9**, pg. 181 (1992).
20. D. J. Heeger, *Curr. Dir. Psychol. Sci.*, **3**, pg. 159 (1994).
21. J. M. Foley, *J. Opt. Soc. Am. A*, **11**, pg. 1710 (1994).
22. L. Itti, C. Koch, and J. Braun, *J. Opt. Soc. Am. A*, **17**, pg. 1899 (2000).
23. D. H. Brainard, *Spatial Vis.*, **10**, pg. 443 (1997).
24. F. L. Van Nes and M. A. Bouman, *J. Opt. Soc. A*, **57**, pg. 401 (1967).
25. H. Roehrig, In Y. Kim and S. C. Horii (Eds.) *Handbook of Medical Imaging Vol. 3.*, Display and PACS, SPIE Press, pg. 155 (2000).

## 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 Ph.D. degree in Experimental Psychology from Northeastern University in 1991. He previously worked at the University of Rochester as a postdoc fellow and the University of Houston as a research associate. He joined Eastman Kodak Company in 1998. His current research is to apply human vision knowledge to image and information processing.

**Sharon Field** received a BA degree in Psychology with Honors from the University of Rochester in 1994. She currently works as a Research Technician in the Corporate Design and Usability lab at Eastman Kodak Company. Her primary interests are in human computer interaction and image-quality evaluations.