

Halftone Quality Evaluation Using Color Visual Models

*Xiao-fan Feng, Jon Speigle and Atsuhisa Morimoto
Sharp Laboratories of America, Inc
Camas, Washington*

Abstract

We conducted a monitor-based subjective study on multilevel halftone. Observers in a halftone preference experiment selected stimuli in a paired comparison design. Images were halftoned using multilevel stochastic screening and multilevel error diffusion. Stimuli were presented on a calibrated monitor from distances equivalent to viewing 300 and 600 dpi prints. We applied Thurstone's law of comparative judgments to derive an image quality scale from the paired comparison data. The derived scale values were compared to summary measures computed from variants of a color difference model (CVDM).¹ CVDM is a multi-channel model having multiple orientation and spatial frequency mechanisms. CVDM was modified to better model the human visual system. An eye OTF front-end is added and CSF filtering and cortex transform are performed in the LAB space. We compare predictions of CVDM with S-CIELAB² and observed slightly better correlation between the models and subjective scale values. On a local scale, the visible difference maps produced by the multi-channel model appear to capture worming and texturing artifacts better than S-CIELAB.

Introduction

Digital halftoning is a process of using a limited number of grays or colors to give a perception of a continuous-tone image. Two classes of halftoning algorithm are error diffusion (ED) and frequency-modulated (FM) screening. In error diffusion, output dots are generated when the sum of the input pixel value and a propagated error signal exceeds a threshold. In FM screening, pixel values are compared to a spatially varying threshold array (TA) where the thresholds are designed to produce an image with high-pass or "blue" noise characteristics. In comparison to FM screening, error diffusion tends to produce images with smoother tone gradation because dot spacing can be adaptively varied as a function of image content. FM screening tends to produce slightly grainier output with the benefit of reduced computation. The dominant artifacts for FM screening include graininess and color mottle. Error diffusion exhibits these to a lesser extent, but can also introduce worming and midtone texturing artifacts that are caused by the inhomogeneous method in which errors are propagated.

Our goal is to develop objective visual quality metrics that can characterize halftone quality as a function of system and algorithm parameters. A quantitative method for evaluating halftoned images would permit optimizing algorithms because it could establish relative quality. One quality metric is the frequency-weighted mean squared error (FWMSE) which takes into account the human visual system's relative sensitivity to different spatial frequencies. To compute this metric both the contone and halftone images are filtered with an approximation of the luminance CSF. The MSE is then computed between these filtered images and is used as the quality metric. The FWMSE metric, however, fails in some applications. For example, it is unable to predict the midtone texturing associated with error diffusion; and several studies have reported weak correlations with subjectively rated image quality.³

More recently, a spatial extension of CIELAB (S-CIELAB) was used to predict halftone texture visibility.⁴ The model incorporates the luminance/chrominance CSFs of human eye and inherits the uniform perception from CIE Lab color space. The model did reasonably well for black-white patterns ($R^2=0.89$), but the overall correlation is still low ($R^2=0.46$). The authors hypothesized that the low correlation could be because S-CIELAB does not include provision for masking or orientation specific effects.

The color visual difference model (CVDM)¹ combines the multi-resolution and masking components of Daly's monochromatic visible difference predictor (VDP)⁵ and S-CIELAB. The VDP not only models the overall spatial frequency response of the visual system, but also masking between patterns of similar orientations and spatial frequencies. This masking plays an important role in perceiving complex images. In this paper, the original CVDM is modified to better model the human visual system. The visual frequency sensitivity is modeled with a linear optical eye model combining a contrast sensitivity function in the non-linear Lab space. We apply this chromatic multi-channel model to multilevel halftoned images. We present a subjective experiment in which a subjective scale was derived for two types of multi-level halftone algorithms: ED and FM. Reasonable correlations are achieved between the summary metrics and the subjective data. We also compared CVDM results with those from S-CIELAB; we obtained a slightly higher correlation and better texture prediction on a local scale.

Subjective Experiments

Stimulus

Stimuli were halftoned by a multilevel stochastic screen and error diffusion algorithm. Seven types with output levels from 2 to 8, of each algorithm were used. The continuous-tone input was a 320 x 320 CMY grayscale image as shown in Figure 1. Patch levels linearly sampled the 0 to 255 range. The C, M, and Y stochastic screens were generated to have maximal decorrelation between output color dots in order to have reduced graininess. The error buffer in the ED algorithm was randomly initialized to avoid correlating CMY planes. Otherwise, ED was Floyd-Steinberg error diffusion. The output halftoned image was scaled to 640 x 640 by replication in order to reduce monitor MTF effects.

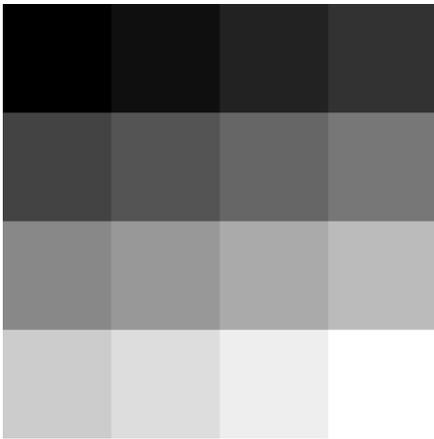


Figure 1. Original target used in subjective experiment

Images were rendered for display on the monitor using the following procedure. The input and halftoned images were defined as linear CMY. The halftone output levels were uniformly spaced in this CMY space. The CMY halftoned image was converted to linear RGB as

$$r = 255 - c, \quad g = 255 - m, \quad b = 255 - y$$

These RGB values were treated as linear sRGB values and mapped to XYZ using the matrix for linearized sRGB-to-XYZ. This resultant image is relative XYZ values with a luminance range of 0 to 1. The Barco was calibrated so that the relationship between RGB settings values and absolute XYZ was known. Displaying the relative XYZ image first required converting the relative XYZ values into absolute XYZ values. This was done by linearly mapping the [0-1] luminance range onto the absolute luminance range of the monitor. The absolute XYZ image was then mapped to RGB using the monitor calibration.

Procedure

Pairs of halftoned images were presented side-by-side in a gray surround on a black background. The monitor was

framed in a black cardboard surround and was viewed in a dimly lit room. Monitor resolution was set to 1600 x 1200 dpi. Observers viewed the display from distances that made the display equivalent to viewing 300 and 600 dpi prints at 14", corresponding to Nyquist frequencies of 37 and 73 cycles per degree. The separation between images was 100 pixels and the surround was 100 pixels from the borders of each image. Observers indicated which image was preferred (in this case, closer to uniform patches) using a keyboard.

The method was complete paired-comparison. The number of pairs was $91 N(N-1)/2$, where N is the number of samples (N=14). Each pair was replicated 4 times in a session, bringing the total to 364 stimuli per condition. Presentation was randomized.

10 Landolt-C tested observers participated in the experiment including the authors. Half of the observers started with the 300 dpi condition while half started with the 600 dpi condition.

Analysis

We represent the results of the preference experiment as a preference matrix, P. Each element, p_{jk} , indicates the proportion of times that the jth stimulus was preferred over the kth stimulus. We set entries $p_{ji} = 0.5$ and $p_{jk} = 1 - p_{kj}$. We used a generalized linear model to fit Thurstone's case V so that the preference matrix was converted to subjective scale values.⁶⁻⁸ Thurstone's law is based on the assumption that location on the perceptual scale is related to the confusability between stimuli. Assuming that stimulus responses are normally distributed, the separation between responses can be written as

$$R_j - R_k = z_{jk} \sqrt{\sigma_j^2 + \sigma_k^2 - 2r_{jk} \sigma_j \sigma_k}$$

where z_{jk} is the z-value corresponding to proportion, p_{jk} . In case V, the assumptions are made that the standard deviations are equal and that the responses are uncorrelated. Dropping a scale factor, the relation reduces to

$$R_j - R_k = z_{jk}$$

The matrix of z-values, Z, indicates the perceptual separation between stimuli. The mean of each row of Z can be used as the final estimate of the scale value.

One difficulty with this method involves the z-values that correspond to stimulus pairs where preference was 100 or 0%. These transform to +/-infinity. For the 300 dpi results, there are quite a few of these cases. For this reason, only the 600 dpi data are analyzed in this paper.

Visually-Based Halftoning Metrics

We applied two visual models to the images used in the experiment: S-CIELAB and CVDM. Both models account for the reduced sensitivity of the visual system to high spatial frequency patterns. This sensitivity is typically summarized as the inverse of the contrast threshold for detecting luminance (Y) and chromatic isoluminant (R/G and Y/B) sinusoidally-varying patches. Figure 2 presents Mullen's measurements of these curves.⁹ In general the

color spatial sensitivity functions are low-pass while the luminance sensitivity function is band pass and peaked at 2-3 cycles/degree. The luminance channel has a higher cutoff frequency.

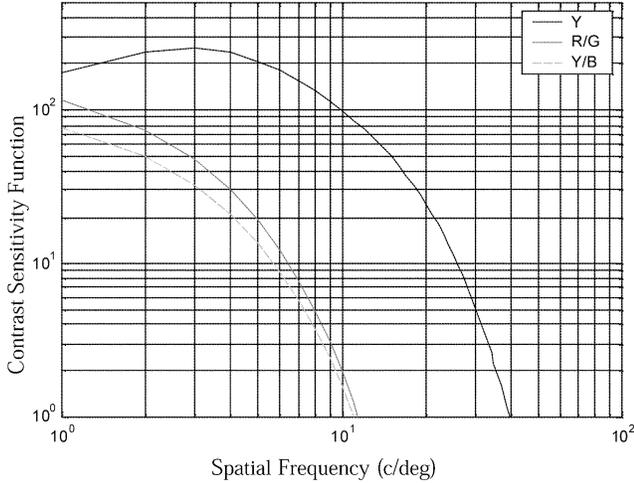


Figure 2. Contrast sensitivity functions of luminance (Y) and chrominances (R/G, and Y/B)

In S-CIELAB, a pair of color images are converted into a linear opponent color space and filtered by a set of spatial contrast sensitivity functions. The resultant images are then converted back to the CIE LAB colorspace and the traditional ΔE^* metric is applied. The end result is that high frequency color and luminance variation is reduced or removed from the image.

CVDM extends S-CIELAB by incorporating a multi-resolution representation and visual masking. Similarly to a number of monochromatic models,⁵ the luminance and opponent color signals are filtered by their CSFs and then decomposed into multiple orientation and narrowband spatial channels. Masking is introduced by reducing the visibility of distortions according to the amount that the oriented narrowband cortex channel is excited. This masking serves to predict the reduction of visibility of distortions in non-uniform areas.

The CVDM model has been previous applied to predict color mis-registration artifact with reasonably accuracy.¹⁰ But in using the model for color mis-registration, text quality, and grating detection studies, we encountered two difficulties:

- 1) CVDM uses the same opponent color space for its spatial filtering as S-CIELAB. Filtering in the OPP space sometimes produces out-of-gamut colors which can lead to large ΔE^* values.
- 2) The CSFs used in CVDM and S-CIELAB are low pass filters with a peak sensitivity of 1. This means the just noticeable difference is about one ΔE^* unit. CSF measurement data show that the luminance CSF peaks at about 3-5 cycles per degree with a peak detectable

difference is less than 0.5 ΔE^* in a normal office viewing condition.^{11,12}

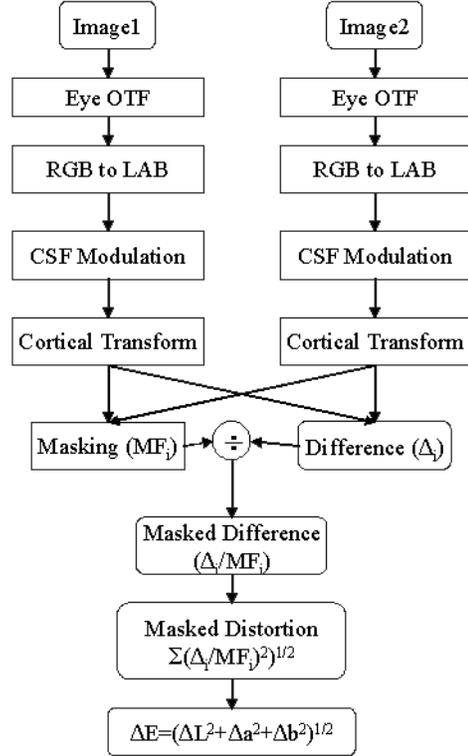


Figure 3. Flow chart of modified CVDM

To avoid these shortcomings, we have modified the CVDM as shown in Figure 3. Instead of performing the CSF filtering and cortex transform in the opponent space, both are performed in the LAB space. This eliminates the production of out-of-gamut colors. This processing order is closer to the human visual pathway, where the retinal image is blurred by the optics of the eye. The retinal image is then sampled by the cones and converted to a neural signal with a nonlinear response. The neural signals are encoded in opponent space in the retina. This nonlinear sensor function and opponent space is modeled with the RGB to LAB color conversion. CSF filtering is performed in the LAB space. After CSF filtering, the L^* and a^* and b^* images are decomposed by a cortex transform,⁵ in which each color channel are decomposed into a number of more narrowly-tuned frequency/orientation bands. Each band represents the responses of the early cortical mechanisms that is tuned in both spatial frequency and orientation. A masking image (MF_i) and difference image (Δ) is computed for each band. This difference is divided by the masking image to form the masked difference for that band (Δ/MF_i). The visual difference between the two images is calculated from the masked differences of each band by Minkowski summation. An exponent of 2 was used in this study. And finally, a standard ΔE error formula is used to calculate the color appearance difference.

For the luminance OTF of the eye, we use Barten's model with a pupil diameter of 3 mm.¹² The chromatic OTF is modeled with defocus arises from axial chromatic aberration.¹³ A complete OTF describes the spatial filtering at each wavelength. Because we are considering a monitor display with known primaries, it is possible to approximate the OTF using spatial filters on each RGB channel. Figure 4 shows the OTF of the eye approximated using the spectral power distribution of a Barco monitor. The OTF of red and green channels are the same, while the blue channel OTF is significantly more low-pass.

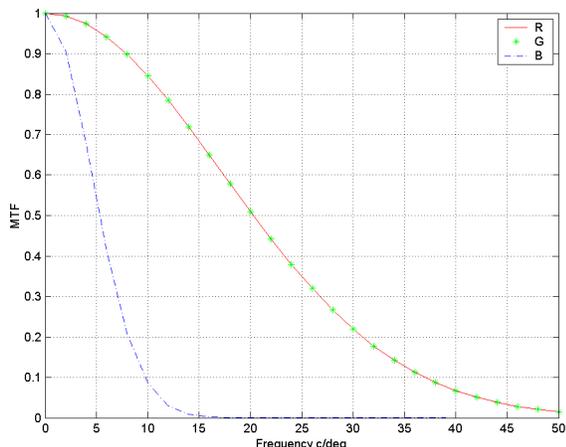


Figure 4. OTF of human eye for Barco monitor primaries

The luminance CSF is derived from Barten's CSF model without the optical OTF of the eye. The peak sensitivity occurs at near 5 cycles per degree and is set to 260 for a luminance of 500 cd/m² and picture angular size of 5 degrees. The low frequency portion of the CSF is modified so that it remains constant from 0.9 c/deg to DC. For most office lighting conditions, this will yield DC CSF of about 100. This will translate to one ΔE^* in L for large patches, matching the CIELAB L*. At 3 to 5 c/deg, the minimum detectable ΔE^* is about 0.4, which agrees with recent measurements by Klassen et al.¹¹ Chrominance CSFs are modeled as lowpass filters based on Mullen's data. These CSF filters are shown in Figure 5.

We tuned the masking exponent using a set of Gabor stimuli displayed on a field of band-passed white noise. The masking parameter was adjusted until the model predictions matched the subjective appearance of the targets. We found the masking exponent to be 0.8 for both luminance (L) and chrominances (a and b). There is no masking in the base-band since masking is accounted for in the nonlinear conversion from RGB to LAB.

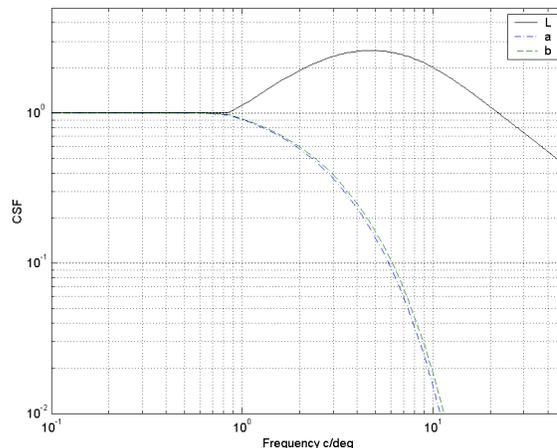


Figure 5. CSF of L, a, and b without optical OTF

Correlation with Subjective Data

The halftone images used in the visual experiments were evaluated using S-CIELAB and CVDM. The original contone image is used as the reference image. Both models output ΔE^* maps indicating the visual difference between the halftoned image and reference images. Figure 6 shows the original contone and error diffusion image and illustrates the ΔE^* maps from the two visual models. Both models capture the worming artifacts. CVDM predicts the texturing visibility produced by ED while the S-CIELAB map does not (e. g. in the bottom right patch shown).

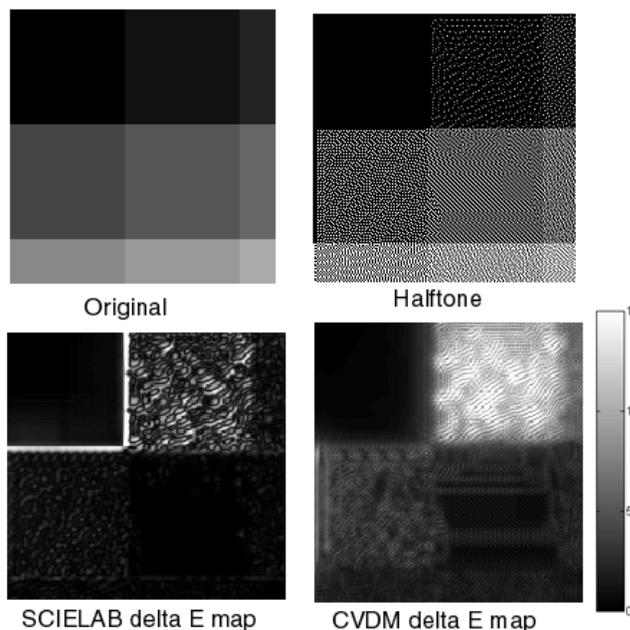


Figure 6. Predicted visual difference maps (upper left, original, upper right: halftoned image)

In the error diffusion images, a sharpening effect and a startup delay artifact occurs at the edges of some patches. S-CIELAB identifies this edge effect in the visible difference map while the CVDM maps do not. However, the effect was not noticeable in the images when viewed on the monitor from the observers' position. This difference between the models is probably due to the inclusion of masking effects in the CVDM model.

In order to perform a correlation study between the visual model and subjective experimental results, we convert the visual difference map to a single number. We considered several summary measures of the error distribution: Minkowski summation with exponents of 2, 3, and 4; and the 50th and 10th percentiles. We found that Minkowski summation with exponent of 2 (root-mean-square (RMS)) yielded the best correlation. Figure 7 plots the RMS ΔE^* of the visual models against the derived subjective quality scale. Each point represents a different halftone algorithm. The solid lines indicate the best linear regression fit of the model data to the subjective score. The correlation coefficients are 0.95 for CVDM and 0.89 for S-CIELAB.

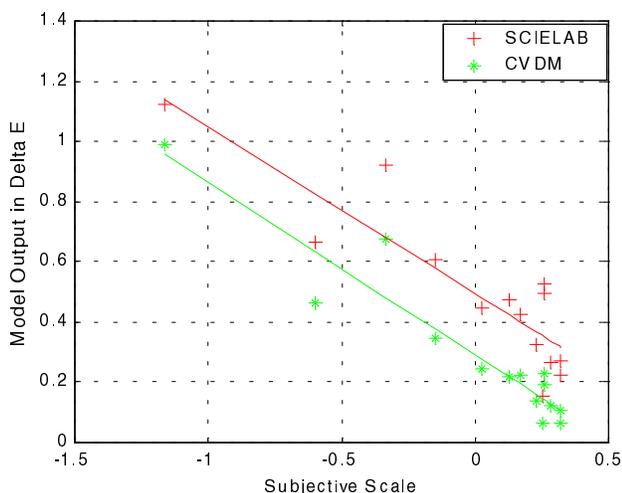


Figure 7. Correlation of subjective scale to model output for the 600 dpi condition

The model output RMS ΔE^* values are converted to subjective scale values using the regression fit. These scale values are plotted against halftone method used in Figure 8. The error bars indicate the 95% confidence interval for the subjective scale. 11 out of 14 CVDM predictions are within the 95% interval while 8 out of 14 S-CIELAB predictions are within the 95% interval.

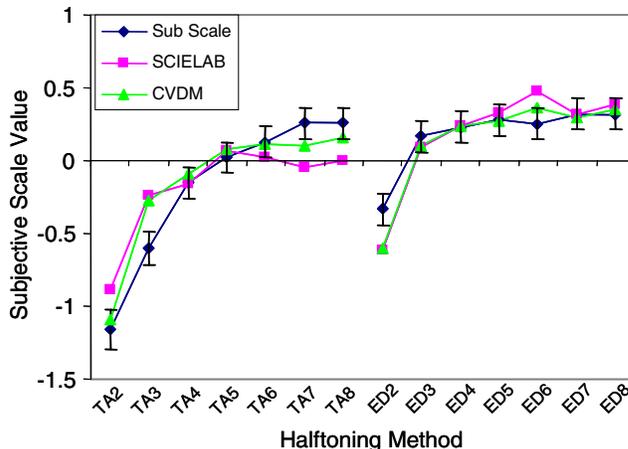


Figure 8. Comparison of model predictions with subjective scale for the 600 dpi condition.

Conclusion

We conducted a monitor-based subjective experiment by presenting halftoned images to observers. Images were presented pair-wise and observers indicated which image they thought smoother. The images represented a multilevel stochastic screen algorithm and multilevel Floyd-Steinberg error diffusion. A subjective scale was derived from the paired comparison data using Thurston's law of comparative judgment and this scale was correlated with the model-based metrics. The correlation was good for both S-CIELAB (0.89) and CVDM (0.95).

We modified the color visual difference model so that it is conceptually closer to the visual system; the modification also fixed two problematic issues with an earlier version of the model. The new model correctly predicts halftone artifacts that are difficult for the S-CIELAB single channel model to predict.

References

1. Jin, E., X.F. Feng, and J. Newell, The development of color visual difference model (CVDM), in Proc. of IS&T PICS, pp. 154. (1998).
2. Zhang, X. and B.A. Wandell, "A spatial extension of CIELAB for digital color image reproduction", SID Symposium Technical Digest, 27, pp. 731-734 (1996).
3. Yu, Q. and K.J. Parker, Quality issues in blue noise halftoning, Proc. of SPIE, vol. 3300: pp. 376-385. (1998).
4. Zhang, X., D.A. Silverstein, J.E. Farrell, and B.A. Wandell, Color image quality metric S-CIELab and its application on halftone texture visibility, in Proc. of SPIE, (1997).
5. Daly, S. The visible differences predictor: an algorithm for the assessment of image fidelity, in Digital Images and Human Vision, A.B. Watson, Editor. MIT Press: Cambridge, MA. pp. 179-206, (1993).
6. Thurstone, L. L. The Measurement of Values. Univ. of Chicago Press, Chicago (1959).

7. Guilford, J. P. *Psychometric Methods*. McGraw-Hill, New York (1954).
8. Critchlow, D. E. and M. A. *Psychometrika*, 56, pp. 517-533, (1991).
9. Mullen, K T. J. *Physiol.*, 359, pp. 381-400, (1985).
10. Feng, X. and S. Daly, Vision-based strategy to reduce the perceived color misregistration of image capturing devices, to be published in the *Proc. of IEEE*, January, (2002).
11. Klassen, R. V. and N. Goodman, Human chromatic contrast sensitivity: exploration of dependence on mean color. *Proc. of IS&T/SID 8th Color Imaging Conference*, pp. 31-37, (2000).
12. Barten, P. G., *Contrast sensitivity of the human eye and its effect on image quality*. Ch. 3, pp. 25-64, HV Press, (1999).
13. Marimont D. H. and B. A. Wandell., *J. Opt. Soc. Am. A*, 11, pp. 3113-3122, (1994)

Biography

Xiao-fan Feng received the B.S. degree in 1983 from Zhejiang University in Hangzhou, China, ME degree in

electro-optics from the Chinese Academy of Science in 1986, MS and Ph.D. degrees in imaging science from Rochester Institute of Technology in 1990 and 1995 respectively. From 1993 to 1997, he worked at Xerox Corporation in Webster NY. Since 1997 he has worked at Sharp Labs of America on image processing algorithms, display quality optimization, and visual modeling.

Jon Speigle received his B.S. degree in Aeronautical Engineering from M.I.T. in Cambridge, MA, in 1991 and a Ph.D. in Experimental Psychology from the University of California, Santa Barbara, in 1997. Since 1997 he has worked at Sharp Laboratories of America in Camas, WA, on algorithms related to digital color copiers and printers. He is a member of the IS&T and Optical Society of America.

Atsuhisa Morimoto received both B.S. and M.S. degrees in Electrical Engineering and Computer Science from Nagasaki University, Japan. He joined Sharp Corporation, Japan in 1997 and moved to Sharp Labs. of America, Inc. in 2000. His research interests include image quality and human vision. He is a member of the IS&T.