Perception and Modeling of Halftone Image Quality Using a High-Resolution LCD

Changmeng Liu, Garrett M. Johnson, and Mark D. Fairchild, Munsell Color Science Laboratory, Rochester Institute of Technology, Rochester, New York, USA
Gustav Braun, Eastman Kodak Company, Rochester, New York, USA

Abstract
An experiment was performed to test the usability of a high-resolution LCD display for examining image quality of halftone algorithms. This experiment was also designed to evaluate the usefulness of an image appearance model in predicting the visibility and/or the perceptibility of noise in various halftone imaging techniques. In this process a paired comparison psychophysical experiment was developed to collect subjective quality data on noise levels for the various halftoning techniques studied. Next an image appearance model (iCAM) was used to create an objective noise metric. This appearance model was applied with the addition of a spatial filter approximating the human CSF. The experimental results show that the simulated halftone image quality is similar to expectations from experience with printed images and that an image difference metric can predict the general trends in algorithm preference.

Introduction
"Digital halftoning is a method of reducing the number or gray levels or colors in a digital image while maintaining the visual illusion that the image is still continuous tone." [1] A halftoned image is typically displayed on media that cannot reproduce a continuous range of levels or colors. There are many types of halftoning technologies that have been developed over the last 40 years. The most visible characteristic of halftone algorithm is typically the halftone structure. There are generally two types of halftone structure; amplitude modulated (AM) and frequency modulated (FM). Cluster dot dithering algorithms is an example of an technique that generates an AM halftone structure. FM halftone structures can be found in blue-noise dithering, dispersed dot dithering and error diffusion. Error diffusion is an adaptive algorithm, so there is no fixed pattern for any given gray level.

Previous research[1] and practical experience shows that blue-noise dithering can eliminate moiré, generate more image detail and sharpness, and provide smoother gradations without contours. That process is especially effective for devices with low spatial resolution. As such, blue-noise dithering is often considered to be better than cluster dot dithering and disperse dot dithering for low resolution. Error diffusion can reduce quantization errors by distributing errors into local neighborhoods. It has similar properties as blue-noise dithering. Additionally there is some amount of edge enhancement inherent in the error diffusion algorithm. As such, these techniques are found to often have better image quality performance than simple blue-noise dithering.

Printers have different strengths and weaknesses compared to soft-copy displays, as well as compared to different print technologies. There are many issues related to real printing processes such as dot gain, registration of three, four, or multi-color separations, noise level of the printer, addressability of the printer, and repeatability of the printer. In this research only dot gain was considered during generation of the halftoned images, because the purpose of this research is evaluation of softcopy simulation of general image quality of halftone algorithms but not a direct test of any specific printer. Dot gain is a term describing an increase in darkness of a halftone print as compared to what was intended by the binary digital pattern that was produced.

An additional goal of this experiment was to evaluate the ability of an image appearance model to predict the perceived noise levels across various halftone imaging techniques. To achieve these goals a psychophysical experiment was performed to obtain a subjective quality scale of noise perception for a typical set of halftoning algorithms. In order to assess the halftoning processes themselves, independent of printing artifacts such as dot-placement errors, differential gloss, and banding, it was decided that the halftoned images would be simulated on a high resolution LCD display.

Although the LCD display has different properties from real printers, it is hoped that it is possible to simulate the halftoned image on a high-resolution LCD display through proper characterization. In this research, each dot in the halftoned image was simulated by round ink drops with dot gain simulation. There are some advantages to test the quality of halftone algorithm by using softcopy of the halftoned image. First of all, it can be less expensive and time-consuming. Secondly, all individual printer variance can be eliminated from the softcopy; therefore the quality result only depends on the halftone algorithm themselves, but not on the device characteristics. Thirdly, it is easy to design the psychophysical experiment to test the image quality and is easy for observer to perform the experiment. There are two obvious disadvantages of using softcopy to test halftone algorithm. The first is the inherent gamut differences between the display technologies. In addition it is very difficult to simulate the high spatial resolutions that printers are capable of on a typical desktop CRT or LCD display. In this research, a psychophysical experiment was implemented to test if the softcopy of halftoned image can provide image quality scale that matches practical experiences.

The scale of overall quality of the halftoned images was based on how "pleasing" the images looked to observers. An ideal image quality model should predict the observer's response. iCAM[2] is one type of model applicable to image appearance, image rendering, and image quality specifications and evaluations. In this research, iCAM served as an image quality model that considers the spatial properties of the image. The IPT color space[3] was used in iCAM to specify the image attributes. The overall Euclidean difference in IPT is referred to as $\Delta \mathrm{Im}$, which can be considered the overall image difference. Image quality metrics can be derived from image difference metrics. For these image quality metrics, viewing-distance-dependent spatial filtering and spatial frequency adaptation is applied in a linear orthogonal color-space and then differences are computed in the nonlinear IPT space. Euclidean summations of these differences can be used as an overall image difference map and then various summary statistics can be used to predict different attributes of image difference and quality. The contrast sensitivity function in the iCAM image quality model serves to modulate spatial frequencies that are not perceptible, and enhance certain frequencies that are most perceptible. Spatial frequency adaptation essentially decreases sensitivity to certain frequencies based upon information present in the image.[4] Since halftoned images contain a large amount of fixed spatial frequency components, spatial frequency adaptation should be considered in the quality model.

In previous research, Zhang et al.[5] used the S-CIELAB model to predict the halftone image quality with some success. It was expected that the iCAM model, which is based in part on the S-CIELAB structure, should perform similarly if not better.

Generating Halftone Images

In this experiment the halftoned images were created using a spectrally based halftone-dot generator. This process simulated physical dot overlap using round ink drops that were scaled to overfill the raster on a 400-dpi grid by approximately 120 percent. The spectral reflectances of the CMYK ink dots were scaled such that the integrated spectral reflectance over a one-inch (400x400 pixel) solid area scaled to CIELAB values close to typical measured printed samples. Using these scaled dot reflectances a CMYK color characterization target was processed through the model to produce a set of CIE Illuminant D50 referred L*a*b* values for each of the halftoning techniques studied. An ICC profile was generated for each halftoning technique using the Kodak Color*flow* ProfileEditor software using a consistent UCR approach for all of the halftoning techniques.

Using the ICC profiles and the halftone model parameters established in the color-characterization phase outlined above, a set of sRGB images were first processed to CMYK code values and then to binary halftoned L*a*b* images using the spectrally based halftone-dot generator. The images were then rendered to the LCD display device using a colorimetric characterization of the display. Out of gamut L*a*b* values were simply clipped to display gamut in the RGB code value space of the display. In general, this clipping process had a very minimal impact on the colorimetry of the images, since all images began in the sRGB space, and the clipping was consistent across all of the various halftoning techniques.

The halftoning techniques used in this study were selected to cover a wide range of spatial frequency characteristics. Three of the five techniques were classified as *blue-noise* processes [6 7 8 9 10] (e.g., a color decorrelated blue noise dithering process, a dot-on-dot blue noise dithering process, and a Floyd Steinburg error diffusion process.) The two remaining halftoning techniques were a center growing cluster-dot[11] and a Bayer dithering process.[12] The cluster-dot algorithm was an 80 line-per-inch (lpi) screen on a 400 dot-per-inch (dpi) raster. The screen angles selected for this technique were 15, 135, 0, and 45 degrees for the C, M, Y, and K channels respectively. The CMYK continuous tone images were halftoned using the Adobe Photoshop 7.0 *Image->Mode->Binary* conversion function. The Bayer dithering was performed using a 16x16 pixel dither matrix where the same dither matrix was applied to all four of the image channels.

Experiment

An exploratory psychophysical experiment was designed to evaluate different halftone algorithms based on the spatial properties, such as the smoothness, uniformity, sharpness, contour artifacts, contrast and noisiness. The softcopy halftoned images described above were used in this experiment. The test targets were displayed on an IBM T221 high-resolution (200 dpi) LCD display. Two different viewing distances were used in an attempt to simulate different resolutions of the hardcopy halftoned images. A paired comparison experiment was performed. The observers were asked to choose which of two images had better image quality based on the spatial properties of the images. Observers were asked to ignore any large-scale color differences, if any, that were present in the images.

Image Selection and Evaluation Consideration

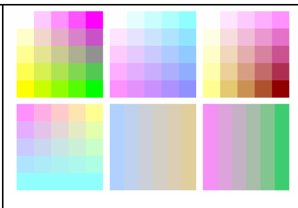
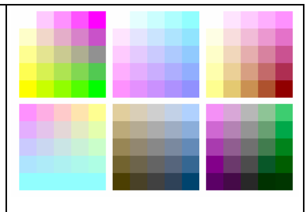
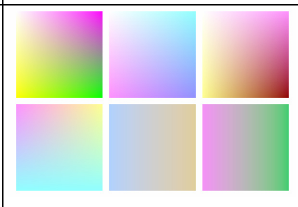
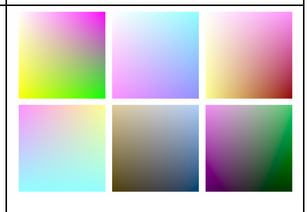
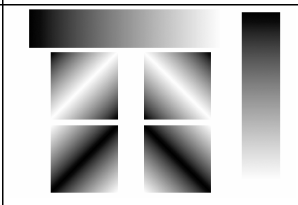
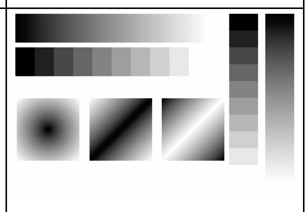
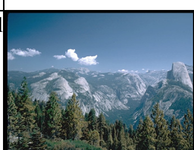
It is important to carefully select the image scenes when comparing halftoning algorithms. To test the smoothness and contour artifact, the test-targets should include the gray and color sweep in different directions. In order to test the uniformity, noisiness, sharpness and contrast, both gray and color patches should be included in the target images. To test the overall perception of complex scenes, complex pictorial colored images should be included in the target set.

Based on these considerations, nine images were selected as the test images. These images included two colored-patch images, two colored-sweep images, two gray-target images that contained both gray patches and gray sweeps of various directions, and three pictorial images. Table 1 shows the nine test images.

Psychophysical Experiment

An IBM T221 LCD was used to display the softcopy halftoned images. This 22" LCD display is 3840 by 2400 pixels. The native resolution of the LCD is 200 dpi. The native white point (x, y, Y) of the LCD was (0.3010, 0.3374, 250cd/m2). The LCD display was characterized with the colorimetric characterization model presented by Day et al. [13]

Table 1. Target Images

Color Patch Targets			
Color Sweep Targets			
Gray Targets			
Pictorial Images			

The IBM LCD allowed a 2400 by 1600 image displayed at one time. The physical size of each displayed image was 12" by 8". The image was presented with a 20% gray of the native LCD white point background in a darkened room. The paired comparison experiment was separated into two sessions. The viewing distance was 40 inches for the first session and 120 inches for the second one. We assumed that the normal viewing distance for hardcopy prints was 20 inches. The softcopy image with 200 dpi native resolution will approximately subtend the same visual angle as a 400-dpi image when viewed at 40 inches, which is twice the normal viewing distance. And the image will simulate 1200-dpi printer when viewed at 120 inches, which is 6 times the normal viewing distance.

Five different halftoning processes were evaluated in this experiment. They were blue noise dithering process (BN), dot-on-dot blue noise dithering process (BDDoD), Floyd Steinburg error diffusion process (ED), cluster-dot dithering (CD), and Bayer dithering (BD). The abbreviation in the parenthesis will be used in the remainder of this paper to refer to the corresponding halftone process. There were nine different test targets and five different halftoning processes. The observers were asked to compare the image quality among the same target processed with different halftoning techniques. There were a total of 90 comparisons for each observer. The presenting order of these 90 pairs was randomized. Observers' decisions were only based on the spatial properties of the image and not on the color difference of the image in order to minimize the gamut differences or color characterization problems of the algorithms. Twenty observers with varying imaging experience served as subjects for both sessions. All observers had normal color vision and normal or corrected-to-normal visual acuity when they performed the experiment.

The original continuous image was encoded as sRGB image, so the image color gamut should be mostly inside the LCD's gamut. The halftoned image were binary L*a*b* images. They were defined under CIE Illuminant D50. These L*a*b* images were transformed to the LCD display white point using a linear von Kries type transformation[14, 15] from CIECAM02 with a complete adaptation factor (D=1). The chromatically adapted L*a*b* images were transformed to RGB images through the LCD colorimetric characterization model.

Results and Discussion

For this paired comparison experiment, Thurstone's Law of Comparative Judgments, Case V, was used to analyze the results. The proportional observer data were converted into an interval scale of image quality. Due to the large image quality difference between some halftoning processes, there were some zero-one proportion results in the experimental data. Those caused the unanimous judgments problems. Morrisey's incomplete matrix solution[16] was used in the data analysis to account for this problem. This solution is a linear regression technique to fill in the incomplete z-scale value matrix.

Experiment Results

Figure 1 shows the interval scale along with 95% confidence limits for each target in 400-dpi simulation (with 40 inches viewing distance). The interval scale in Figure 1 can be interpreted as an image quality scale because of the objective of the psychophysical experiment. The results in Figure 1 indicate that image dependency is present, but not very large for the 400-dpi simulations across all five halftoning processes. This also indicates that the four different target image types, which are colored patches, colored sweeps, gray targets, and pictorial images, are highly correlated with each other.

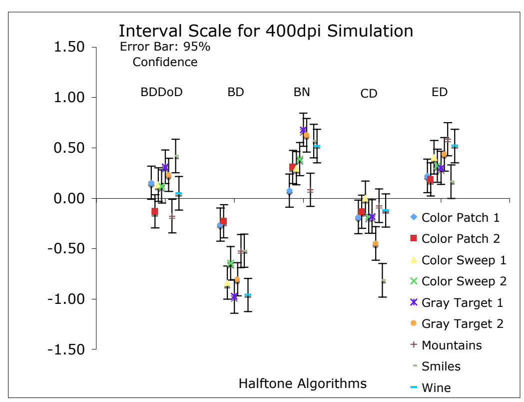


Figure 1. Interval Scores of halftoned image for nine target images in 400-dpi simulation

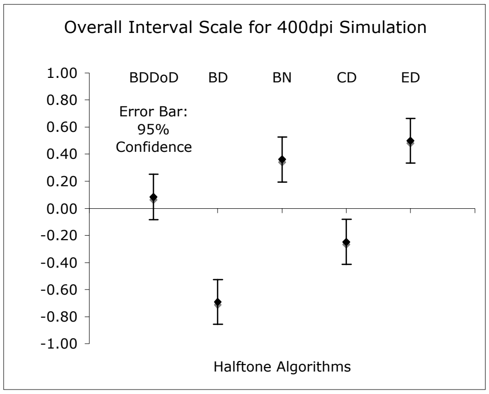


Figure 2. Average Interval Scales of halftoned image for all targets in 400-dpi simulation

Figure 2 shows the averaged interval scales along with 95% confidence limits across all nine targets for the 400-dpi simulation. The experimental result shows that all three blue-noise processes, which are the blue noise dithering process (BN), dot-on-dot blue noise dithering process (BDDoD) and error diffusion process (ED), are significantly better (95% confidence) than the cluster dot dithering (CD) and disperse dot dithering (BD). From practical experience, blue-noise dithering has been shown to eliminate moiré, generate more image detail and sharpness, and provide smoother gradations without contours. And it is especially effective for low resolution. The experimental results match these practical experiences. Within the blue-noise processing, the error diffusion is slightly higher quality than blue noise dithering but there is no significant different between these two process. This is reasonable because the error diffusion algorithms also have an edge enhancement effect. The interval scale value of the blue-noise dot-on-dot process is less than the blue-noise and error diffusion. The blue-noise dot-on-dot algorithm tends to show noise levels higher than the straight blue-noise dithering because it uses the same dither matrix for the CMYK channels. The blue-noise dithering process uses decorrelated masks that form the halftone pattern dot-off-dot for about 1/3 of the tone scale.

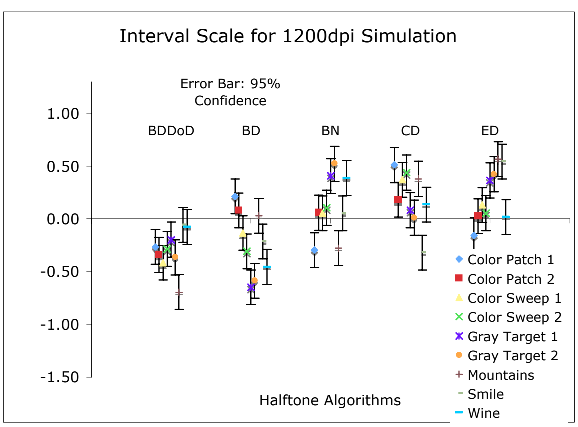


Figure 3. Interval Scores of halftoned image for nine target images in 1200-dpi simulation

Figure 3 shows the interval scale along with 95% confidence limits for each target under 1200-dpi simulation (at 120 inches viewing distance). Figure 4 shows the averaged interval scales along with 95% confidence limits among all nine target images in 1200-dpi simulation. The experimental results show that there was more confusion among the subjects at 1200 dpi viewing distance, though the error bars tend to indicate that the disperse dithering and the dot-on-dot blue noise process were of lowest quality. For the 1200-dpi simulation, the quality scale for the cluster dot dithering and disperse dot dithering is significant increased compared to the same algorithms in the 400-dpi simulation. This makes sense because the spatial frequency structure of the images is increased by increasing the viewing distance. This is equivalent to moving the low spatial frequency component in disperse dot dithering and cluster dot dithering to higher frequencies that are less visible to the human visual system (HVS).

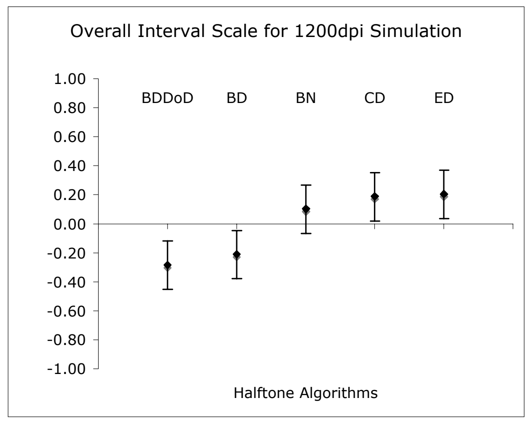
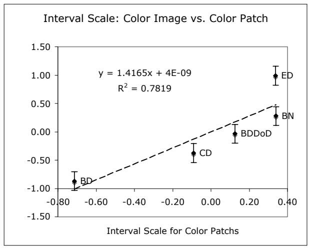


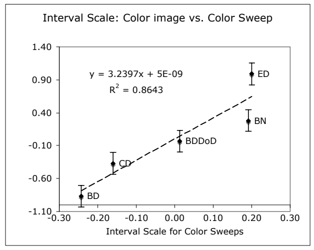
Figure 4. Average Interval Scales of halftoned image for all targets in 1200-dpi simulation

Correlation between Different Types of Target

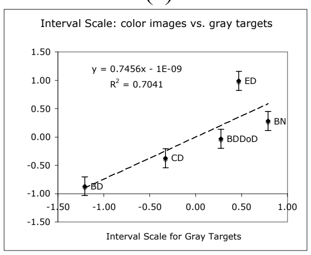
Figure 5 shows the correlation between the interval scales of the different types of test targets under 400-dpi simulation. Figure 5(a) shows the correlation between the pictorial images and color patches. Figure 5(b) shows the correlation between the pictorial images and color sweeps. Figure 5(c) shows correlation between the pictorial images and gray targets. In all three cases, the dashed line represents the best fit line through the quality scales of all five halftoning processes. This indicates that the error diffusion algorithm is an outlier in all three cases. If error diffusion is removed, the pictorial image interval scale is highly correlated to the other four types of target images. The bias of the error diffusion away from the fitting line indicates that error diffusion always has better performance than predicted by the test-targets, when it is applied on the pictorial images. This result suggest that the edge enhancement effect of error diffusion has more of an impact on pictorial images than other types of targets, while other halftone processes have no or less edge enhancement effect. The error diffused pictorial image appears sharper than the other four halftoning processes, especially for the lower resolution simulation. Pictorial images contain more high frequency information than color patches or color sweeps; therefore the sharpening effect is more easily detected in the pictorial images than in the color patches or color sweeps.



(a)



(b)



(c)

Figure 5. Correlation between different types of target images for 400-dpi simulation. (a) Interval Scale correlation between pictorial image and color patches. (b) Interval Scale correlation between pictorial image and color sweeps. (c) Interval Scale correlation between pictorial image and gray targets.

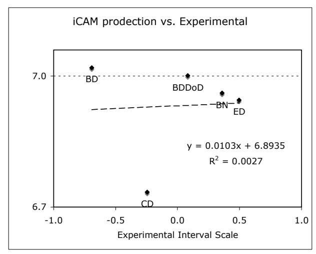
For the 1200-dpi simulation, the correlation between the pictorial images and other types of targets shows same trend as 400-dpi simulation, though it is not as strong. The error diffusion still shows bias from the fitting line and the quality scales of the pictorial images are under predicted by the test targets.

Predicting and Modeling Halftoned Image Quality

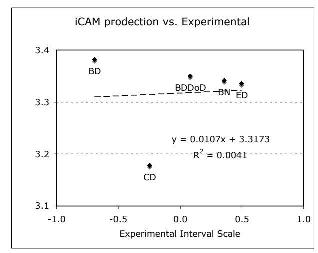
The image appearance model iCAM was used to predict the halftoned image quality scales. Spatial filters that approximate the contrast sensitivity functions of the human visual system were first applied. These filters were subjected to a spatial frequency adaptation, based upon the frequencies present in each image. The IPT color space used in iCAM is valid only for CIE Illuminant D65. The original continues tone image is encoded as sRGB image with a D65 whitepoint. The halftone images were transformed to CIELAB values under a D50 whitepoint, so a chromatic adaptation transform is needed to transfer the halftoned image from D50 to D65. The image difference was then calculated between the input continuous tone image and each of the halftoned images.

The default output of iCAM is a Euclidian color difference similar to CIE ΔE^*_{ab}. A lightness difference was also calculated in this research because the observers were asked to judge the image quality based only on the spatial properties. Image quality is thought to be inversely proportional to the average image difference from the continuous image, so the reciprocal of the image difference is used as a metric of image quality. It has also been shown that large color differences or artifacts can overwhelmingly determine the image quality in some circumstance, while the average image difference does not. Therefore the average and 95th percentile of color and lightness difference are calculated.

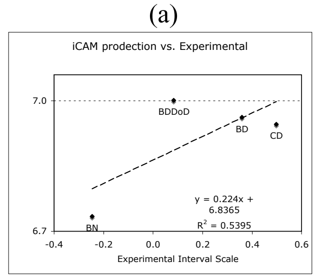
Figure 6 shows the correlation between iCAM color difference prediction and the experimental interval scale for the 400-dpi simulation. Figure 6 (a) shows the average color difference prediction; Figure 6 (b) shows the 95th percentile color difference prediction. Figure 6 (c) and (d) plot the same relationship as in Figure 6 (a) and (b) except a single algorithm determined to be an outlier, the Bayer Dither (BD) has been removed. The reciprocal of the color difference prediction is used in all the plots in order to positively correlate with the image quality interval scale. After removing the outlier BD algorithm, the results show that iCAM can predict that the blue-noise processes (BDDoD, BN, and ED) have higher quality than the cluster dot dithering process (CD). The color difference of the outlier, BD, has been over predicted by iCAM. This large difference may be caused by an overall color shift in the BD images, which was measured by the image difference metric but ignored by the observers. The observers were instructed to focus mainly on spatial intensity differences.



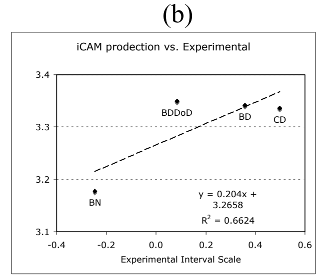
(a)



(b)

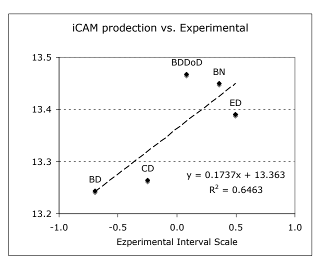


(c)

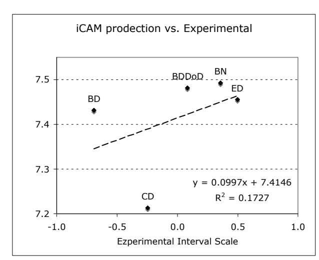


(d)

Figure 6. Fitting iCAM color difference prediction and the experimental interval scale for 400-dpi simulation. (a) Fitting the average color difference and experimental interval scale. (b) Fitting 95% color difference and experimental interval scale. (c) As same as (a) but remove the outlier BD. (d) As same as (b) but remove the outlier BD.



(a)

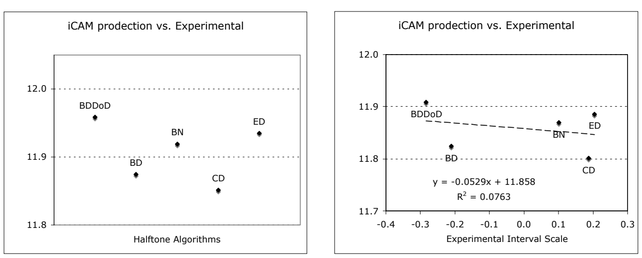


(b)

Figure 7. Fitting iCAM lightness difference prediction and the experimental interval scale for 400-dpi simulation. (a) Fitting the average lightness difference and experimental interval scale. (b) Fitting 95% lightness difference and experimental interval scale.

Figure 7 shows the correlation between the iCAM lightness difference prediction and the experimental interval scale for the 400-dpi simulation. Figure 7 (a) shows the average lightness difference prediction; Figure 7 (b) shows the 95% lightness difference prediction. Lightness prediction result shows the same trend as the color difference prediction without the obvious outlier. The difference between the blue-noise processes and the ordered dithering is higher in lightness difference prediction than in color difference prediction. This suggests that the lightness difference predictions fit the experimental data better than the color difference. This is reasonable because the observers judged the image quality mainly based on the spatial property of the image, and were told to ignore large color differences. The result also shows that the 95th percentile does not improve predictions when compared to the average color or lightness differences.

Figure 8 (a) shows the iCAM average color difference prediction for five halftone algorithms for the 1200-dpi simulation. Figure 8 (b) shows correlation between iCAM model lightness difference prediction and experimental interval scale for 1200-dpi simulation. In this case, iCAM was unable to fit the experimental data well. It is interesting to note from Figure 8 that the two algorithms that are predicted poorly are the cluster dot (CD) and blue-noise dither dot-on-dot (BDDoD). The cluster dot image difference is consistently too high, while the BDDoD is too low. The possible reasons for the failure is that the CSF low-pass filter used in iCAM is not appropriate for this viewing distance. This discrepancy might also be explained by approximate size constancy that has recently been observed in image quality experiments using increased viewing distances to simulate increased spatial resolution. The iCAM model does not yet predict such an effect.



(a) (b)

Figure 8. (a) iCAM predicted color difference for all five algorithms for 1200-dpi simulation. (b)Fitting iCAM lightness difference prediction and the experimental interval scale for 1200-dpi simulation.

Conclusions

A paired comparison psychophysical experiment was developed to simulate and evaluate the performance of five halftoning algorithms using softcopy on a high-resolution LCD display. The results shows the blue-noise processes have better image quality than the ordered dithering processes under low resolution (400-dpi). Image quality of cluster dot dithering and disperse dot dithering was increased to the same level as blue-noise processes when the resolution increased from 400-dpi to 1200-dpi. Error diffusion shows that it works better in complex pictorial image than the simple color patches or sweeps because of its edge enhancement properties. These experimental results agree with image quality result from previous research and practical experience, suggesting that soft-copy displays can adequately simulate hardcopy prints. iCAM image appearance model was used to predict the psychophysical experiment result. For the 400-dpi simulation, iCAM predicts that the blue-noise processes have better performance than the cluster dot dithering and dispersed dithering. For the 1200-dpi simulation, iCAM failed to predict the average experimental results well, perhaps because inappropriate CSFs were used for the large viewing distance.

Acknowledgements

This research was supported by Eastman Kodak Company and the NYSTAR CAT in Electronic Imaging Systems.

References

1. G. Sharma, Digital Color Imaging Handbook, Chp. 6, CRC Press LLC, (2003)
2. M. D. Fairchild and G.M. Johnson, The iCAM framework for image appearance, image differences, and image quality, *J. of Electronic Imaging* 12, 126-138 (2004)
3. F. Ebner and M.D. Fairchild, Development and testing of a color space (IPT) with improved hue uniformity, *IS&T/SID 6th color imaging Conference,* Scottsdale, 8-13 (1998)
4. G.M. Johnson and M.D. Fairchild, On contrast sensitivity in an image difference model, *Proc. of IS&T PICS Conference*, 18-23 (2001)
5. X. Zhang, D. A. Silverstein, and J. E. Farrell, Color Image Quality Metric S-CIELAB and Its Application on Halftone Texture visibility, Proceedings of the IEEE CompCon, (1997)
6. R. W. Floyd and L. Steinberg, An adaptive algorithm for spatial gray-scale, *Proc. Soc. Inf. Disp.* 17/2, 72-77 (1976)
7] K. Knox, Evolution of Error Diffusion, *Proc. SPIE*, vol. 3648, 448-458, 1999
8. R. Ulichney, Dithering with blue noise, *Proc. IEEE,* vol. 76, no. 1, pp. 56-79, 1988
9. T. Mitsa and K.J. Parker, Digital halftoning technique using a blue-noise-mask, *J. Opt. Soc. Am. A,* vol. 9, pp. 1920-1929 (1992)
10. J. F. Jarvis, C. N. Judice, and W. H. Ninke, A survey of techniques for the display of continuous tone pictures on bilevel displays, *Comput. Graph. Image Process.* 5, 12-40 (1976)
11. R. Ulichney, Digital Halftoning, The MIT Press, Cambridge (1987)
12] B.E. Bayer, An optimum method for two level rendition of continuous-tone pictures, *Proc. IEEE Int. Conf. Commun., Conference Record*, pp. (26-11)-(26-15), (1973)
13] E.A. Day, L.A. Taplin, and R.S. Roy, Colorimetric Characterization of a Computer-Controlled Liquid Crystal Display, *Color Res. Appl.* (2004)
14. M.D. Fairchild, A Revision of CIECAM97s for Practical Applications, *Color Res. Appl. 26, 418-427* (2001)
15. N. Moroney, M.D. Fairchild, R.W.G. Hunt, C.J. Li, M.R. Luo, and T. Newman, The CIECAM02 Color Appearance Model, *IS&T/SID 10th Color Imaging Conference, Scottsdale,* (2002)
16. P. Engeldrum, Psychometric Scaling: a Toolkit for Imaging Systems Development, Imcoted Press, Winchester, (2000)

Author Biography

Changmeng Liu received his B.S. degree in Electronic Engineering Department from TianJin University (China) in 1996 and M.S. degree in Computer Science Department from the Rochester Institute of Technology in 2002. He is a Ph.D. candidate in the Munsell Color Science Laboratory at the Chester F. Carlson Center for Imaging Science at the Rochester Institute of Technology. His current interests include color appearance, image quality, color management, and digital Halftoning.