Evaluation of HDR Tone Mapping Algorithms using a High-Dynamic-Range Display to Emulate Real Scenes

Jiangtao Kuang, Rod Heckaman, Mark D. Fairchild, Rochester Institute of Technology, Rochester, New York

Abstract

Current HDR display technology approaches the dynamic range of the fully adapted human vision system. As such, this technology has potential for performing as a surrogate for real-world scenes in the evaluation of the accuracy of high dynamic range (HDR) algorithms that map HDR scenes to the more common, everyday display technology having limited dynamic range. Clearly, use of HDR display technology has the benefit of simplicity in experimental design while maintaining the high dynamic range of the original scene. To evaluate this potential use for HDR displays, seven published versions of well-known tone mapping algorithms were benchmarked for rendering accuracy against each of four real-world scenes constructed in the lab and their corresponding renderings on the Munsell Color Science Laboratory's (MCSL) HDR display. The results between those obtained from the MCSL HDR display and those obtained from the real-world scenes are in good agreement thus validating the HDR display's potential as an evaluation tool in this context.

Introduction

In the last decade, many tone-mapping algorithms have been developed to reproduce the high dynamic range (HDR), real-world scenes onto display devices that are only capable of low dynamic range luminance. When a new algorithm with its corresponding tone mapping algorithms is proposed, it is necessary to benchmark its performance against existing algorithms using commonly accepted methodologies. The most straightforward of these methodologies is to directly compare the tone-mapped image with its real-world counterpart. However, it is difficult to maintain repeatability of such real-world scenes in a well-controlled environment to say nothing of including people or outdoor vistas in the lab.

Currently, HDR display technology approaches the dynamic range of the fully adapted human vision system and, as such, offers the potential for performing as a surrogate for real-world scenes in the evaluation of the accuracy of the tone-mapping operators. Furthermore, this potential was demonstrated by Ledda, *et. al.*, 2005)[1] in a similar evaluation of tone mapping algorithms using a HDR display. Clearly, use of this technology has the benefit of simplicity in experimental design while maintaining the high dynamic range of the original scene.

The aim of this paper is to validate this potential. To this end, seven algorithms - the bilateral filter,[2] photographic reproduction,[3] histogram equalization,[4] iCAM,[5] iCAM06[6] and commercial software tools, Exposure & Gamma, and Local Adaptation from Photoshop CS2 - were evaluated according to a previously published methodology[7] for rendering accuracy against four real-world scene constructed in the lab and each of their renderings on the MCSL HDR display.

The MCSL HDR Display

The MCSL, projector-based HDR display was used in this evaluation as its dynamic range (over five orders of magnitude) approaches the range the fully adapted, human visual system. This display was built based on the technology developed at the Structured Surface Physics Laboratory of the University of British Columbia and ultimately Brightside Technologies, Inc.

The Brightside technology was first introduced in the form of a DLP projector modified to project only a modulated luminance channel and a LCD panel that, in turn, modulates the projector beam into three RGB channels. The result is a very bright image – 2,700 cd/m^2 as reported by Brightside and a very low measured black level giving contrast ratios of 54,000:1. [8]

In the Brightside configuration, perfect alignment between the projector's pixels and those of the LCD panel is not possible, and a moiré pattern across the viewing field can be a problem. Because the Brightside display is intended for more general use, they have chosen to defocus the projector so that the pattern is not visible. To restore sharpness, the luminance channel is split between the projector and the LCD where it is inverse-filtered spatially thereby de-saturating the display's color channels and reducing color gamut.[7.8]

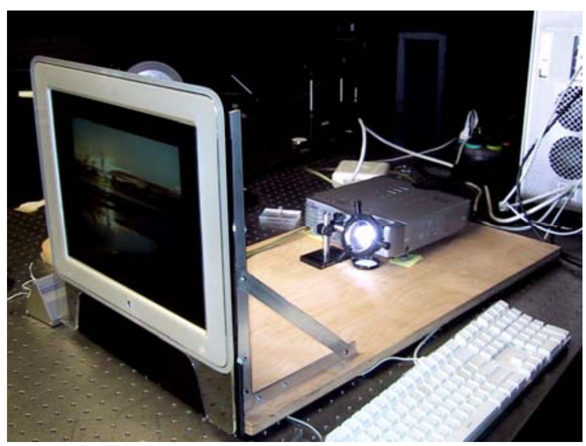

Figure 1: MCSL HDR Display

The MCSL version (Figure 1) is intended for experimental purposes with one observer whose viewing position remains relatively constant, and moiré between the projector pixels and the LCD pixels has not been seen as a problem. Hence, the projector is focused on the LCD relieving the LCD of the burden of supplying a luminance component for sharpness preservation, and maximum color gamut is available.

MCSL HDR Display Characterization and Performance

The MCSL HDR display was characterized according to the Equation 1[9,10] using two series of ramps – a projector series with the LCD full on and an LCD RGB series in digital counts with the projector full on. display was characterized to within an average DE94 of 1.0 and a standard deviation of 0.67 to the CIE color matching functions for the 1931 observer and illuminant D65 using a LMT C 1210 Colorimeter. A LMT C 1210 Colorimeter was used to measure the XYZ data for each of the ramps according to the CIE color matching functions for the 1931 observer and illuminant D65.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = PM \begin{bmatrix} R \\ G \\ B \\ 1 \end{bmatrix} \quad (1)$$

The vector RGB represents the augmented, scalar input values to the LCD panel obtained from look up tables (LUTs) in RGB digital counts versus scalar value as derived from the measured LCD ramps with the projector fully on. The factor P is the scalar attenuation of the full output of the projector obtained from the projector LUT in scalar luminance versus projector digital counts as measured from the projector ramp. Figure 2 plots the projector scalar value or attenuation factor P in units of relative luminance and the effective LCD RGB scalars with the projector fully on as a function of digital counts.

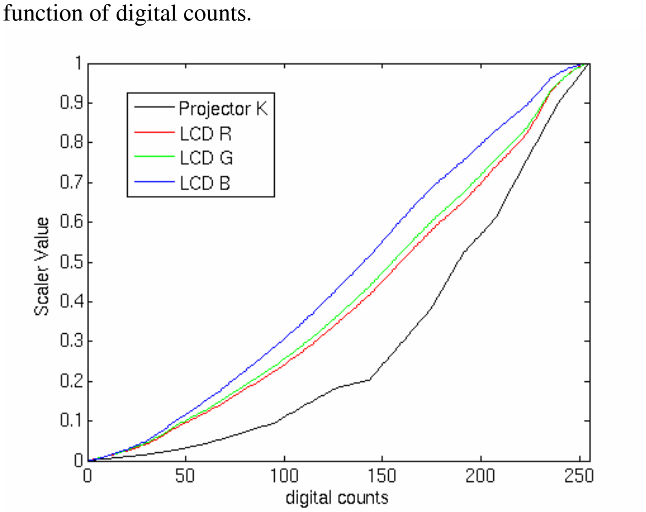

Figure 2: Projector scalar value in relative luminance and effective LCD RGB scalar values, projector full on, as a function of digital counts

The matrix M (Equation 1 and 2) performs the transformation of the RGB and projector scalar values to CIEXYZ tristimulus values. The matrix is constructed from the measured values of maximum $XYZ_{RGB,max}$ and minimum $XYZ_{RGB,min}$ for each channel of the display when backlit by the full output of the projector (P=1.0).

$$M = \begin{bmatrix} X_{r,\max}-X_{k,\min} & X_{g,\max}-X_{k,\min} & X_{b,\max}-X_{k,\min} & X_{k,\min} \\ Y_{r,\max}-Y_{k,\min} & Y_{g,\max}-Y_{k,\min} & Y_{b,\max}-Y_{k,\min} & Y_{k,\min} \\ Z_{r,\max}-Z_{k,\min} & Z_{g,\max}-Z_{k,\min} & Z_{b,\max}-Z_{k,\min} & Z_{k,\min} \end{bmatrix} \quad (2)$$

The corresponding chromaticities and absolute luminances as derived from the measured chromaticities are given in Table 1 for each channel of the display with the backlight or projector fully on.

Table 1. The maximum CIE chromaticities, x and y, and absolute luminance Y for each channel of the display and the display's white point and black point with the projector full on.

	x	y	Y(cd/m2)
R	0.6333	0.3528	278.89
G	0.3423	0.5920	1224.55
B	0.1427	0.0923	218.49
White Point	0.2959	0.3295	1721.93
Black Point	0.2991	0.3446	13.01

At maximum attenuation of the projector output, the overall dynamic range of the display is computed as 114,000:1 with a computed minimum black level of 0.015 cd/m^2.

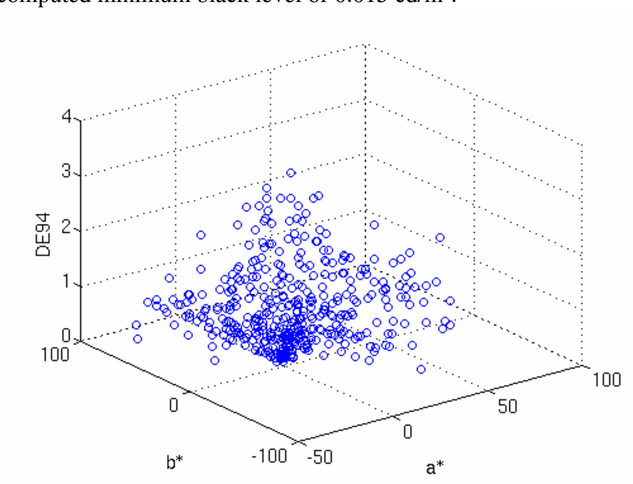

Figure 3. *Scatter plot of CIEDE94 versus CIE a* and b* from 400 random samples*

Figures 3 illustrates a scatter plot in CIE a*b* of CIEDE94 for 400 randomly sampled, measured XYZ data and their predicted values from the characterization. For this data, the mean CIEDE94 was 1.05 with a standard deviation of 0.70, and the distribution of CIEDE94 values seem, for all practical purposes, independent of their value in a*b*.

In the rendering of an image for display, the XYZ image data is first linearly scaled to the entire dynamic range of the display. These scaled XYZ values are then converted to projector and LCD RGB scalars using the inverse of Equation 1. Because the projector scalar P and the LCD scalars RGB are not uniquely determined, the additional constraint imposed on the projector is that it always assumes as much of the burden of producing luminance as possible so that color gamut is preserved. Figures 4 illustrate the result in digital counts as a function of log metric lightness for each of the projector channel (K) and the LCD channel (R=G=B).

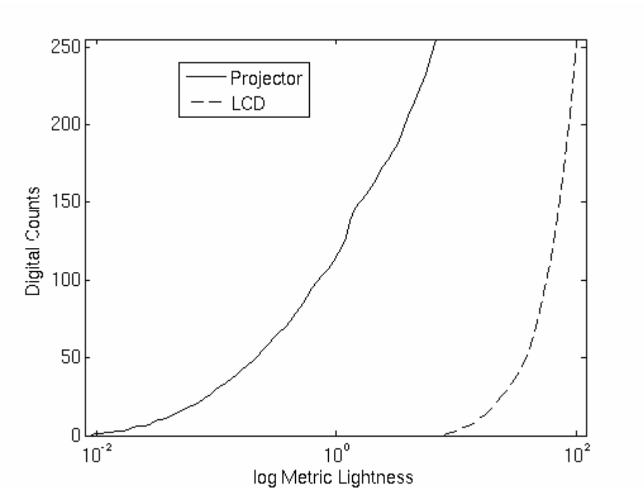

Figure 4. *MCSL HDR display projector and LCD (RGB) digital counts as a function of log metric lightness*

Finally, Figure 5 illustrates the gamut of the MCSL HDR display in CIELAB where the display's white point set to its maximum luminance.

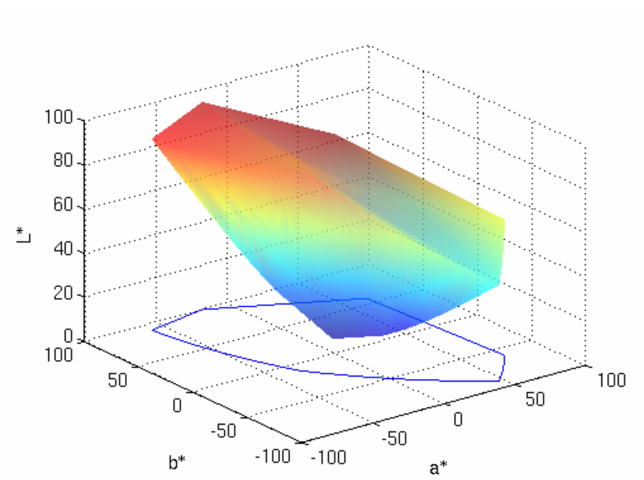

Figure 5. *MCSL HDR Display gamut in CIELAB*

Experimental

Four HDR real-world scenes (Figure 6) with a variety of dynamic ranges and spatial configurations[7] were designed and constructed in the lab. These scenes were then captured by a colorimetrically characterized Nikon D2X digital camera using the multiple exposure method.[11]

Two psychophysical experiments were conducted using the method of paired comparison. In both experiments, the HDR algorithm's, rendered results were displayed on a colorimetrically characterized 23-inch Apple Cinema HD LCD Display with a maximum luminance of 180 cd/m2 on a gray background with a luminance of 20% of the adapting white point. Twenty-three observers with normal color vision took part in the experiments.

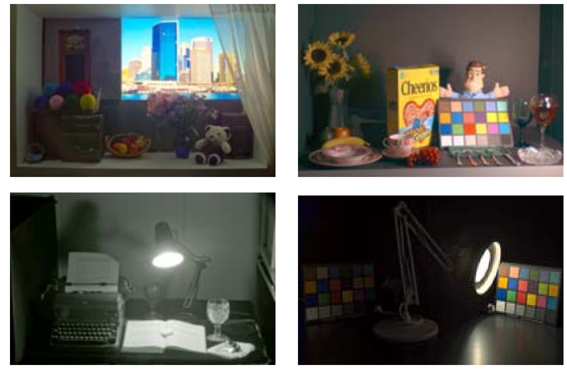

Figure 6. *Experimental scenes: (a) window (b) breakfast (c) desk (d) Double Checkers*

In the first experiment (Figure 7), tone-mapped images displayed on a desktop, low dynamic range LCD monitor were compared against the real-world scenes. Participants were asked to stand in a position where the viewing angles for the physical scenes were the same as those for the images on the display. For every seven pairs of evaluation, they were obligated to look and remember the appearance of the scene for at least 30 seconds, and return to the display to make their evaluation after a 20-second of adaptation period.

Figure 7. *Experimental setup for accuracy evaluation using real-world scenes*

In the second experiment (Figure 8), the HDR images were linearly scaled and rendered on the HDR display as described in the above to serve as the surrogate for real-world scenes in the first experiment. Observers were asked to compare the appearance of a pair of simultaneously displayed, tone-mapped images on the desktop, low dynamic range LCD monitor with the corresponding HDR image on the HDR display and select which of the two monitor images more closely resembled the one on the HDR display. They were allowed to look back-and-forth between the monitor and the display for their judgments.

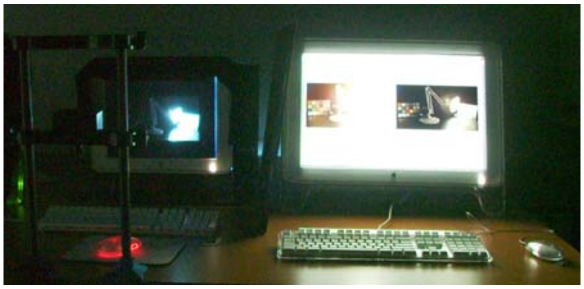

Figure 8. *Experimental setup for tone-mapping algorithm evaluation using HDR display*

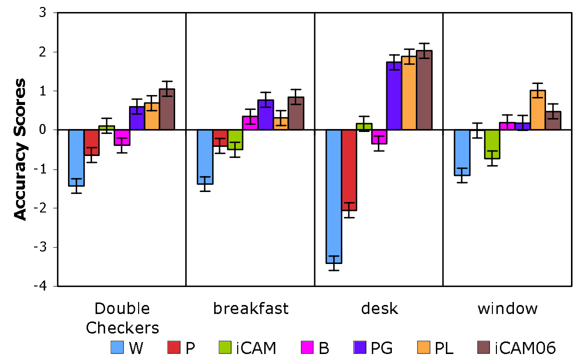

Figure 10. *Accuracy scores for 4 test HDR scenes by scene*

Results and Analysis

Accuracy Evaluation using Real Scenes

Figure 9 shows the average overall accuracy scores for the four test scenes. These results indicate how well the algorithms reproduce the appearance of the physical scenes. The interval scale with 95% confidence level was generated using Thurstone's Law, Case V. The overall results show that iCAM06 is ranked first, but not significantly better that the other two Photoshop methods. This group of algorithms performed significantly better than other algorithms.

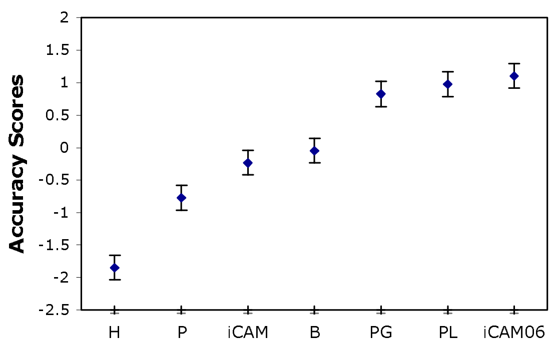

Figure 9. *Overall accuracy scores of tone-mapping operators for 4 real-world scenes*

The results for individual scenes (Figure 10) provide more insight. The test algorithms are separated into three groups: iCAM06 and two Photoshop methods have all positive scores over the test images, the photographic reproduction and histogram adjustment all have negative scores, and the bilateral filter and iCAM do not have the same homogeneity as other algorithms.

Accuracy Evaluation using HDR Display

Figure 11 plots the overall results obtained by using an HDR display as surrogate to the physical scene. The results show a similar pattern to those in Figure 9 where iCAM06 performs significantly better overall. One discrepancy is found where iCAM has a statistically significant higher score than the bilateral filter when evaluated against the HDR display whereas, while virtually statistically equivalent as their confidence intervals overlap, their average scores are reversed when evaluated against the physical scene.

In this regard, it was observed that the images rendered on the HDR display were slightly less colorful than the corresponding physical scenes and that overall contrast was slightly higher thereby sacrificing local area contrast in the shadows and highlights. The fact that iCAM generates tone-mapped images with lower colorfulness and highlight and shadow contrast than reality[7] might explain this discrepancy as, while not statistically significant, their average scores are nevertheless reversed. The results for individual image are shown in Figure 12.

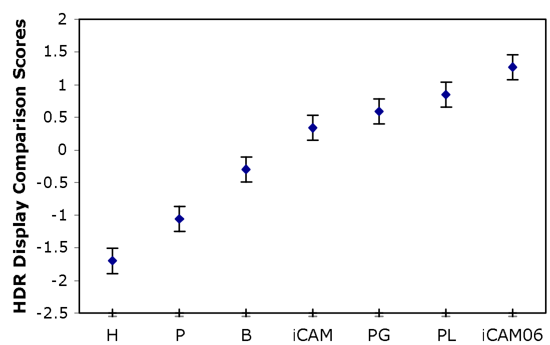

Figure 11. *Overall scores of tone-mapping operators for 4 test images on HDR display*

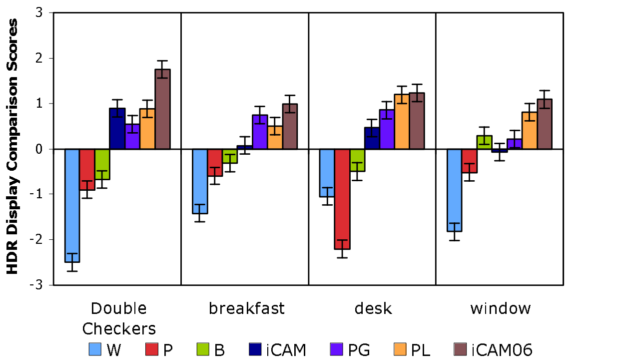

Figure 12. *Accuracy scores for 4 test images using HDR display by image*

Summary Results and Conclusions

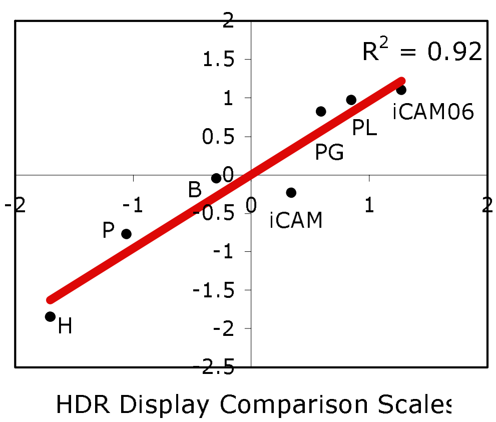

Figure 13. *Comparison of accuracy evaluation results from comparison with images on an HDR display and direct comparison with real-world scenes*

The overall tone mapping accuracy scales from the HDR display comparison experiment are plotted in Figure 13 against those from the comparison with the real-world scene to investigate the validity of using a HDR display as a surrogate for the real-world scene. A linear regression, as shown, illustrates that the scales from these two experimental methods correlate well with each other, with a coefficient of determination of 0.92.

These results validate the application of an HDR display for evaluating the accuracy of tone-mapping operators instead of building actual scenes in a lab environment. It provides many benefits of simplicity in experimental design and the opportunity for testing a large a variety of images such as outdoor scenes and scenes with people that could not be easily built in the lab.

From the last section of this paper, the results for the iCAM and bilateral filter algorithms switched their ranks in the accuracy evaluation between the two experimental methods. As noted, this inconsistency was attributed to colorfulness and contrast differences between the real scene and its reproduction by the HDR display. This inconsistency is, in fact, being addressed as a part of a larger MCSL effort in understanding and optimizing color and tone reproduction in HDR media.

References

[1] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, Evaluation of tone mapping operators using a High Dynamic Range display, *ACM Transactions on Graphics, Volume 24,* Issue *3,* pg. 640-648 (2005)

[2] F. Durand and J. Dorsey, Fast Bilateral Filtering for the Display of High-Dynamic-Range Image, *Proceedings of ACM SIGGRAPH*, Computer Graphics Proceedings, Annual Conference Proceedings, pg. 257-266 (2002)

[3] E. Reinhard, M. Stark, P. Shirley and J. Ferwerda, Photographic Tone Reproduction for Digital Images, Proceedings of ACM SIGGRAPH, Computer Graphics Proceedings, Annual Conference Proceedings, pg. 267-276 (*2002*)

[4] G.W. Larson, H. Rushmeier and C. Piatko, A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes, IEEE Transactions on Visualization and Computer Graphics, pg. 291-306 (1997).

[5] G.M. Johnson and M.D. Fairchild, Rendering HDR images, *IS&T/SID 11th Color Imaging Conference*, Scottsdale, pg. 36-41 (2003)

[6] J. Kuang, G.M. Johnson, M.D. Fairchild, iCAM06: A refined image appearance model for HDR image rendering, Journal of Visualization and Communication, doi:10.1016/j.jvcir.2007.06.003 (2007)

[7] J. Kuang, H. Yamaguchi, C. Liu, G.M. Johnson, M.D. Fairchild, Evaluating HDR Rendering Algorithms. ACM Transactions on Applied perception, Volume 4, Issue 2 (2007)

[8] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, High dynamic range display systems, ACM Transactions on Graphics, 23(3), pgs. 760-768 (2004)

[9] R.S. Berns, S.R. Fernandez, L. Taplin, Estimating Black-Level Emissions of Computer-Controlled Displays, Munsell Color Science Laboratory, Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology (2002)

[10] F.A. Imai, D. Kui, and R.J. Motta, Colorimetric characterization of a HDR Display System, *Proceedings of 10th Congress of the International Colour Association,* Granada (2000)

[11] M.A. Robertson, S. Borman, R.L. Stevenson, Dynamic range improvement through multiple exposures, IEEE International Conference on Image Processing, Volume 3, pg. 159-163 (1999)

Author Biographies

Jiangtao Kuang *received his B.S. degree in optical engineering from Zhejiang University, China (2000) and his PhD in imaging science from Rochester Institute of Technology (2006). Since then he has worked in OmniVision Tech. Inc. at Sunnyvale, CA. His work has focused on the research and development of high-dynamic-range CMOS image sensor, digital image processing algorithms and image quality.*

Rodney L. Heckaman *is a fourth-year Ph.D. student and Macbeth-Engel Fellow in Image Science, Rochester Institute of Technology. His work focuses on perceptual gamut, brilliance, and surround with application to high-dynamic-range displays. He graduated in 1968 from the Ohio State University in engineering physics with postgraduate work performed at the University of Rochester's Institute of Optics and Harvard University in Finance and retired after 32 years of service at the Eastman Kodak Research Laboratory.*