Rendering Non-Pictorial (Scientific) High Dynamic Range Images

Sung Ho Park and Ethan D. Montag
Munsell Color Science Laboratory, RIT
Rochester, New York, USA

Abstract

This research integrates the techniques used for the display of high dynamic range pictorial imagery for the practical visualization of non-pictorial (scientific) imagery such as remote sensing, medical imaging, astronomical imaging, etc. for data mining and interpretation. Nine algorithms were utilized to overcome the problem associated with rendering high dynamic range image data to low dynamic range display devices, and the results were evaluated using psychophysical experiments. Two paired-comparison experiment judging preference and scientific usefulness and a target detection experiment were performed. The paired-comparison results indicate that the Zone System algorithm performs the best on average and the Local Color Correction method performs the worst for both paired-comparison experiments. The results show that the performance of different encoding schemes depend on the type of data being visualized. The correlation between the preference and scientific usefulness judgments ($R^2 = 0.31$) demonstrates that observers tend to use different criteria when judging the scientific usefulness versus image preference. The result of the target detection experiment illustrates that the detectability of targets in an image is greatly influenced by the rendering algorithm due to the inherent differences in tone mapping among the algorithms.

Introduction

One possible aim of realistic image rendering or reproduction is the creation of images that share identical appearance attributes as a real scene. The real world exhibits a wide range of luminance values. The human visual system is capable of perceiving this wide range of dynamic scenes spanning five orders of magnitude and adapting more gradually to over nine orders of magnitude, which is facilitated by local adaptation that allows regions of various luminance levels to be viewed essentially simultaneously. Recent advances in high dynamic range capturing systems[1-3] make it possible to capture a highly detailed range representation of the scene and later process the data in order to select the image that better fulfills the given requirements. However, since a typical desktop display, such as CRTs and LCDs, is only capable of displaying two orders of magnitude of dynamic range, the question is then how can we reproduce and visualize such HDR images in a standard output device.

More recently, concern has grown in the visualization and scientific communities over the use of scientific imagery, its interpretation, and the relation of the data to its interpretation. Novel techniques are also required for imagery captured from non-visual sources such as remote sensing, medical imaging, astronomical imaging, etc. The goal of this study is to integrate the techniques used for the display of HDR pictorial imagery for the display of non-pictorial imagery while searching for perceptually based schemes for encoding this imagery that facilitate its interpretation. By applying these same HDR processing techniques developed for pictorial imagery, it is hypothesized that more information can be conveyed because local perceptual contrast in a wider range of the scene will be preserved by automatically adjusting the luminance and chromatic contrast in the image based on the image content.

Much research has been done to develop algorithms that are capable of recreating a truthful rendition of high dynamic range image onto lower dynamic range displays.[4-7] Unlike pictorial imagery, the truthfulness of the displayed non-pictorial imagery cannot be evaluated by comparison with the original scene. Instead, the usefulness of the display lies in the ability of the user to visually interpret and use the data. The term, non-pictorial, refers to scientific imagery captured outside the visible wavelength region or of objects not accessible to the human eye, such as hyperspectral data captured by spacecraft or aircraft, astronomical images captured using non-visible wavelengths, or characteristics of human tissue obtained in medical imaging. Since the main focus of this project is to test algorithms for the display of non-pictorial HDR imagery that is univariate, the visualization of multidimensional data is not of concern in this study.

There are three aspects of this study: 1) The development and implementation of HDR algorithms including some used for HDR pictorial imagery 2) The psychophysical evaluation of these algorithms in rendering this non-pictorial imagery, and 3) The psychophysical measurement of the effect of tone and contrast mapping on target detection. The results from the evaluation aspect will be used as feedback to help improve the algorithms used to encode the data. Two psychophysical experiments were conducted to evaluate these algorithms. The goal of the psychophysical testing was to determine which algorithms lead to visual preference and better data mining and interpretation.

Non-Pictorial (Scientific) Imagery

Five different sources of scientific imagery were utilized in this study, and one pictorial image was also included for comparison. They are briefly described in Table 1. Histograms of the images and thumbnails of each image processed by the iCAM are shown in Figure 1. A processed radar image was cropped to 930(rows) x 800(columns) in order to display the image in true size.

Table 1. Information of the Imagery Exploited.

Image Type	Source	Max digit	Size
Astronomical	Hubble Space Telescope	65455	1000x650
Medical	Magnetic Resonance	1655	256x256
Hyperspectral	AVIRIS	28175	614x512
Radar	AIRSAR	16384	1485x2161
Infrared	WASP	2302	640x510
Pictorial	Memorial Church	65536	768x512

Algorithms

Since the issues of realistic tone mapping were introduced, many algorithms have been proposed to overcome the problem of displaying HDR image. In order to simulate the realistic perception of world luminance levels on a standard output device, some algorithms utilize perceptual data based on psychophysical experiments, and others exploit a mathematical approach to simply compress the luminance range with aim of obtaining the maximum visibility on the display device and without considering the perceptual aspects of visual system. In any cases where tone reproduction attempts to simulate reality, one of the most important factors for rescaling the high dynamic range to fit into the smaller output dynamic range is that the final image maintains the lightness integrity of the original scene.

Nine algorithms (Linear Mapping,[9] Sigmoid-lightness rescaling,[9] Localized Sigmoid Mapping[8] Spiral Rendering,[10] Photoshop (Auto-levels),[11] iCAM,[7,12,13] Local Color Correction,[14] Fast Bilateral Filtering,[5] and Zone System[4]) primarily proposed for the display of HDR pictorial imagery were implemented for the display of non-pictorial imagery. The nine proposed algorithms varied from a simple linear scaling factor to more complete high end solutions, which take into account complex perceptual attributes. In other words, they vary from simple global (spatially uniform) mapping to complex multi-scale local (spatially varying) mapping to imitate the visual system. Inverse display characterization was applied at the end of each algorithm to account for inherent device nonlinearity before displaying. Controllable parameters for each algorithm were set as stated and recommended in its reference. See Reference 8 for more detailed description of each algorithm.

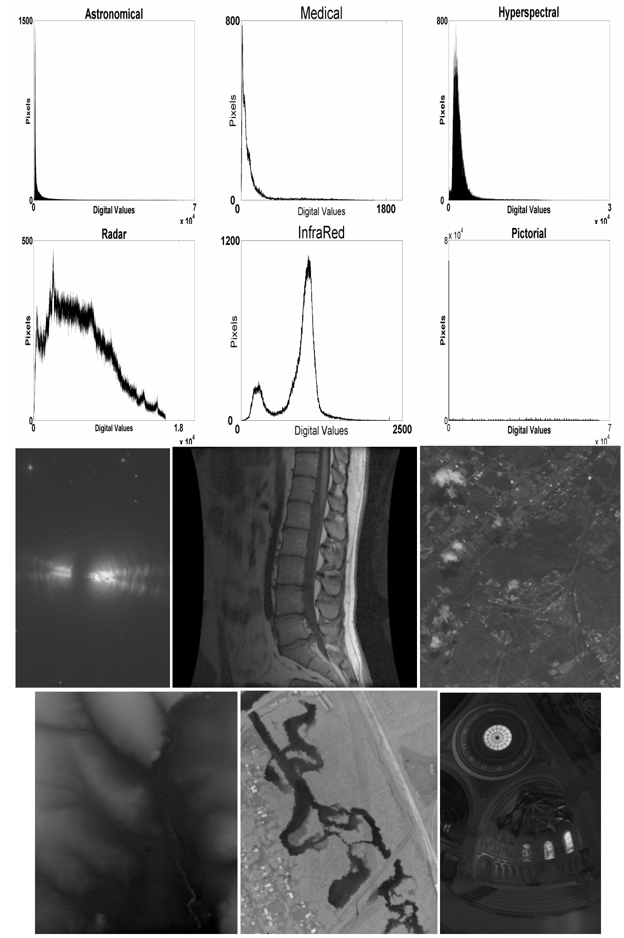

Figure 1. Histograms and thumbnails of the imagery exploited in the study. Top row: Astronomical, Medial, and Hyperspectral image, Bottom row: Radar, Infrared, and Pictorial image.

Psychophysical Experiments

The psychophysical experiments were conducted on a colorimetrically characterized 23" Apple Cinema HD flat-panel LCD display connected with an Apple Power Mac G4 dual 1GHz processor.[17] Since this project deals with non-pictorial imagery, the fidelity of the processed images cannot be judged by comparison with the original scene. Instead, three psychophysical experiments were carried out to measure the effect of the different algorithms on the perception of the various images. More detailed description of each experiment can be found in Ref. 8.

Paired-Comparison Experiments

Two experiments were conducted to judge both the observers' preference and the scientific usefulness of the images in a paired-comparison paradigm. The goal of the first experiment was to determine which encoding schemes rendered the high dynamic range images in more preferable way. In this task, 25 observers were instructed to choose the image that they preferred in each pair. In the second experiment, the same stimuli were used but the observers were instructed to choose the image in each pair that they considered to be "more scientifically useful." Observers were allowed to use their own criteria for making these judgments.

Target Detection Experiment

A third psychophysical experiment was performed to measure how the change in contrast tone mapping due to the various algorithms affected the detection of a target as measured by the amplitude of the target in the raw image data. This experiment used a two-alternative forced-choice method of constant stimuli to find the threshold for detecting embedded noise target in the Medical image. The task of target detection can be considered as a way of determining the change in detectability of a "tumor" embedded in the image.

Thresholds, in terms of the original digital values in the image data, were measured for three different targets in the image. Each of the targets had a different spatial size and location in the image. The targets consisted of random noise in a Gaussian envelope. For each target, a series of images were precomputed with different amplitudes of noise added to the original image data. The targets were placed at the three different lightness areas, dark-, mid-, and high-tone areas separately, and the images were processed with each algorithm. The experiment was analyzed using Probit analysis to determine the corrected-for-chance 50% threshold for target detection.

Results and Discussion

Paired-Comparison Experiments

The paired-comparison data was converted to interval scales for analysis by employing the Thurstone's Law of Comparative Judgments (Case V).[18] For the preference task, observers were asked to choose which of the two images they preferred in terms of overall image quality. For scientific usefulness, no specific criteria were given to observers. They had to decide what is meant by "scientifically useful," which may have introduced some difficulty in deciding what criteria to use. The image preference and judged scientific usefulness of all images are shown in Figure 2. The error bars on all plots were calculated in terms of interval scale units for a 95% confidence interval.[19] Both figures indicate that performance of each algorithm depends on the image type. Comparison of these graphs also shows the different pattern of response between the two tasks. This distinction is more apparent in average performance data shown in Figure 3. The low correlation between the two sets of results demonstrates that the observers were using different criteria for the two tasks. As shown in Figure 4, the data has an R-square value of 0.305.

The images processed using the Zone System were judged high both in preference and scientific usefulness. The Local Sigmoid function showed the most prominent changes between the two tasks. Observers did not prefer the images processed by the local sigmoid algorithm but they found that it revealed data that were judged to be more scientifically useful.

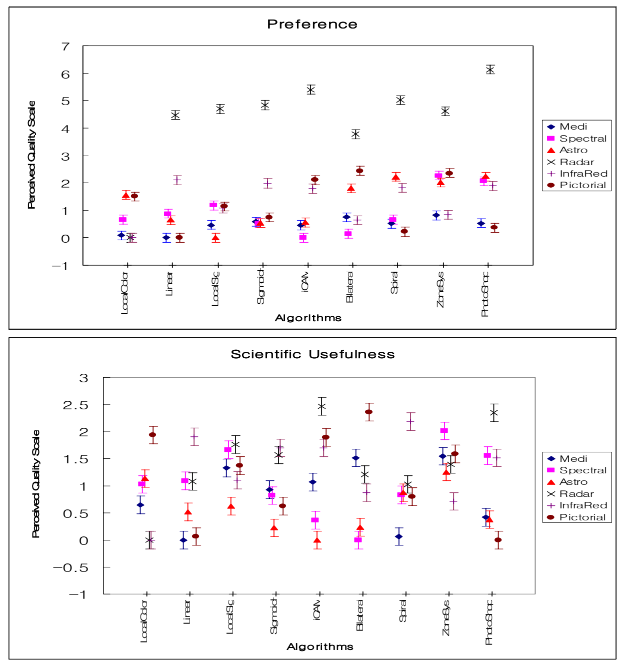

Figure 2. Results of Paired-Comparison experiments.

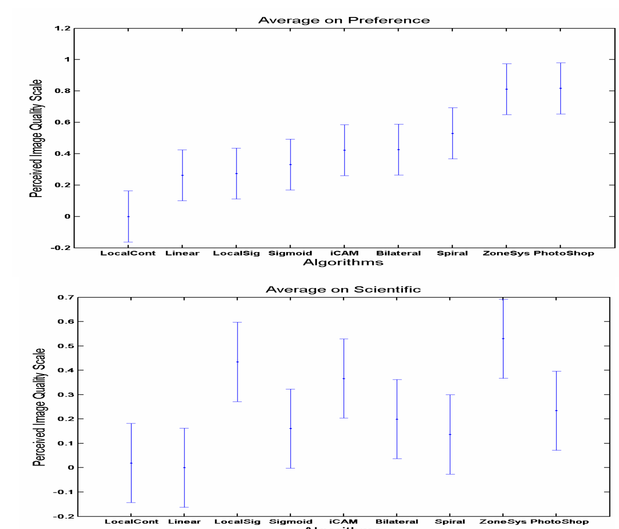

Figure 3. Average performance of Paired-Comparison experiments: Preference (left), scientifically useful (right).

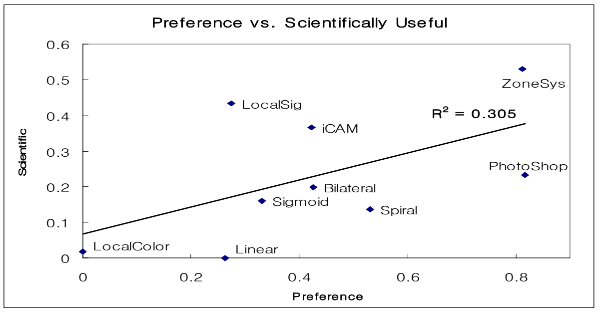

Figure 4. Plot of Preference vs. Scientific Usefulness.

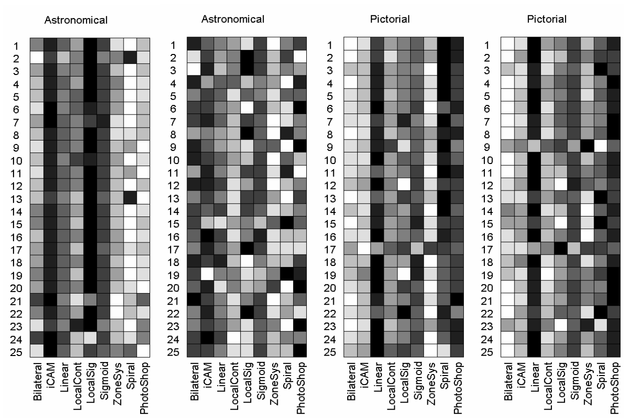

Figure 5. Schematic diagram for Astronomical and Pictorial image: Preference (left), Scientific Usefulness (right).

Individual variability for the Astronomical image is plotted using diagrams that show the observer's response patterns in Figure 5. Individual observer data is shown along the rows and the columns represent the algorithm types. A box with a lighter shade indicates that the algorithm in that column was chosen more frequently in the experiment than the other algorithms. Therefore, white boxes show often chosen algorithm types and black boxes show rarely chosen types. The apparent stripe pattern is the indication of consistent responses among the observers. The Astronomical image, Figure 5, left, illustrates that observers agreed on their preference judgments but not on their judgments of scientific usefulness. By contrast, Figure 5, right, shows similar individual agreement on both preference and scientific usefulness task for the Pictorial image.

The Zone System performed well for the majority of the tested image. Nevertheless, it did not achieve the same result for the Infrared image. As is illustrated in Figure 6, performance of algorithms can be divided into two groups. Except for iCAM, algorithms with local contrast feature behave worse than the one without so that simple linear mapping renders the image better. Figure 7 shows the plot of the Infrared image's pixel by pixel values for the linear rendering versus the Zone System (left) and versus iCAM (right). The Linear vs. Zone System and Linear vs. iCAM plots show how the algorithm with spatial filtering rendered the image compared to one without. For this particular image, the relationship between Linear and Zone System can be explained by a simple gamma curve. The shadow and highlight areas are more compressed in the image processed by Linear method than Zone System. However, comparing to iCAM, only the shadow regions are more compressed and other regions are linearly related. These results are different depending on the spatial structure of the image.

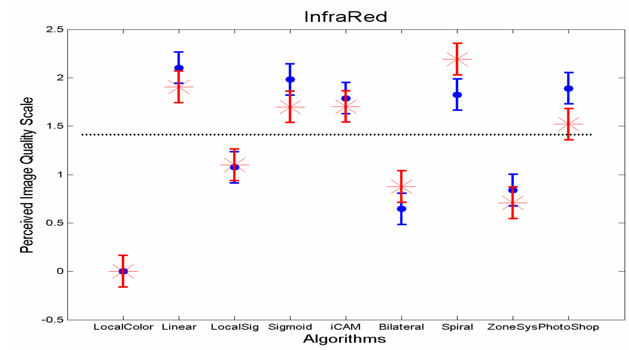

Figure 6. Plot of paired-comparison result for Infrared image: Dot (•) represents Preference results and Asterisk (*) for Scientific usefulness.

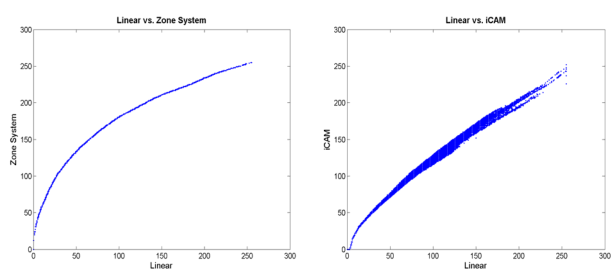

Figure 7. Comparing processed Infrared image: Linear vs. Zone System and Linear vs. iCAM.

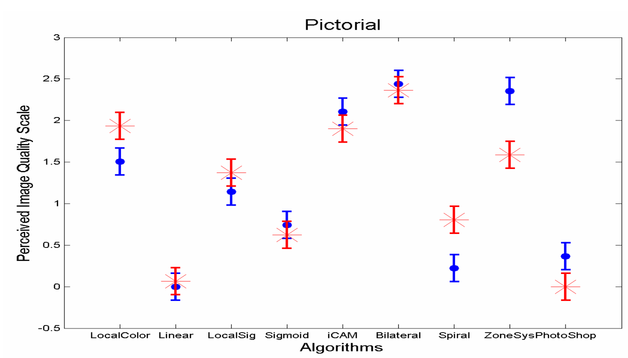

Figure 8. Plot of paired-comparison results for Pictorial image: Dot (•) represents Preference results and Asterisk (*) for Scientific usefulness.

Photoshop shows the worst performance for the pictorial image (Figure 8), though it performs well on average (Figure 3). This result might be explained by the image histogram of the image. The histogram of the image, see Figure 1, shows that the majority of pixels are located at extremely low ends and only a small number of pixels are dispersed over the complete range. Photoshop tends to produce better results with images that have a wider distribution of pixel values, such as the Radar image. A simple method for rendering HDR pictorial imagery is to apply a 99 percentile clipping and a gamma correction. These techniques are simple but powerful enough to obtain acceptable reproduction.

The Spiral encoding is the only algorithm that adds color to the image. This algorithm can be treated as a linear L* mapping since the digital values were first mapped linearly to L* values, and then, chroma and hue values were add to the monochrome image. Observers tend to favor color over monochrome image when tone mapping is acceptable. However, this tendency diminishes when judging the scientific usefulness (see Figure 3). Due to the limitation of tone mapping, spiral encoding can't reveal much hidden information. If other tone mapping techniques can be combined with color, the performance might show a possible increase.

The Local Color Correction algorithm is the worst method to use for rendering the HDR images on average and especially for the Radar and InfraRed image. However, this method performed well for the Pictorial and Astronomical image. It is better than algorithms with global mapping but not good enough to compare with algorithms with local contrast mapping function. The performance of iCAM is neither excellent nor bad. The results are somewhat expected since iCAM is intended to render a pictorial scene truthfully rather than enhancing it. The aim of iCAM is accurate prediction of a variety color appearance phenomena that mimic the human perception. Experiment on the accuracy, which is not possible for scientific imagery, can be conducted to support this hypothesis by employing pictorial imagery.

Target Detection Experiment

The target detection experiment was conducted to measure the detectability of an embedded noise target in the Medical image to demonstrate the effect of the algorithms on target detection. It is obvious that the spatial structure and tone scale mapping of the images and their resultant renderings will introduce distortions that will effect target detection. Therefore, the characteristics of targets in the image should be taken into account when determining the appropriate rendering algorithm. In theory, better algorithms will allow detection of targets with low amplitude regardless of the surrounding local contrast.

The noise-targets were first obtained by creating normally distributed random noise, and then, multiplying with a Gaussian envelope to reduce the sharp edges. The size of the Gaussian filter was set to 5, 8, and 10 pixels on 15x15, 20x20, and 30x30 noise patch for high-, mid-, and dark-tone regions respectively. The amplitude of noise was varied depending on the image type and was optimally set in seven steps. Each target image was presented randomly with the corresponding rendered algorithm without the target 60 times for one subject. The threshold results are shown in Table 2.

The threshold was set at the corrected 50% probability of detection. Lower thresholds indicate better detection of the noise at small amplitudes. As it is illustrated by the Table 2 and Figure 9, the results are different depending on the target size and location. There is no clear correspondence between these threshold values and the results from paired-comparison experiments. For the high-tone area, the Local Sigmoid method has the best detectability with low threshold value and the Zone System was the worst, which is somewhat opposite from the paired-comparison results. However, for the mid- and dark-tone area, the Zone System and the Local Sigmoid method shows the best detectability with the lowest threshold values and the Linear and the Spiral method shows the highest threshold values representing the worst detectability. These results closely coincide with the scientific usefulness paired-comparison results. In general, the effects seen with the targets embedded in mid- and dark-tone area have smaller thresholds, are more similar to each across algorithm than the high-tone region target, and closely correspond with the results of scientific usefulness paired-comparison experiment.

Table 2. Results of Target Detection Experiment.

Algorithm	High	Mid	Dark
Linear	42.83	21.81	28.86
iCAM	70.77	14.44	4.83
Sigmoid	41.89	16.99	13.13
Spiral	34.79	17.87	43.52
Local Sigmoid	16.55	8.67	4.68
Local Correction	90.64	15.11	4.16
Bilateral	71.77	12.02	3.76
Zone System	87.75	11.15	3.74
PhotoShop	32.86	15.41	17.18

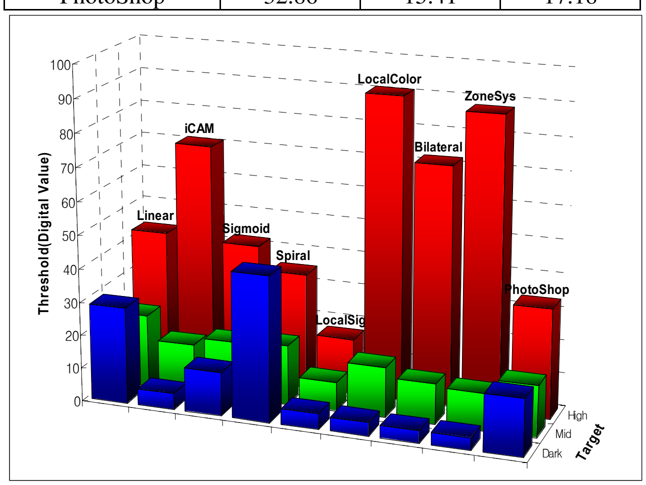

Figure 9. Bar graph of target detection experiment result

Conclusion

Nine algorithms were used for rendering pictorial HDR image from a variety of scientific imagery. Two paired-comparison psychophysical experiments were performed to evaluate which algorithms produced the most preferred images and images that were considered scientifically useful.

Although the Zone System has the best performance on both average preference and scientific usefulness, the results of the paired-comparison experiments suggest that different encoding schemes might be useful depending on the data type. There was little correlation between preference and scientific usefulness indicating that observers used different criteria for the two tasks. The observers demonstrated substantial individual variation in their judgments.

The effects of image distortion introduced by the rendering algorithms in the third experiment were investigated using a noise target threshold detection paradigm. The results of high-tone area target indicate that the detectability does not strictly correspond with the results of the paired-comparison experiment. However, the threshold results of mid- and dark-tone area target show somewhat close relationship with the scientific usefulness paired-comparison experiment.

References

1. P. E. Debevec and J. Malik, Recovering High Dynamic Range Radiance Maps from Photographs, *Proceedings of SIGGRAPH 97*, 369-378 (1997).
2. F. Xiao, J.M. DiCarlo, P.B. Catrysse and B.A. Wandell, High Dynamic Range Imaging of Natural Scenes, *IS&T/SID 10th Color Imaging Conference*, 337-342 (2002).
3. S.K. Nayar and T. Mitsunaga, High Dynamic Range Imaging: Spatially Varying Pixel Exposures, *Proceedings of IEEE CVPR*, (2000).
4. E. Reinhard, M. Stark, P. Shirley, and J. Ferwera, Photographic Tone Reproduction for Digital Images, *Proceedings of SIGGRAPH 02*, 267-276 (2002).
5. F. Durand and J. Dorsey, Fast Bilateral filtering for the display of High-Dynamic Range Images, *Proceedings of SIGGRAPH 02*, 257-266 (2002).
6. J. M. Dicarlo and B.A. Wandell, Rendering High dynamic Range Images, *Proceedings of SPIE Image Sensors* 392-401 (1999).
7. G. M. Johnson and M.D. Fairchild, Rendering HDR Image, *IS&T/SID 11th Color Imaging Conference* 36-41 (2003).
8. S. H. Park, Rendering Non-Pictorial High Dynamic Range Images, MS Thesis, RIT, Rochester, NY, 2004.
9. G.J. Braun and M.D. Fairchild, Image Lightness Rescaling Using Sigmoidal Contrast Enhancement Functions, *Journal of Electronic Imaging*, Vol. 8, 380-393 (1999).
10. E.D. Montag, The Use of Color in Multidimensional Graphical Information Display, *IS&T/SID 7th Color Imaging Conference*, 222-226 (1999).
11. Adobe Creative Team, *Adobe Photoshop 7.0 Classroom in a Book*, Adobe Press, USA, (2002)
12. M.D. Fairchild and G.M. Johnson, Meet iCAM: An image color appearance model, *IS&T/SID 10th Color Imaging Conference*, 33-38 (2002).
13. M.D. Fairchild and G.M. Johnson, The iCAM framework for image appearance, image differences, and image quality, *Journal of Electronic Imaging,* in press (2004).
14. N. Moroney, Local Color Correction Using Non-Linear Masking, *IS&T/SID 8th Color Imaging Conference*, 108-111 (2000).
15. C. Tomasi and R. Manduchi, Bilateral Filtering for Gray and Color Images, *Proceedings of IEEE Int. Conference on Computer Vision*, 836-846 (1998).
16. F. J. J. Blommaert and J. B. Martens, An Object-Oriented Model for Brightness Perception, *Spatial Vision* 5, 1, 15-41 (1990).
17. E. A. Day, L. A. Taplin, and R. S. Berns, Colorimetric Characterization of a Computer-Controlled Liquid Crystal Display, *Color Research & Application*, in press (2004).
18. P. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development*, Imcotek Press, Winchester, MA (2000).
19. E. D. Montag, Louis Leon Thurstone in Monte Carlo: Creating Error Bars for the Method of Paired Comparison, *IS&T/SPIE Symposium on Electronic Imaging: Science and Technology*, SPIE (2004).
20. S. Nishisato, Elements of Dual Scaling: An Introduction to Practical Data Analysis, Lawrence Erlbaum Assoc., Hillsdale, N.J (1994).

Biographies

Sung Ho Park is currently a second year M.S. student in Color Science program at MCSL, Rochester Institute of Technology. His current research interests lie in visualization and rendering. He will receive his M.S. in August 2004. For his Master's thesis, he has been working with Dr. Ethan Montag on rendering non-pictorial HDR images. He received his B.S. in Printing System Management also from RIT.

Ethan D. Montag received his Ph. D. in Experimental Psychology in 1991 from UCSD working in color vision. He is an Assistant Professor at RIT's Center for Imaging Science where he pursues work in color science in the Munsell Color Science Laboratory. His current interests include image quality, color gamut mapping, color vision, color tolerance measurement and the use of color information display. Dr. Montag is a member of OSA, ICVS, ISCC, and IS&T.