Comparative Study of Metrics for Spectral Match Quality

Francisco H. Imai, Mitchell R. Rosen and Roy S. Berns
Munsell Color Science Laboratory, Rochester Institute of Technology
Rochester, New York
e-mail: imai@cis.rit.edu, rosen@cis.rit.edu and berns@cis.rit.edu

Abstract

The selection of metrics for spectral matches is fundamental to MVSI (multi-channel visible-spectrum imaging) otherwise known as spectral imaging. The metrics used for spectral matches can impact everything from the selection of the filters used for multi-channel capture to the evaluation of the spectral estimation. However, there is, as yet, no consensus on which metric should be applied for spectral matches. The purpose of this research is to compare various metrics that have been used for spectral matches. The metrics for spectral comparison were categorized in four classes: CIE color difference equations, spectral curves difference metrics, metamerism indices and weighted spectral metrics. Here we show an analysis of the appropriateness and weakness of each metric. We compare their use for various types of spectral mismatches resulting from problems in imaging calibration, out-of-gamut colors and those due to metamerism.

Introduction

It is widely accepted that multi-channel visible spectrum imaging (MVSI)[1] is the only way to ensure a color match for all observers and across changes in illumination, as result of estimating the spectral information of the scene from the captured channels. This type of imaging is particularly critical for high-end color reproduction such as artwork reproduction[2,3] and telemedicine.[4] There are various techniques for capturing images used for spectral estimation. For instance, one may use wide or narrow spectral bandpasses for image acquisition channels.[5] MVSI research is impacted by a wide range of disciplines including filter design,[6] target analysis,[7] image processing,[8] image compression[9] and multi-ink printing.[1] The design of system components relies on the use of cost functions for evaluating the expected quality of estimated spectra. Also, for comparing and evaluating spectral reproduction systems, metrics are needed to evaluate the closeness of spectral matches. There is, as yet, no consensus on which metric is best for evaluating spectral matches. It was found by Imai *et al.* that depending on the shape and magnitude of spectral curves, one metric could perform better than another.[5] Therefore this research has the objective of comparing existing spectral-match metrics with emphasis on highlighting their limitations and usefulness.

Metrics for Spectral Comparison

From the literature, metrics for spectral matching quality tend to fall within the following categories:

I. CIE Color Difference Equations

CIE committees have developed color-difference equations including CIELUV,[10] CIELAB,[10] CIE94[11] and the most recently, CIE2000.[12] Although the aim of the color difference is not to evaluate spectral curves matches, metrics such as CIELUV, CIELAB and CIE94 have been used as cost functions or to evaluate spectral estimation accuracy.[13-15] Since color difference equations consider the response of human visual system under controlled illumination and observation conditions they can give a good clue about the color matching. However, color difference equations are prone to produce bad correlation to spectral matches, particularly for metameric pairs.

II. Spectral Curves Difference Metrics

Another approach to comparing spectral curves is based on computation of spectral curve differences.

A. Root mean square error – RMS error is a very simple metric that has been used for spectral estimation evaluation for many studies.[14,15]
B. Hernández-Andrés *et al.* have suggested a goodness-of-fit coefficient (GFC) to test reconstructed daylight spectra.[16] The GFC is based on the inequality of Schwartz and it is described by the equation (1)

$$GFC = \frac{\left|\sum_j R_m(\lambda_j) R_e(\lambda_j)\right|}{\sqrt{\left|\sum_j \left[R_m(\lambda_j)\right]^2\right|}\sqrt{\left|\sum_j \left[R_e(\lambda_j)\right]^2\right|}} \quad (1)$$

where $R_m(\lambda_j)$ is the measured original spectral data at the wavelength λ_j and $R_e(\lambda_j)$ is the estimated spectral data at wavelength λ_j. GFC $\geq$ 0.999 and GFC $\geq$ 0.9999 are required for respectively good and excellent spectral matches.

III. Metamerim Indices

A metamerism index compares the extent to which two spectra are different between a reference condition and a test condition under different illuminants and observers.

A. Fairman proposed a metamerism index using parameric decomposition.[17] In this method, the test spectrum is corrected spectrally until an exact tristimulus equality is achieved under a reference condition. The metameric index is a CIE color-difference equation for a test illuminant and observer. The CIE refers to this type of index as a "special index of metamerism."[18]
B. Viggiano's perception-reference method compares radiance ratio spectra. This index (M_V) is computed[19, 20] as shown in Equation (2)

$$M_V = \sum_{\lambda=1}^{n} w(\lambda)\|\Delta\beta(\lambda)\| \quad (2)$$

where $\Delta\beta(\lambda)$ is the difference between the two spectra and $w(\lambda)$ are the weights computed as follows:

$$w(\lambda) = \sqrt{\left(\frac{dL^*}{d\beta(\lambda)}\right)^2 + \left(\frac{da^*}{d\beta(\lambda)}\right)^2 + \left(\frac{db^*}{d\beta(\lambda)}\right)^2} \quad (3)$$

It is a refinement of a spectral-based metameric index based on a weighted sum of the absolute differences between two spectra proposed by Nimeroff and Yukov.[21] A spectral comparison index of 3 units is considered an excellent match.[22] The CIE refers to this type of index as a "general index of metamerism."[18]

IV. Other Weighted rms Metrics

It is possible to weight spectral reflectance factor rms error between reference and test curves in a way that consider some properties of human visual system. The general weighted rms error equation is shown as follows:

$$wrms = \sqrt{\frac{\sum_{\lambda=1}^{n}\left(\sqrt{w(\lambda)}\Delta\beta(\lambda)\right)^2}{n}} \quad (4)$$

where $w(\lambda)$ is the weight, n is the number of the wavelengths, $\Delta\beta(\lambda)$ is the difference between the two spectra.

A. Inverse of the reference spectra – In this case we are considering that it is more important to weight spectral data with small magnitude than the ones with larger magnitude because human visual system is more sensitive to mismatches in dark colors than light colors. The inverse relationship is shown in Equation (5).

$$w_{invR}(\lambda) = \frac{1}{R_m(\lambda)} \quad (5)$$

where $R_m(\lambda)$ is the measured reference spectral data.

B. Diagonal of matrix [***R***] – Cohen developed a mathematical technique, known as Matrix [***R***] based on Wyszecki's hypothesis that any stimulus can be decomposed into a fundamental stimulus with identical tristimulus values and a residual metameric black.[23] The matrix [***R***] can be easily calculated from the matrix ***A*** of weights for the reference illuminant and observer as shown in equation (6).

$$[\boldsymbol{R}] = A * inv(A^t * A) * A^t \quad (6)$$

where t denotes transposed of the matrix. The diagonal of the matrix [***R***] can be used as the weighting function for the rms calculation as shown in equation (7).

$$w_{diagR}(\lambda) = diag([R]) \quad (7)$$

It follows that there is one set of weights for each combination of illumination and observer. For example, if we consider D65 illuminant and 10 degree observer we have the weight curves shown in Figure 1. Figure 1 shows that the diagonal of matrix [***R***] biases the rms error calculation in a fashion that gives more importance to the wavelengths that correspond to higher sensitivity in the human visual response for a specific combination of illuminant and observer.

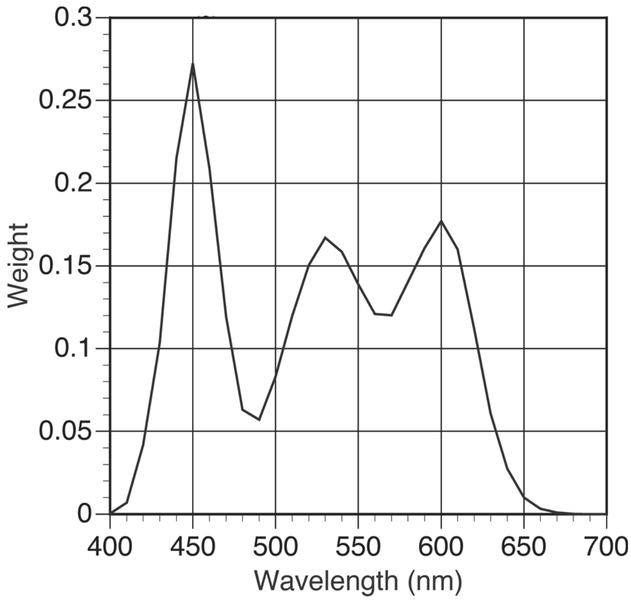

Figure 1. Weighting functions calculated from matrix [R] for D65 illuminant and 10 degree observer.

Experimental

Typical spectral curves of paints taken from previous investigations were used as test cases for evaluation of spectral metrics. Each metric candidate was observed in its response to deliberate departures from perfect matches. In particular, cases of mismatches due to shifts in amplitude for dark and light colors as well as metameric pairs showed interesting results.

The first experiment considered shifts in spectral amplitude. Figures 2a and 2b show a spectral mismatch for a blue paint by a magnitude shift of respectively 0.01 and 0.02 in reflectance factor units. This kind of mismatch can happen due to problems in calibration of the imaging system used to get signals used for spectral estimation. Figure 2c shows a case of a light blue color mismatch that was produced by shifting the case in Figure 2a by 0.5 reflectance factor units in order to compare the metrics for dark and light color mismatches.

Table I summarizes the evaluation of the spectral match metrics for a theoretical perfect spectral match and for the mismatched pairs represented in the Figures 2a, 2b and 2c.

Table I. Comparison of the spectral fit metrics for various reflectance pairs shown in Figures 2a, 2b and 2c.

Metric for spectral match	Perfect Match	Pair Figure 2a	Pair Figure 2b	Pair Figure 2c
CIELAB (D65, 10 degree observer)	0	4.4	8.1	0.6
CIE2000 (D65, 10 degree observer)	0	2.2	4.3	0.4
Spectral error rms factor	0	0.01	0.02	0.01
GFC	1	0.998	0.995	0.999
Fairman Metamerism Index (D65, A, 10 degree observer, DE2000)	0	0.7	1.4	0.2
M_v (D65, 10 degree observer)	0	11.4	22.7	2.7
wRMS (Inverse Reflectance)	0	0.047	0.094	0.013
wRMS (Diagonal Matrix [***R***])	0	0.003	0.006	0.003

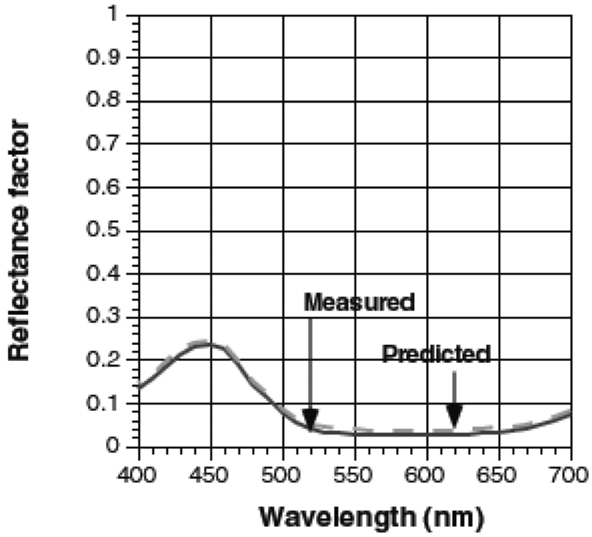

Figure 2a. Dark blue mismatch with 0.01 shift in amplitude

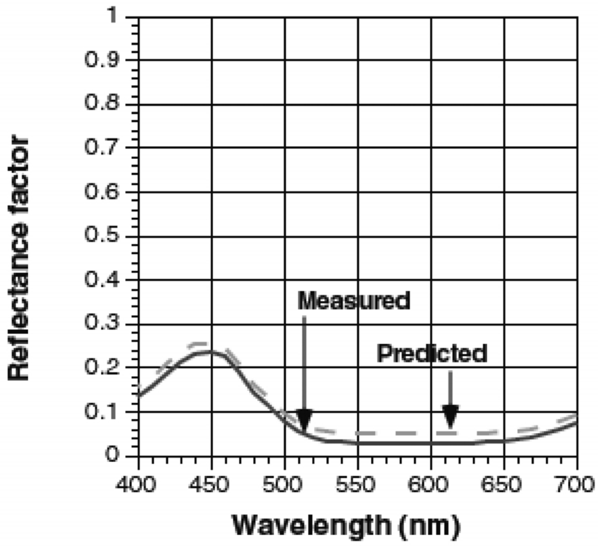

Figure 2b. Dark blue mismatch with 0.02 shift in amplitude

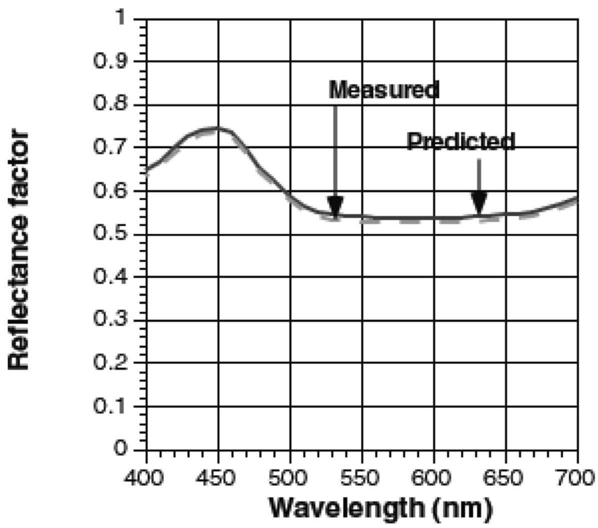

Figure 2c. Light blue mismatch with 0.01 shift in amplitude.

In the second experiment the metrics were evaluated against three pairs of metameric matches. Figures 3a, 3b and 3c show three spectral mismatches which cross each other three to four times. Table II summarizes the evaluation of the spectral match metrics for the pairs represented in the Figures 3a, 3b and 3c.

Discussions

From Table I it is shown that when comparing mismatches with a magnitude shift of 0.01 and 0.02 reflectance factor, all the metrics tested are scalable except the goodness-of-fit coefficient (GFC). However, from the comparison of dark and light blue colors with identical shape and shifted by 0.01, there was a difference among the metrics. Spectral curve difference metrics (GFC and rms error) and the weighted rms using the diagonal of the matrix [***R***] were unaffected by the change from dark to light, whereas the color difference equations, the weighted rms using the inverse of the reference spectra and metamerism indices presented smaller errors for a light color than for the dark because these metrics consider this aspect of human visual system in their calculations.

Table II considers curves that are highly spectrally distinct from one another. The color difference equations show that these spectra are close metamers under D65 illuminant and for the 10 degree observer. Spectral curve difference metrics (GFC and rms error) as well as the metameric indices and weighted rms metrics do detect significant mismatches for these pairs.

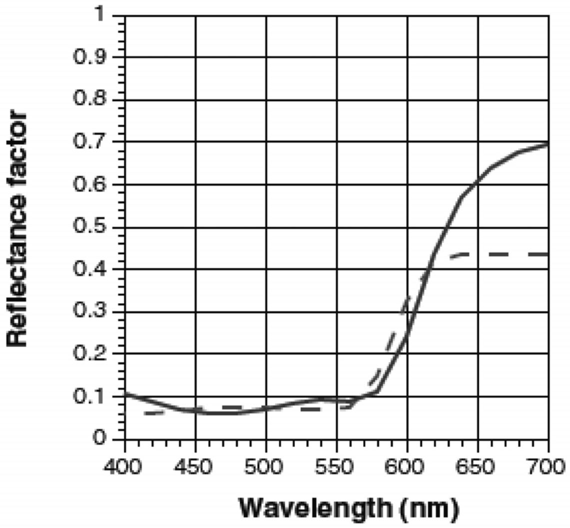

Figure 3a. First metameric pair

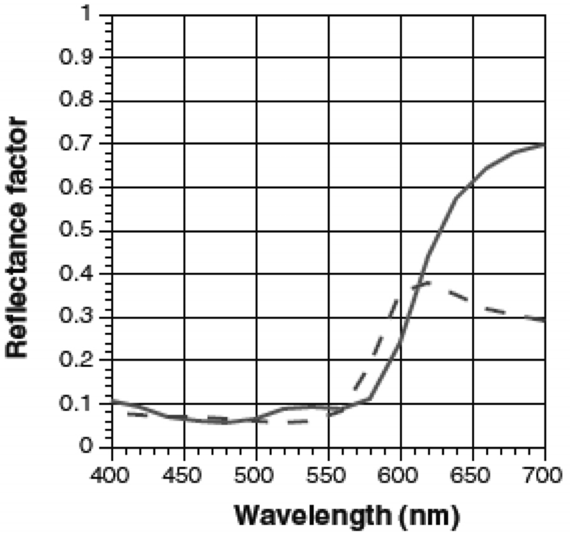

Figure3b. Second metameric pair

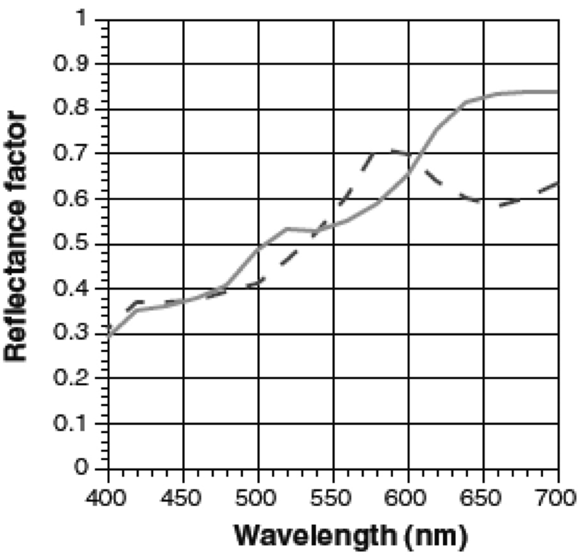

Figure 3c. Third metameric pair

Table II. Comparison of the spectral fit metrics for various metameric pairs.

Metric for spectral match	Spectral pair 1	Spectral pair 2	Spectral pair 3
CIELAB (D65, 10 degree observer)	0.04	0.07	0.02
CIE2000 (D65, 10 degree observer)	0.02	0.04	0.02
Spectral error rms factor	0.108	0.170	0.124
GFC	0.977	0.925	0.983
Fairman Metamerism Index (D65, A, 10 degree observer, DE2000)	0.8	2.2	3.9
M_v (D65, 10 degree observer)	18.6	25.3	14.8
wRMS (Inverse Reflectance)	0.176	0.227	0.1415
wRMS (Diagonal Matrix [R])	0.012	0.019	0.021

Table III. Comparison of the various metrics for spectral quality.

Metric for spectral match	Advantage	Disadvantage	Application
CIELAB	Consider human vision	Doesn't detect metamerism	When metamerism is not a issue
CIE2000	Consider human vision	Doesn't detect metamerism	When metamerism is not a issue
Spectral error rms factor	Easy to calculate and it is general	Doesn't consider human vision	Comparing physical stimuli
GFC	Easy to calculate and it is general	Doesn't consider human vision	Comparing physical stimuli
Fairman Metamerism Index	Gives difference in familiar units	Need specific sets of illuminant and observer	Gives a measure of metamerism for specific conditions
M_v	Consider human vision and difference between dark and light colors	Result units are not very intuitive	Could be a candidate for a specific spectral match maetric
wRMS (Inverse Reflectance)	Different weight for dark and light colors	Doesn't consider human cones sensitivities	Could be a candidate for a general spectral match metric
wRMS (Diagonal Matrix [R])	Spectral rms that considers human cones sensitivities	Doesn't consider differences between dark and bright colors	Could be a candidate for a specific spectral match maetric

Table III summarizes the results. The color difference metrics such as CIELAB and CIE2000 are based on human vision but they are prone to give good matches for metameric pairs and therefore should only be used when metamerism is not an issue such as when displaying a color on monitor. The spectral difference metrics such as rms error and GMC are easy to calculate but don't consider aspects of human vision and therefore they are more useful for comparing mismatches to physical measurements without the evaluation by human subjects. The Fairman metamerism index is a very useful metric to compare two spectra under two different illuminants but it is not a general metric in the sense that illuminants need to be specified. The weighted rms using the inverse of reflectance puts more weight on darker colors than light colors however it does not consider the cone sensitivities. On the other hand the weighted rms using the diagonal of the matrix [$\boldsymbol{R}$] although considering the human cone sensitivities does not make distinction between light and dark colors. The spectral comparison index (M_v) presents both properties of different weights for differences in lightness and consideration of the human cone sensitivities but its scale defies intuition. .

Conclusions

Based on this study, there is no metric that can be recommended as conclusively superior to others for all purposes. Until more is known, metric choices should be made based on appropriateness to applications, for example as detailed in Table III. We recommend that a combination of the metrics should be used to explore particular advantages of each metric. To gain more insight into the usefulness of the metrics, it is imperative to evaluate the spectral match metrics within a psychophysical experiment comparing their use on color patterns and complex images.

References

1. F. H. Imai, M. R. Rosen, D. Wyble, R. S. Berns and D. Tzeng, Spectral reproduction from scene to hardcopy I: Input and Output, in IS&T/SPIE Conference on Color Imaging: Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications II, M. M. Blouke, J. Canosa, N. Sampat, Eds., Proc of SPIE 4306, 2001, pp. 346-357.
2. Y. Miyake, Y. Yokoyama, Development of Multispectral and Color Imaging System for Recording of Art Painting, in *IS&T/SPIE Conference on Color Imaging: Device-Independent Color, Color Hardcopy and Graphic Arts IV,* G. B. Beretta, R. Eschbach, Eds., *Proc. of SPIE* **3648**, 1999, pp. 190-192 and 218-225.
3. H. Maître, F. J. M. Schmitt, J. P. Crettez, Y. Wu, J. Y. Hardeberg, Spectrophotometric image analysis of fine art paintings, in *Proc. IS&T/SID Fourth Color Imaging Conference: Color Science, Systems and Applications*, IS&T, Springfield, VA, 1996, pp. 50-53.
4. N. Tsumura, Y. Miyake, F. H. Imai, Medical Vision: measurement of skin absolute spectral-reflectance-image and the application to component analysis, in *Proc. of the Third International Conference on Multispectral Color Science MCS'01*, Eds. M. Hauta-Kasari, J. Hiltunen, J. Vanhanen, 2001, pp. 25-28.
5. F. H. Imai, M. R. Rosen and R. S. Berns, Comparison of spectrally narrow-band capture versus wide-band with a priori sample analysis for spectral reflectance estimation, *Proc. of in IS&T/SID Eighth Color Imaging Conference*, IS&T, Springfield, VA, 2000, pp. 234-241.
6. M. J. Vrhel, H. J. Trusell and J. Bosch, Design and realization of optimal color filters for multi-illuminant color correction, *J. Electron. Imaging* **4** 6-14 (1995).
7. J. P. S. Parkkinen, J. Hallikainen, T. Jaaskelainen, Characteristic spectra of Munsell colors, *JOSA A* **6** 318-322 (1989).
8. M. Rosen, M. Fairchild, G. Johnson, D. Wyble, Color Management within a Spectral Image Visualization Tool, *Proc. of IS&T/SID Eight Color Imaging Conference*, IS&T, Springfield, VA, 2000, pp.75-80.
9. M. Hauta-Kasari, J. Lehtonen, J. Parkkinnen, and T. Jaaskelainen, Spectral image compression for data communications, in *IS&T/SPIE Conference on Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts VI*, R. Eschbach, G. G. Marcu, Eds., *Proc of SPIE* **4300**, 2001, pp. 42-49.
10. CIE Publication 15.2, *Colorimetry*, 2nd ed., Commission Internationale de L'Éclairage, Vienna, Austria (1986).
11. CIE Publication 116, *Industrial Colour-Difference Evaluation*, Commission Internationale de L'Éclairage, Vienna, Austria (1995).
12. CIE Publication 142, *Improvement to Industrial Colour-Difference Evaluation*, Commission Internationale de L'Éclairage, Vienna, Austria (2001).
13. M. J. Vrhel, R. Gershon, L. S. Iwan, Measurement and analysis of object reflectance spectra, *Color Res. Appl.* **19** 4-9 (1994).
14. Haneishi, H., Hasegawa, T., Tsumura, N., Miyake, Y., Design of color filters for recording artworks, in *Proc. of IS&T's 50th Annual Conference*, 1997, pp. 369-372.
15. F. H. Imai, R. S. Berns and D. Tzeng, A comparative analysis of spectral reflectance estimated in various spaces using a trichromatic camera system, *J. Imaging Sci. Tech.* **44** 280-287 (2000).
16. J. Hernández-Andrés and J. Romero, Colorimetric and spectroradiometric characteristics of narrow-field-of-view clear skylight in Granada, Spain, *JOSA A* **18** 412-420 (2001).
17. H. S. Fairman, Metameric correction using parametric decomposition, *Color Res. Appl*. **12** 261-265 (1997).
18. R. S. Berns, Billmeyer and Saltzman's Principles of Color Technology, John Wiley & Sons, 2000.
19. J. A. S. Vigianno, The comparison of radiance ratio spectra: assessing a model's "goodness of fit", in *Advanced Printing of Conference Summaries: SPSE's 43rd Annual Conference*, Springfield, VA: SPSE, The Society for Imaging Science and Technology, 1990, pp. 222-225.
20. J. A. S. Viggiano, A perception-referenced method for comparison of radiance ratio spectra and its application as an index of metamerism, Proc. of AIC Color 01, The 9th Congress of the International Colour Association, Rochester, NY, 2001, in press.
21. I. Nimeroff, J. A. Yurow, Degree of metamerism, *JOSA* **55** 185-190 (1965).
22. J. A. S. Viggiano, Personal communication, 2001.
23. Cohen, J. B., Visual color and color mixture, University of Illinois Press, 2001.

Biography

Dr. Francisco Hideki has a Ph.D. in Imaging Science from Chiba University in Japan. Since September 1997 he has been working at Munsell Color Science Laboratory at Rochester Institute of Technology. His research has been focused on high-spatial resolution multi-spectral image capture and spectral reconstruction. He was named as the recipient of the 1998 Itek Award for the best student paper in 1997 by The Society for Imaging Science and Technology.