Color Management: New Roles for Color Transforms

Michael H. Brill
National Information Display Laboratory
Sarnoff Corporation, Princeton, New Jersey 08543-8619

Abstract

This paper discusses two color-management ideas whose time may have come: the use of image context in a color-management system (CMS); and a procedure for checking the effectiveness of a CMS with minimal recourse to the assumptions underlying the CMS itself. Both these ideas use color transforms, in different ways. The content-based CMS idea was introduced four years ago, based on finding the minimum color difference through a simple metric. A new metric derived from a vision model is now introduced. Color verification should also involve color transforms, to accommodate different white points. A procedure is described for such verification, which uses an illuminant/reflectance model to define error-free color transmission.

Introduction

Color management can be viewed as having two complementary goals, both of which should involve color transforms. The most familiar goal is to effect device-independent color reproduction. The second goal, less glamorous but equally necessary, is to measure the degree of success in achieving the first goal. I will present here a forward-looking discussion of both color reproduction and color verification, in which color transforms figure in prominent but different ways. *(1) Color reproduction.* Some success has been achieved in color reproduction by estimating and inverting device profiles (input-output relationships), and mapping colors pixel-by-pixel to and from a device-independent color space. This success might be improved by incorporating a vision model with spatial as well as chromatic dependencies. In this way, a color-management system could be developed that makes use of image content as well as device profiles. Several such vision models have been proposed, including Sarnoff's JND model.[1,2] I will summarize this model and suggest how it might be used to optimize color reproduction based on image content. *(2) Color verification.* What constitutes "perfect" color reproduction? Replication of tristimulus values seems neither possible nor even desirable. The design white point of the output device must be considered. I propose here an ecological definition of "perfect" reproduction: render certain test colors as if they were ordinary reflectances under an ordinary light with the chromaticity of the display white point. For test colors, I nominate eight reflectances that have low saturations (so as not to encounter gamut limitations of a particular device). Errors in such colors should uncover digital communication problems (which are rampant in today's mix-and-match world), without mistaking them for intrinsic gamut or other analog problems. Once the white point and gamut of the output device are addressed in this way, it will be possible to examine the fruits of our color-management labors, as it were, in dispassionate daylight.

Content-Driven, Visually Optimized Color Management

Ideally, a color-management system (CMS) converts the colors seen via device 1 (e.g., a CRT) to perceptually equivalent colors via device 2 (e.g., a color printer). If the color gamuts of the two devices are not the same, some of the more vivid colors rendered by device 1 may not be producible by device 2. In that case, a remapping (in some CIE color space) of *all* the colors rendered by device 2 will make this shortcoming less conspicuous. Conventional CMS technology performs this mapping directly from each pixel value, without recourse to the content of a particular image. But the perceptibility of color differences is very sensitive to image context. For example, color differences between textured materials are far less conspicuous than color differences between untextured ones. This means that the current CMS technology does not take full advantage of the properties of human vision. It may be important to field a vision-optimized CMS that is driven by image content. Furthermore, it is worth asking whether a single image-wide gamut transformation is rich enough to effect visual optimization, or if a spatially dependent set of control variables can also be defined.

The idea of content-based color-management was introduced by Amantea, et al.[3] The algorithm developed in the earlier work had three new features: (1) Control on the output color image was effected through a gamut transformation with a number of parameters to be controlled by the image content, but which operated globally over the whole image. (2) To assess the visual significance of a particular gamut transformation before final rendering, the initial and final images were compared in a few dimensions—energies in six spatial-frequency bands and in all CIELUV coordinates. (3) The gamut-parameter values were adjusted so as to minimize the difference between the image-energy vectors. The software had the basic architecture of Figure 1, but the energy-vector comparison was a stand-in for the Sarnoff Vision Model shown in the figure.

An improvement on the original content-based idea would be to replace the energy-vector comparison with a more detailed image comparison using a metric based on modeled just-noticeable differences (JNDs). The Sarnoff

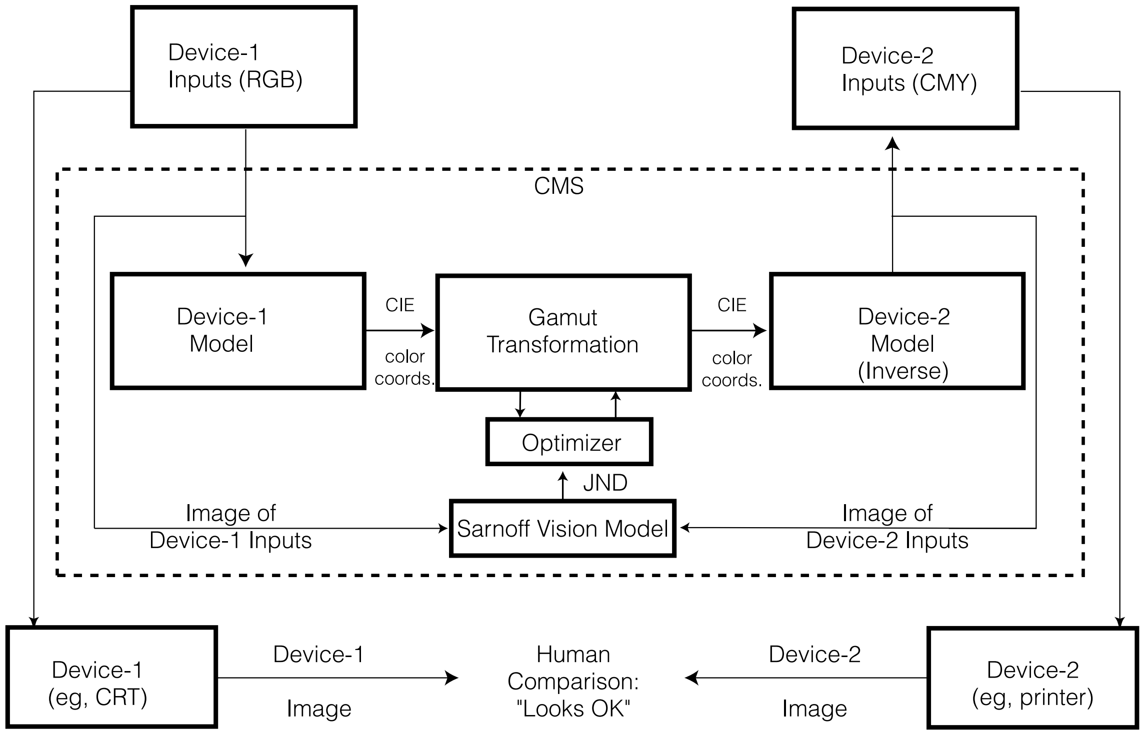

Figure 1. Content-Driven Color-Management System The CMS (in dashed box) contains the vision model and an optimizer. Note that the vision model must also contain forward Device-1 and Device-2 models. The diagram for a conventional CMS is the same as shown, but without either a vision model or an optimizer.

Vision Model[1,2] fulfills the conditions of a tested vision model based on JNDs. This model was developed partly under the auspices of NASA and ARPA to optimize LCD displays, and partly to create a new industrial product to evaluate visual fidelity of image-compression technologies.

The JND Model evolved in the following context. In 1948, Schade[4] at RCA found a useful frequency-domain metric for image quality on a CRT. Carlson and Cohen[5] refined this approach by partitioning the one-dimensional power spectrum of an image into several frequency bands, subjecting the computed contrast in each band to a static nonlinearity, and then comparing the results between two images as a metric of their visible difference. The model of Watson et al.[6] generalized the linear filtering stage to two dimensions, but does not have a point-nonlinearity after the filtering stage, and hence is accurate only at stimulus levels near detection. Daly[7] addressed this limitation by applying a nonlinear masking function after contrast computation, but has not yet generalized it from luma to chroma. Other similar models that include color [8,9] seem to predict color appearance, but as yet have no masking functions and no temporal dependence.

In this context, the JND Model is a unified approach with spatio-chromatic dimensions and masking. It takes in two images (or image sequences) and produces a single metric of perceptual differences between them, these differences being quantified in units of the modeled human just-noticeable difference (JND). The JND Model was first developed for static, achromatic images,[1] and later generalized to spatiotemporal and color domains.[2] The model is calibrated to fit sine-wave detection and discrimination data for chroma and luma. It also predicts psychophysical data for which it was not calibrated.

Inputs to the JND model are two images (one of which is shown in Figure 3)[†]. For each image, there are three digital data sets, characterizing the color control variables of a device. In Figure 2, the device shown is a digitally driven CRT, with inputs R', G', and B'. Front-end processing transforms these data to luma Y and chroma (CIELUV u*, v*).

Luma processing in the JND model accepts two images (test and reference) of luminances Y, expressed as fractions of the maximum (either of the display or of a reference-white reflector). First, a point nonlinearity (which depends on overall light level) effects luma compression. Next, each sequence is filtered and down-sampled using a Gaussian pyramid operation[10] to generate a range of spatial resolutions for subsequent filtering operations. Then contrast arrays (local differences divided by local sums) are calculated at each pyramid level, and scaled to be 1 when the image contrast is at the human detection threshold. Finally, these scaled contrast arrays are subjected to masking nonlinearities (to desensitize in the presence of image "busyness") and compared between test and reference to produce a JND map.

Similar processing occurs in each of the chroma images u* and v*, for which contrast is defined as a local spatial difference at each pyramid level. A point nonlinearity with influences from the luma channel effect the masking stage, and then test and reference are compared to produce a chroma JND map. The chroma and luma JND maps are

[†] text shown in shaded box should be disregarded

each available as output, together with a summary value to model an observer's overall rating.

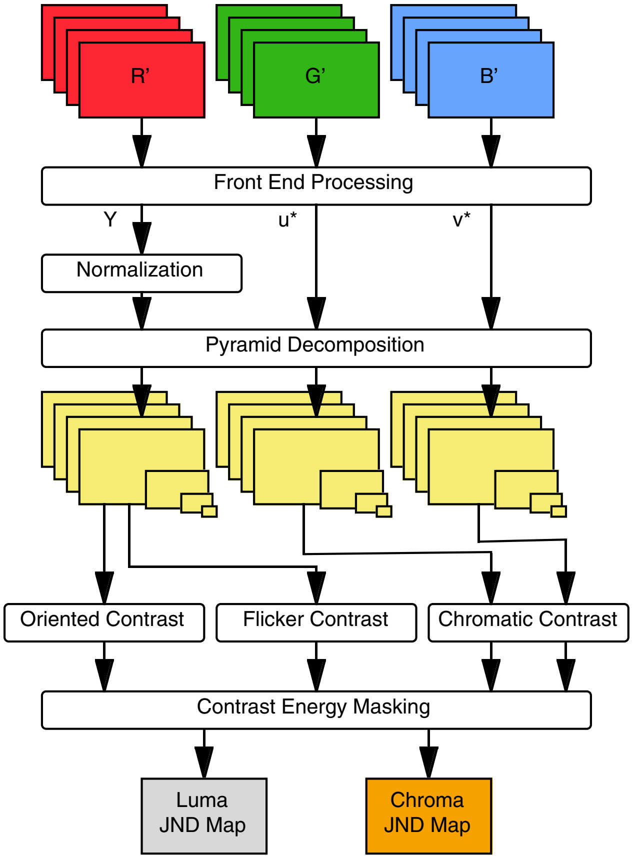

Figure 2. Architecture of the Sarnoff Vision Model. Note that one further step, the single-number summary of the JND map, is not represented in this figure.

The present form of the JND model compares two image sequences through the same device—one a compressed (distorted) version of the other. However, the distortion function can be replaced by a second device model (e.g., for a CMY or CMYK color printer) apropos of a color-management system. Whereas the current system has only one device model that processes both distorted and undistorted digital images, a CMS version would use two device models.

Using the JND model in a CMS promises to provide a minimum of visual distortion of the gamut parameters. The richness of the gamut parameters as control variables might also be enhanced, perhaps by making them depend on spatial location within the image or on spatial resolution of a pyramid-decomposition of the image. The problem of designing such dependencies into the mapping, and of optimizing them, are challenging problems for the future.

Color Verification

Given a CMS operating to achieve device-independent colors, it is important to be able to *verify* the colors—i.e., to determine by measurement whether given digital inputs produce output light with correct CIE XYZ tristimulus values. I will discuss verification in the context of electronic displays (VDUs), but it could be generalized to hardcopy devices as well. For displays, verification is not a simple "electrons-in, light-out" measurement, because modern display devices carry digital-media inputs directly to light outputs without an easy access to the "electrons-in, light-out" part of the process. Recent standards documents[11] have addressed CRT verification in a limited way through the measurement of certain responses to controlled RGB DAC drivers, but have not dealt with subtle digital incompatibilities "upstream" of the DAC.[12] To assess correctness of a complete system, one must feed a VDU to be tested with the same digital inputs (text pattern) as is a standard VDU deemed to be "calibrated", and then note the degree of colorimetric equivalence of the outputs. Besides the test pattern, one must also transmit enough information to compute a "correct", or target, set of tristimulus values for each color in the test-pattern set.

I describe here a procedure for verifying VDU colors that should detect digital incompatibilities, without mixing them up with gamut mismatches or with white-point incongruencies. Accordingly, a test pattern of nine full-screen patches is chosen to lie inside any foreseeable device gamut, but to subtend enough chromaticity for color verification. I choose Color-Rendering-Index (CRI) reflectances under a typical daylight spectrum with the same tristimulus values as the monitor white, and the monitor white itself. The chromaticities generated in this way are close enough to the monitor white that they are unlikely to be outside any device gamut—hence allowing a check for mistakes of the digital coding that might otherwise be confused with gamut-mapping problems. Also, the use of reflectances and model illuminants automatically compensates for change of the white point of the calibrated display relative to that of the standard display. The white-point compensation allows a user to test a display that has, say, a correlated color temperature of 9300 K. Such displays are plentiful, and must not be excluded from consideration. [An earlier approach[13] is similar to ours in using reflectance-based test colors, but assumes a D65 white point, and thereby sacrifices generality. Also, the standard is not a real VDU, but a device model.]

Based on these criteria, here is a procedure to verify test-display colors. In this procedure, off-line processing is to be performed once for all time, test-color preparation using a standard VDU is to be performed each time a new digital communication protocol is implemented, and test-VDU verification is to be performed each time a new device is introduced to a new protocol, and as needed after that. The procedure is outlined below.

A. OFF-LINE PROCESSING

Step 0. Compute tristimulus values of color-rendering-index (CRI) reflectances as if illuminated by daylight-eigenvector spectra. This computation uses: (a) The 1931 CIE XYZ color-matching functions[14] $x_j(\lambda_k)$, where $j = 1$ for X, etc. and λ_k assumes the 31 values from 400 to 700 nm at 10-nm increments; (b) the principal-component spectra of daylight[14] $S_0(\lambda_k)$, $S_1(\lambda_k)$, and $S_2(\lambda_k)$; and (c) reflectance spectra $r_i(\lambda_k)$ of the first eight Munsell reflectances used to compute the CRI.[15] Given these ingredients, compute the tristimulus values of each CRI reflectance plus white ($r_o(\lambda_k) =1$) under the three daylight principal-component spectra (eigenvectors). The result is a set of 81 values. Nine of these values define a 3×3 matrix **A** whose mj element is the j'th tristimulus value of the *m*'th daylight eigenvector:

$$A_{mj} = \sum_{k=1}^{31} S_m(\lambda_k)\, x_j(\lambda_k). \qquad (1)$$

The other 72 of these values comprise a $3 \times 3 \times 8$ array **B** whose mji element is the j'th tristimulus value of the i'th CRI reflectance times the *m*'th daylight eigenvector:

$$B_{mji} = \sum_{k=1}^{31} r_i(\lambda_k)\, S_m(\lambda_k)\, x_j(\lambda_k). \qquad (2)$$

The 81 numbers comprising arrays **A** and **B** are circulated with the digital standard for use in computing target tristimulus values for each device.

B. TEST-COLOR PREPARATION USING STANDARD VDU

Step 1. Select a standard VDU and measure its white-point tristimulus values. These values are the 1931 XYZ values $\mathbf{X}_{nR} = (X_{nR}, Y_{nR}, Z_{nR})$. At present, CRTs are well understood and readily used in this context.

Step 2. Compute CIELUV values of CRI reflectances under a model light. The model-light spectrum is the linear combination of daylight eigenvectors that has the same tristimulus values $\mathbf{X}_{nR}$ as the test-VDU white point. First, compute the coefficients in that linear combination:

$$(a_0, a_1, a_2) = \mathbf{X}_{nR}\, \mathbf{A}^{-1}, \qquad (3)$$

where **A** was computed in Step 0. Then, compute the target XYZ tristimulus values of the eight CRI-based test patches i under the model light:

$$\mathbf{X}_{iR} = (X_{iR}, Y_{iR}, Z_{iR}) = \sum_{m=0}^{2} a_m\, (B_{m1i}, B_{m2i}, B_{m3i}), \qquad (4)$$

where B_{mji} were computed in Step 0. Finally, compute target CIELUV coordinates[14] $(L^*_{iR}, u^*_{iR}, v^*_{iR})$ from $\mathbf{X}_{iR}$ and white point $\mathbf{X}_{nR}$. The choice of CIELUV is based on its historical use by the display industry.

Step 3. Select digital colors (that will be presented to all test devices). Arrange digital inputs to the standard device that result in measured CIELUV values $(L^*_{iR}{}', u^*_{iR}{}', v^*_{iR}{}')$, relative to white point $\mathbf{X}_{nR}$, that match the target values $(L^*_{iR}, u^*_{iR}, v^*_{iR})$ as closely as possible. The criterion of closeness here is $\Delta E < 3$ CIELUV units—the minimum perceptible color difference for colors that are not spatially next to each other.

Note: Step 3 is difficult because it involves iterative adjustment of the digital values until the correct tristimulus values are obtained. Fortunately, Step 3 needs to be done only for the standard VDU. One way of performing Step 3 is to use an accurate model of the standard VDU to make a good first estimate of the required digital values for each test-pattern color, and then to refine this estimate by experiment.

C. COLOR VERIFICATION OF TEST-VDU

Step 4. Measure XYZ values $\mathbf{X}_{nT} = (X_{nT}, Y_{nT}, Z_{nT})$ of the test-VDU white point.

Step 5. Compute CIELUV values of CRI colors under model test-light. The test light is the linear combination of daylight eigenvectors that has the same tristimulus values $\mathbf{X}_{nT}$ as the test-VDU white point. First, compute the coefficients in that linear combination:

$$(b_0, b_1, b_2) = \mathbf{b} = \mathbf{X}_{nT}\, \mathbf{A}^{-1}, \qquad (5)$$

where **A** was computed in Step 0. Then, compute the target tristimulus values of the eight CRI-based test patches i under the model test light:

$$\mathbf{X}_{iT} = (X_{iT}, Y_{iT}, Z_{iT}) = \sum_{m=0}^{2} b_m\, (B_{m1i}, B_{m2i}, B_{m3i}), \qquad (6)$$

where B_{mji} were computed in Step 0. Finally, compute CRI target CIELUV coordinates from $\mathbf{X}_{iT}$ and white point $\mathbf{X}_{nT}$.

Step 6. Measure test-pattern colors on the test device. Let the digital values derived from the standard VDU (step 3) drive the test VDU, measure the XYZ tristimulus values $(X'_{iT}, Y'_{iT}, Z'_{iT})$ of these 8 colors, and convert them to CIELUV using white point $\mathbf{X}_{nT}$, for comparison with the target values in Step 5.

Step 7. Compute CIELUV ΔE values for the measured test colors relative to the target values. Each ΔE value should be less than 10 for the color-transfer to be called successful. The value 10 comes from the observed variations of CIELUV coordinates across VDU screens.

Outlook

Elements of both the above ideas have been empirically tested. When evaluated against color-video sequences corrupted by compression distortions, the JND model correlates well with subjective rating data for these same sequences. When tested with a standard and four test-VDUs fed with the same digital RGB values, the color-verification procedure yields ΔE values that are acceptably small for the standard VDU, but not for the test-VDUs. Further work will determine whether the 10-unit criterion is realistic, and to what extent a CMS can ameliorate the results.

It might seem inconsistent to give color verification a white-point transform that is not the same as the CIELUV compensation inherent in the JND model. However, verification implies a constraint among visual *inputs* (i.e., naturalness of scenes) and the JND model implies a constraint based on visual *processing*. This is a basic conceptual difference. To equate these constraints by imposing a "vision-like" white-point correction on the test colors (such as the NTSC correction, that is Von-Kries in the tristimulus basis corresponding to the phosphor primaries) entails the following difficulties: (a) Colors in a natural scene do not transform this way; and (b) if vision itself performs a Von-Kries correction for white point, then artificial pre-correction should be unnecessary. Besides being intrinsically important, the conceptual difference noted here has the advantage that it does not bias the verification paradigm in favor of any standard human-vision models. Imposing the condition that white-point change should be the same as an illuminant change on reflectances seems the safest way to specify a "ground truth" for inter-device color reproduction for which digital sanity checks are desperately needed.

Acknowledgement‡

I thank the National Information Display Laboratory (NIDL) for sponsorship of the color verification procedure, and particularly Mike Grote of NIDL and Art Cobb of MIMA for reducing the procedure to practice. Also, I thank DARPA and NASA for sponsorship of the development of the Sarnoff Vision Model.

References

1. J. Lubin, A visual system discrimination model for imaging system design and evaluation, in E. Peli (ed.),Visual Models for Target Detection and Recognition, World Scientific Publishers, 1995.
2. J. Lubin, M. Brill, and R. Crane, Vision model-based assessment of distortion magnitudes in digital video, presented at the November, 1996 meeting of the International Association of Broadcasters (IAB).
3. R. Amantea, et al., "A content-driven color adjustment system," *IS&T and SID's Color Imaging Conference* (1993), pp. 228-232.
4. O. H. Schade, "Electro-optical characteristics of television systems. I. Characteristics of vision and visual systems," *RCA Review* **9** (1948), 5-37.
5. C. Carlson and R. Cohen, "A simple psychophysical model for predicting the visibility of displayed information," *Proceedings of the Society for Information Display* **21** (1980) 229-245.
6. K. Nielsen, A. Watson, and A. Ahumada, Jr., "Application of a computable model of human spatial vision to phase discrimination," *J. Opt. Soc. Am.* [A] **2** (1985) 1600-1606.
7. S. Daly, The Visible Differences Predictor: An algorithm for the assessment of image fidelity. In A. B. Watson (ed.), *Digital Images and Human Vision*, MIT Press, 1993, pp. 179-206.
8. A. B. Poirson and B. A. Wandell, "Pattern-color separable pathways predict sensitivity to simple colored patterns," *Vision Res.* **35** (1995) 239-254
9. T. Frese, C. A. Bouman, and J. P. Allebach, A methodology for designing image similarity metrics based on human visual system models, in *Human Vision and Electronic Imaging II,*, B. E. Rogowitz and T. N. Pappas, Editors, *Proc. SPIE* Vol. **3016** (1997) 472-483.
10. P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications,* **COM-31**, 532-540 (1983).
11. Commission Internationale de l'Eclairage (CIE), The relationship between digital and colorimetric data for computer-controlled CRT displays, *Publication CIE 122*, Bureau Central de la CIE, 1996.
12. G. Dispoto and M. Stokes, Limitations in communicating color appearance with the ICC profile format, *Proc. IS&T/SID 1995 Color Imaging Conference: Color Science, Systems, and Applications* , pp. 155-159 (1995); (see page 122, this publication).
13. L. Grambow and J. Antkowiak, Evaluation of color-TV displays by means of the Color-Rendering Index, *SID 97 Digest* , pp. 1041-1044 (1997).
14. Commission Internationale de l'Eclairage (CIE), Colorimetry (Second Edition), *Publication CIE 15.2*, Bureau Central de la CIE, 1986.
15. Commission Internationale de l'Eclairage (CIE), Method of Measuring and Specifying Colour Rendering Properties of Light Sources (Second Edition), *Publication CIE 13.2,* Bureau Central de la CIE, 1974.

☆ This paper was previously published in *IS&T/SID 5th Color Imaging Conference Proc.,* p. 78 (1997).

‡ acknowledgement added during the preparation of this volume