TRADEOFFS IN VDU MONITOR CALIBRATION AND IN COLOR CORRECTION

Michael H. Brill*

Abstract: Color calibration of a video display unit (VDU) determines the mapping from inputs (digital in the case of graphic VDUs and analog in the case of television) to tristimulus outputs. Assumptions are necessary to simplify this process. A universal assumption is spatial independence: that tristimulus values at a given monitor location are independent of the activations of other pixels on the monitor. Two other common assumptions are phosphor constancy (the relative spectrum of each phosphor is independent of its input) and phosphor independence (that the input to a phosphor influences the output of only that phosphor). In addition, assumptions are often made about the functional form of the phosphor luminance versus input gun voltage. The more of these assumptions one makes, the easier and more automatic can be the monitor calibration, but with a tradeoff against accuracy. Adjustments of the monitor are made during calibration to help ensure the invariance of the neutral chromaticity under scaling of the gun voltages by the same factor, to help match the gamut of the received colors to the gamut of the phosphors, and (in the case of digital inputs) to ensure adequate color resolution. These adjustments select a "monitor white" chromaticity. Color correction—replete with its own assumptions—then transforms a viewed scene from one illuminant to what it would appear under a light with the chromaticity of monitor white. This paper describes the implications and tradeoffs of all these assumptions.

INTRODUCTION

Much of the color science of video display units (VDUs) deals with two questions: the production of colors with given target tristimulus values, and the reproduction of colors that are either recorded by some other medium (as in television) or produced as part of a simulation of an illuminated scene (as in computer graphics).

Corresponding to the problem of color production is display calibration, the connection between input digital values and the tristimulus values of the output light. (Although such video displays as home television receivers operate without digital data, calibration is most often used in contexts for which digital data drive the phosphors.) The process of VDU calibration is laborious and most VDU users leave the details to the display

* Science Applications International Corporation, 1710 Goodridge Dr., McLean, Virginia, 22102

manufacturer. Often calibration is combined with adjustment of the monitor to achieve objectives of overall color balance, color gamut, and color resolution for the digital inputs. The adjustment is made on the mapping from the input digital values to the voltages that drive the phosphor electron guns. When this mapping is followed by the mapping (over which the calibrator has no control) from gun voltages to output tristimulus values, the result is the complete mapping from digital inputs to phosphor tristimulus outputs. One of the parameters of a calibrated display is the monitor white.

The problem of color reproduction is greatly helped by calibration of the VDU monitor. If a complete calibration can be made on a VDU, its pixels can be used as controlled light sources in visual experiments. In that case the VDU would be a quantitative colorimetric tool. It could also be used to produce light in a scene whose tristimulus values were exactly the same as those created by light reflecting from objects, either real or simulated. In television such a scene is produced from the responses of a video camera to a real scene (in which case the exact reproduction of color depends on the camera sensitivity functions being linear combinations of the human color-matching functions). In computer graphics the scene can be made by simulating surfaces with particular shapes and photometric properties, and simulating light reflecting from these surfaces. In both these applications it may be desired to represent a scene under (known or unknown) illumination as if it were illuminated by a known spectral power distribution. This latter problem is called color correction, and has been the subject of considerable research. In general, the chromaticity of the known spectral power distribution is taken to be the monitor white.

The present paper reviews the assumptions commonly made in monitor calibration and adjustment. As more of these assumptions are made, the required measurements become fewer, but with a definite tradeoff against accuracy. Then, recent models connecting input digital values to phosphor luminance are compared with respect to their implications for monitor adjustment and calibration. Mathematical symmetries in the models sometimes, but not often, simplify the adjustment process. In all cases, the monitor white is an important result of calibration and adjustment. Finally, color correction is discussed, predicated on the white point that is specified during monitor calibration.

Plausible Assumptions made in Calibration

As mentioned in the Introduction, calibration of a VDU involves finding the connection between input digital values and the output tristimulus values. By restricting the characterization of a light to tristimulus values, we have already assumed that using a system like that of the CIE properly represents the observer, and that observer differences are negligible. Granted the CIE standard observer, the strictest interpretation of calibration would involve recording the tristimulus values for all pixels (typically 8 bits of digital input x 3 image planes), for all possible spatial patterns of pixel activations by the three electron guns (typically on a 512x512 screen), and every time the monitor was used. Following the analysis by Brainard (1989), we note that the number of measurements for one such calibration would be $512 \times 512 \times (2^{24})^{512 \times 512}$, or about $10^{1,893,922}$. The resulting lookup tables would fill many replicas of the known universe. Clearly some simplifying assumptions have to be made.

An assumption that is always made is that of spatial independence: that the monitor's output at a particular pixel is a function only of the input values at that location. This simple assumption reduces the number of measurements required to perform monitor calibration to $512x512x(2^8)^3$, or about $4x10^{12}$. Fortunately, spatial independence is not a bad assumption to make about most monitors (Brainard, 1989).

Although $4x10^{12}$ measurements are storable in the known universe, so many measurements are still impractical in calibration. Further assumptions that are standardly made are gun independence and phosphor constancy. Gun independence states that the output of a particular phosphor depends only on the input corresponding to that phosphor, and is independent of the inputs for all other guns. This reduces the number of measurements to 512x512 pixels x 3 measurements/pixel, or about $2.4x10^7$. The factor 512x512 can be massively reduced by making the assumption of <u>approximate homogeneity</u>, whereby the calibration can be interpolated between a few pixels on the monitor.

Phosphor constancy states that the relative spectral power distribution of light from a phosphor is independent of the input digital value that drives the phosphor: only the <u>amount</u> of light varies when the digital value is varied, hence the phosphor chromaticity is constant. The assumption of phosphor constancy does not reduce the number of measurements as we have defined them, but changes the definition of most of the measurements to single luminance values instead of entire spectra.

Brainard (1989) and Berns, et al. (1991) assess these assumptions quantitatively using monitor measurements. Post and Calhoun (1989) suggest that colorimetric accuracy for some applications requires that phosphor constancy and phosphor independence <u>not</u> be assumed. However, if these assumptions are not made a pixel must be measured for all possible digital inputs to the three phosphors, and that is not often practical.

Having made the assumptions of spatial independence, phosphor constancy, and phosphor independence, one is still left with the task of measuring representative pixels on the screen for each one of the three phosphors (with the other two phosphors turned off), and for all possible digital inputs to that phosphor. The number of measurements is thereby reduced to only 3x256 = 768 for each 24-bit pixel that is measured.

If even this number of measurements is undesirable, it can be reduced still further by assuming a particular form (with some undetermined parameters) for the function relating digital input to phosphor luminance output. Only a few measurements may then be necessary to determine the parameter values for the input-to-luminance model of each phosphor. Such a reduced set of measurements is a sufficient calibration for many applications.

Even when all the above assumptions are satisfied, the combined process of adjustment and calibration can be intricate, and is generally accomplished in an iterative way. An initial assignment of digital values to voltages is made, the lookup table from voltages to tristimulus vectors of the individual phosphors is generated from these voltages, these tristimulus vectors are added to give the total tristimulus vector from all three phosphors together, and the set of such tristimulus vectors is used to change the mapping from digital values to voltages. One aspect of the monitor adjustment is to stipulate the monitor reference white.

Models for Phosphor Luminance

Subject to the assumptions of phosphor constancy, spatial independence, and phosphor independence, there are many models for the output luminance Y of a phosphor as a function of the input voltage V. For example, Meyer (1990) modelled the luminance Y of a phosphor by

$$Y = k(a + b\,V^{g}) \quad . \tag{1}$$

where g is the "gamma" of the electron-gun/phosphor combination, and a and b are constants of gun offset and gain. In general, a and b are both positive, and g has a value between 2 and 2.5 (Meyer, 1990). Although Eq. (1) seems to apply only to the luminosity of the light from the phosphor, it is actually a consequence of the following model that relates V to the spectral power distribution E(l) emitted by the phosphor:

$$E(\lambda) = K(\lambda)(a + bV^{g}), \tag{1'}$$

where the k in Eq. (1) is proportional to the wavelength integral of $K(\lambda)$ times the luminosity function. Note that, as a consequence of the assumption of phosphor constancy, the emission spectrum of the phosphor is a product of a function of wavelength and a function of voltage. Hence emission models analogous to Eq. (1') are always implied by luminance equations such as Eq. (1). Also, equations for CIE X and Z of the phosphor are derivable from these analogues of Eq. (1'), and will have the same form as Eq. (1). It should be noted, therefore, that although I refer throughout this paper to "phosphor luminance", I could equally well refer to "phosphor X-ness", "phosphor Z-ness", etc. The luminance coordinate has no special significance.

In contrast to Eq. (1), Berns, et al. (1991) modelled the phosphor luminance as

$$Y = k(a + bV)^{g} \, . \tag{2}$$

Of all the models reviewed here, the model of Berns, et al. seems to have the most detailed physical motivation: The video voltage maps linearly to a grid voltage at the CRT; this voltage, after offset for a cutoff, produces a beam current via a power function; and finally, the beam current is proportional to the phosphor luminance. The resultant mapping from video voltage to phosphor luminance has the form of Eq. (2), with the offset a <u>inside</u> the power function. It should be noted that this form is somewhat different from the form for which gamma is traditionally defined (which is actually the model of Meyer).

Post and Calhoun (1989) tested six models of phosphor luminance, all of them distinct from those of Eqs. (1) and (2):

$$Y = a\,V \tag{3a}$$
$$Y = a + b\,V + c\,V^2 \tag{3b}$$
$$\log Y = a + b \log V \tag{3c}$$
$$\log Y = a + b \log V + c\,(\log V)^2 \tag{3d}$$
$$\log Y = a + b\,V \tag{3e}$$
$$\log Y = a + b\,V + c\,V^2 \tag{3f}$$

Here,the coefficients are obtained by various forms of regression on measurement data. Each of the variables and coefficients, of course, bears an implicit subscript for the phosphor under discussion. Similar equations can, of course, be written for X and Z tristimulus coordinates; the a's and b's will be different, and will correspond to wavelength integrals of the underlying spectral power distributions multiplied by x-bar and z-bar color-matching functions instead of by y-bar.

These models have different implications for the calibration/adjustment process, which are reviewed in the next section.

The Process of Monitor Adjustment

Monitor adjustment has four main objectives: (1) to assure that digitally input whites, grays, and blacks have the same chromaticity—e.g., when the red (R), green (G) and blue (B) digital input values are equal to a single value I, the chromaticity of the resulting light does not depend on the value I; (2) to assure that a neutral chromaticity remains neutral when the "brightness" and "contrast" knobs (voltage gains and offsets) are adjusted on the video display; (3) to assure that there is enough digital color resolution to achieve any likely target color to a sufficient accuracy; and (4) to assure that the gamut of colors spanned by the digital values encompasses most of the target tristimulus vectors, and does not exceed the gamut of the monitor. In the present discussion of monitor adjustment, we assume spatial independence, phosphor constancy, and phosphor independence. Also, we assume that a particular pixel is being calibrated, and therefore do not discuss dependence on the location of the pixel on the screen.

Monitor adjustment as it relates to calibration has three requirements:

a. For R, G, and B phosphors, the model relating electron-gun voltage to phosphor luminance must be either determined from an empirical lookup table or expressed in a mathematical model with empirically fit parameters (such as Eqs. 1, 2, or 3).

b. For R, G, and B electron guns, there must be a digital-to-analog-conversion (DAC) model relating the input digital values I (e.g., from a stored image file) to the electron-gun voltages V. That model might have the form $V = AI + B$. When the voltage-luminance function of (a) above is composed with the functions of the DAC model, the result is a model of phosphor luminances as a function of input digital values. One can visualize the interaction of the DAC model with the voltage-luminance model by imagining the consecutive digital values as beads on a wire represented by the voltage-luminance curve.

c. For each of the DAC models, parameters that I will call <u>control variables</u> must be identified. Control variables are model parameters over which the calibrator of a video display exercises control during a monitor adjustment. For example, in the model $V = AI + B$, the gain A and the offset B might be accessible to control during calibration. If both the gain and offset of each DAC are control variables, then there are six variables that can be adjusted during calibration. Following the analogy in (b) above, adjusting the monitor corresponds to moving the beads on each of the "wires" that represent the voltage-to-luminance functions of the three phosphors.

To illustrate the three parts of the calibration/adjustment model, let us start with the voltage-luminance model of Eq. (2) for each phosphor. The DAC conversion is accurately modelled (Berns, et al., 1991) as the linear function

$$V = A\,I + B\,, \tag{4}$$

where the gain and offset A and B are the control variables. If this DAC model is combined with Eq. (2), then choosing the offset B as $B = -b/a$ will result in the following simple relationship between digital value I and phosphor luminance:

$$Y = k'\, I^{g}\,, \tag{5}$$

where k' is a constant that is not equal to the k in Eq. (1) unless the gain A is chosen to be 1/a.

Suppose the model represented by Eqs. (2), (4), and (5) applies to three phosphors (R, G, and B), all with the same value g but with other parameters (k, a, b) that are possibly distinct. This model and choice of control parameters has the advantage that the chromaticity incurred by three input values I_R, I_G, and I_B would be unchanged if the input values were scaled by the same constant. Scale-invariance is the digital-graphic analogue of "black-and-white compatibility" in television, and automatically ensures the invariance of any chosen monitor-white chromaticity (corresponding to $I_R = I_G = I_B$) under scaling of the digital inputs (I_R, I_G, I_B) by the same factor. [The white-point chromaticity can be selected, without compromising the model of Eq. (5), by changing the relative values of the gain A corresponding to the R, G, and B phosphors.]

A disadvantage of having chosen $B = -b/a$ for each phosphor (a prerequisite of Eq. 5) is that B is now negative, hence the lowest values of I produce values for V that are negative and hence not in the normal operating range of the device. Also, although all chromaticities are invariant under digital scaling, they are not invariant with respect to a change of all three voltages by the same factor, a change that is easily induced by changing the "brightness" knob on the front of the VDU. Of course, inducing the same voltage offset on the three phosphor guns (using the "contrast" knob) also changes the chromaticity.

Solving these problems requires a compromise. Not all chromaticities can be invariant to scaling of all the digital values ($I_i \rightarrow AI_i$) and also to scaling and offset of the voltages ($v_i \rightarrow \mu v_i + \nu$). However, under fairly realistic conditions a chosen neutral chromaticity can have all these scaling features. The analysis in Appendix A develops such conditions for the models of Meyer and of Berns et al., as well as for models (3a) and (3e) of Post and Calhoun.

In all the models considered, the adjustment that ensures invariance of the neutral axis (with respect to I and V) allows no freedom to specify the white point once the black chromaticity is determined. In fact, the chromaticities of white and black must be equal. (Meyer, 1990 noted this requirement for his own model.) It is therefore clear that the monitor black point specified by the display manufacturer enforces the monitor white point. This clarifies an argument that was not made clear in the earlier work of Brill and Derefeldt (1991), which called for a particular choice of monitor white point (the D65 chromaticity).

If the models of Eq. (1) or (2) are compromised (e.g., by the gammas not being the same) or if the control variables A and B in Eq. (4) are not set as prescribed so that input neutrals lie on the neutral axis legislated by the monitor black point, then colors intended to be neutral will not appear as such. Also, if a color to be construed as white is constructed from nonequal digital values, the chromaticity will tend to wander as the digital values are scaled down. This is partly because the same scaling factor applied to three nonequal digital values will cause nonequal roundoffs in achieving the final digital values. In view of these fairly serious consequences of failure to pay attention to the white point of the monitor, we describe in the next section a process intended to transform all colors to the specified monitor white point. In order not to have to exercise this process of color correction beyond its expected domain of validity, the choice of white point enforced by the display manufacturer will therefore be quite important and is discussed in the next section as well.

Color Correction and the Monitor White Point

We now examine the interaction of the monitor calibration and white point with color correction in the representation of colored objects under different illuminants. The discussion will be restricted to the television application, in which color correction attempts to transform between the appearance of a scene under a "studio illuminant" and its appearance under an illuminant consistent with the monitor reference white. However, the discussion applies equally well to color corrections applied to simulated scenes generated by computer graphics.

In television the choice of reference white has been necessitated by two considerations: exact (colorimetric) color reproduction and black-and-white compatibility. Television seeks to reproduce at a receiver the color recorded by the camera. If color reproduction is equated with the reproduction of tristimulus specification, the receiver must reproduce white as closely as possible to the chromaticity of the "studio" illumination, typically reflected from a matte white card. The reproduced color will be affected by, e.g., the illuminant spectral power distribution (SPD) under which the camera views the card, the nonlinearities of the video camera and in the phosphor guns of the VDU, the ambient illumination at the receiver, and gain adjustments made on each electron gun at the receiver.

Having established a reference white subject to all these influences, the television technology must ensure that this white does not change chromaticity when the white card is replaced by a gray or black one. Attention to this detail is necessary for black-and-white compatibility of the television system.

Neither chromatic adaptation nor color constancy removes the need for attention to the chromaticity chosen as the VDU reference white. Although chromatic adaptation is significant, it is never complete; consequently it is possible to detect an overall spectral bias for a collection of aperture colors, even if the perception of this bias is reduced by chromatic adaptation. Similarly, even an observer who recognizes portrayed object colors independent of the illuminant spectrum can still detect illuminant spectral biases (Arend and Reeves, 1986).

Manufacturers of televisions and graphics VDUs have made considerable use of various white standards over the years. But, oddly, not all manufacturers have adopted the same white standard. Rather, some (such as U. S. and Japanese manufacturers of graphical displays, and of some home television receivers) have chosen a standard white with a correlated color temperature (CCT) of 9300 K, and others (such as European manufacturers of home television receivers) have chosen a D_{65} standard, with CCT of 6500 K (Hunt, 1987). Brill and Derefeldt (1991) compared the merits of these different standards.

The choice of a D_{65} standard for VDU white evolved as follows. Forty years ago the N.T.S.C. specified that the reference white of television receivers should have the chromaticity of CIE Standard Illuminant C (Bingley, 1957). After the CIE replaced Illuminant C with D_{65} (whose CCT is quite close to the 6800 K CCT of Illuminant C), television manufacturers followed suit. Given any choice of reference white, it is possible to achieve exact colorimetric color reproduction (i.e., exact reproduction of the tristimulus values of reflected lights) only when the chromaticity of the reference white is the same as that of the studio illumination. However, when Illuminant C was first adopted as a standard, the lighting under which the cameras operated was incandescent near CCT = 3200 K, and not at higher color temperatures because that would have necessitated noisy carbon-arc lamps at the time. Because the studio illuminant color differed greatly from the monitor reference white, the other colors on the monitor would not necessarily resemble the original objects under daylight.

To reconcile the difference between monitor white and studio white involves an implicit assumption that we will call the gain control assumption: If studio light reflected from a white reflectance standard is forced to have a different chromaticity at the television receiver, then all other colors rendered by the receiver should approximate the colors actually seen under a commonly encountered light with the new chromaticity. One way to adjust the reference white is to adjust the gain factors on the three received TV signals. This is multiplicative color correction. The assumption that such an adjustment can simulate or compensate for an illuminant change on objects viewed by a TV camera is mathematically identical to assuming that the visual system can compensate for illuminant change by adjusting the gains of the color receptors—i.e., the model that von Kries adaptation can give color constancy. The limitations of this assumption for von Kries color constancy were discussed earlier (West and Brill, 1982), and are explained in the present context in Appendix B.

It has been noted recently (Trussell and Vhrel, 1991) that a much more satisfactory method than multiplicative color correction can be achieved using principal-component analyses of illuminant and reflectance spectra. The basic theory, originally advanced by Sällström (1973) and Buchsbaum (1980), is summarized in Appendix C. As in multiplicative correction, principal-component correction requires the equivalent of a white card that is used as a color standard at the time of measurement.

Troost and de Weert (1992) conducted a comparison, similar to that of Trussell and Vhrel, between multiplicative and principal-component color correction. However, they arrived at the opposite conclusion. Since the inaccuracies of color correction are less than three MacAdam units in both schemes, Troost and de Weert find that either is acceptable in practical applications in which colorimetric precision is not demanded.

Even if color correction is achieved with acceptable accuracy, more exact color reproduction is possible if the monitor white is as close to the studio white as possible. Early in the history of color television, however, the daylight standard was retained despite the difference between the chromaticities of daylight (monitor white) and studio illumination. This decision was largely motivated by the fact that much lower color temperatures than that of Illuminant C made television pictures look intolerably yellow (Bingley, 1957). It was fortunate that television cameras were later operated under daylight, so the daylight standard allowed more nearly exact color reproduction than was originally possible. These developments justified the N.T.S.C. choice of the Illuminant C chromaticity as a reference white.

CONCLUSION

This paper has summarized some theoretical aspects of color production and reproduction on video display units. The assumptions underlying monitor calibration and adjustment (including the selection of a reference white) were examined in terms of issues related to invariance of neutrals under voltage scaling. Also, two algorithms for color correction were reviewed because of the importance of restoring the color-space axis of object-color neutrals to the monitor neutral axis established during calibration. It can be concluded that accurate calibration of a VDU is very difficult, as is accurate color correction. However, for most applications accuracy is not critical and rapid calibration is essential. The assumptions reviewed in this paper have made calibration and color correction easy enough to serve these applications.

LITERATURE CITED

Arend, L. and Reeves, A.
1986. " Simultaneous color constancy", J. Opt. Soc. Am. A., Vol. 3, pp. 1743-1751.

Berns, R. S., Gorzynski, M. E., and Motta, R. C.
1991. "CRT metrology and colorimetric characterization techniques", Munsell Color Science Laboratory Technical Report (Rochester Institute of Technology, Rochester, NY, September, 1991).

Bingley, F. J.
1957. "Color vision and colorimetry", Television Engineering Handbook, D. G. Fink, Editor (McGraw-Hill, New York).

Brainard, D. H.
1989. "Calibration of a computer controlled color monitor", Color Res. Appl., Vol. 14, pp. 23-34.

Brill, M. H. and Derefeldt, G.
1991. "Comparison of reference-white standards for video display units", Color Res. Appl., Vol. 16, pp. 27-30.

Buchsbaum, G.
1980. "A spatial processor model of object color perception", J. Franklin Inst., Vol. 310, pp. 1-26.

Forsyth, D. A.
1990. "A novel algorithm for color constancy", Int. J. of Comput. Vision, Vol. 5, pp. 5-36.

Hunt, R. W. G.
1987. The Reproduction of Colour, Fourth Edition (Fountain Press, Tolworth, England), pp. 435-438.

Meyer, G.
1990. "The importance of gun balancing in monitor calibration", Perceiving, Measuring, and Using Color, M. H. Brill, Editor, Proc. SPIE, Vol. 1250, pp. 69-79.

Post, D. L. and Calhoun, C. S.
1989. "An evaluation of methods of producing desired colors on CRT monitors", Color Res. Appl., Vol. 14, pp. 172-186.

Sällström, P.
1973. "Colour and physics: Some remarks concerning the physical aspects of human colour vision", Report No. 73-09 (Institute of Physics, University of Stockholm).

Troost, J. M. and De Weert, C. M. M.
1992. "Techniques for simulating object color under changing illuminant conditions on electronic displays", Color Res. Appl., in press.

Trussell, H. J. and Vrhel, M. J.
1991. "Color correction using principal components", Image Processing Algorithms and Techniques II, M. R. Civanlar and R. J. Moorhead, Editors, Proc. SPIE, Vol. 1452, paper 1452-01.

West, G. and Brill, M. H.
1982. " Necessary and sufficient conditions for von Kries chromatic adaptation to give color constancy", J. Math. Biol., Vol. 15, pp. 249-258.

Appendix A: Conditions for Invariance of Neutral Chromaticity under Voltage Scaling and Offset

Assuming phosphor constancy as in the text, let the emission of the phosphor be

$$E(\lambda,\mu v_i) = K_i(\lambda)\, f_i(\mu v_i)\,, \tag{A.1}$$

where v_i (i=1,2,3) are the input voltages corresponding to the monitor white, and μ is a scaling parameter applied to the three white-point voltages v_i.

The tristimulus values X_j (j=1,2,3) of a particular neutral gray are then given by

$$X_j = \Sigma_i\, a_{ij}\, f_i(\mu v_i)\,, \tag{A.2}$$

where aij is the j'th tristimulus value of $K_i(\lambda)$.

Voltage-scaling invariance of the neutral chromaticity is ensured if X_j/X_1 is independent of μ. Transforming the tristimulus coordinates X_j by the inverse of matrix a_{ij} will not affect the μ-invariance of the chromaticity. Hence the chromaticity will be μ-invariant if $f_i(\mu v_i)/f_1(\mu v_1)$ is μ-invariant.

We first consider the form

$$f_i(\mu v_i) = b_i + (c_i + d_i\, \mu v_i)^g\,, \tag{A.3}$$

which includes the models of Berns et al. (1991) and Meyer (1990). Here, b_i, c_i and d_i are phosphor-dependent constants, and the phosphor systems are assumed to have the same g value. Assuming that b_1, c_1, and d_1 are nonzero, the quantity $f_i(\mu v_i)/f_1(\mu v_1)$ is μ-invariant if and only if

$$(b_i/b_1)^{1/g} = c_i/c_1 = d_i v_i/(d_1 v_1). \tag{A.4}$$

We now examine three subcases:

Case 1 (Meyer model): If $c_i = 0$, then the applicable part of Eq. (A.4) is

$$(b_i/b_1)^{1/g} = d_i v_i/(d_1 v_1). \tag{A.5}$$

This is equivalent to Meyer's statement that the chromaticity of the monitor white point must be equal to the chromaticity of the monitor black point. If in addition it is required that voltage offsets $v_i \rightarrow v_i + v$ (as in the "contrast" adjustment on a VDU), then $v_1 = v_2 = v_3$, and Eq. (A.5) simplifies. Of course, if the neutral chromaticity is invariant to gain and offset adjustments of all the voltages at once, the change of the DAC mapping $v = AI + B$ will not affect the neutral chromaticity either.

Case 2 (Berns et al. model): If $b_i = 0$, then the applicable part of Eq. (A.4) is

$$c_i/c_1 = d_i v_i/(d_1 v_1). \tag{A.6}$$

Again, the monitor white-point chromaticity must be equal to the monitor black-point chromaticity. As in Case 1, invariance to voltage offsets $v_i \rightarrow v_i + v$ requires $v_1 = v_2 = v_3$, and Eq. (A.6) is thereby simplified. The remarks about the DAC mapping that applied in Case 1 apply here too.

Case 3: The Post-Calhoun models represented by Eqs. (3a) and (3c) of the main text are special subcases of Case 1 and Case 2 above.

Although not a special case of Eq. (A.3), one more of the models by Post and Calhoun also has reasonable conditions for scale-invariance of the neutral chromaticity. The model of Eq. (3e) implies

$$f_i(\mu v_i)/f_1(\mu v_1) = \exp[\mu(b_i v_i - b_1 v_1)] , \tag{A.7}$$

for phosphor-dependent constants bi. These quantities (and hence the chromaticity) are μ-invariant if $b_i/b_1 = v_1/v_i$. As in Cases 1-3 above, offset invariance is assured by the equality of the voltages v_1, v_2, and v_3.

Appendix B: Spectral Assumptions Underlying Multiplicative Color Correction

For a scene photographed by a TV camera (or simulated by computer graphics) let the illuminant spectral power distribution $S(\lambda)$ be a linear combination of N basis functions $s_k(\lambda)$, with coefficients p_k:

$$S(\lambda) = \Sigma_k p_k s_k(\lambda) \ . \tag{B.1}$$

Denote by < > an integral over visible wavelength $\int d\lambda$, and denote a typical spectral reflectance in the scene by $R(\lambda)$. Finally, denote by $x_j(\lambda)$ the color-matching functions of the standard observer—linearly transformed such that the spectra for the phosphors of the VDU are the primaries. Then the tristimulus values of reflectance R under illuminant S are

$$<SRx_j> = \Sigma_k p_k <s_k R x_j> \quad . \tag{B.2}$$

If a white card in the scene has reflectance $W(\lambda) = 1$, then the tristimulus values of the white card under illuminant S are

$$<Sx_j> = \Sigma_k p_k <s_k x_j> \quad . \tag{B.3}$$

Because the color-matching functions are those for which the phosphors are the primaries, it follows that scaling the phosphor gains over the whole scene corresponds to separate scaling of the tristimulus values (in the basis of x_j). The operation of multiplicative color correction in a television system hence amounts to dividing each of the phosphor outputs (tristimulus values of reflectance R in Eq. B.2) by the corresponding tristimulus value of the white card in Eq. (B.3), and then multiplying the result by $<S'x_j>$ for the illuminant S' that has the chromaticity of the monitor white point. (The illuminant spectrum S' is also assumed to be a linear combination of s_k). The goal of such a rescaling operation is to produce "predicted tristimulus values" that are the same as those that would have come about from changing the illuminant from the studio illuminant S to the monitor-white-point illuminant S'.

The assumption that this scaling of phosphor gains is the same as an illuminant change on the photographed reflectances is accurate if the scaled tristimulus values

$$\Phi_j = \Sigma_k p_k <s_k R x_j> / [\Sigma_k p_k <s_k x_j>]. \tag{B.4}$$

are independent of the p_k (whether these coefficients are those of the monitor or studio illuminant). It was shown by West and Brill (1982) that Φ_j are illuminant-invariant if and only if $R(\lambda)$ is orthogonal to a "forbidden subspace" spanned by the 3N functions of wavelength

$$\Phi_{kj}(\lambda) = [<s_k x_j> s_l - <s_l x_j> s_k] x_j(\lambda) \quad . \tag{B.5}$$

This condition is very stringent (West and Brill, 1982), and is satisfied only for reflectances that occupy a small gamut in tristimulus space. Hence the gain-control assumption is not particularly justified in compensating for illuminant spectral power distributions.

In advancing the above argument, one should note a counterargument by Forsyth (1990). Forsyth proved that multiplicative gain control—or von Kries adaptation—is unique among color-constancy theories in placing no restrictions on $R(\lambda)$—hence the subspace of Eq. (B.5) is empty. However, the complete freedom of reflectances—and hence Forsyth's proof—depends on the color-matching functions $x_j(\lambda)$ being delta functions in wavelength. It turns out that the color-matching functions of the CIE standard observer are not narrow compared to wavelength variations of illuminant and reflectance spectra, and hence are not well approximated by delta functions. This means that von Kries adaptation may not be optimal in achieving color constancy.

Appendix C: Theory of Principal-Component Color Correction

Let illuminant spectral power distributions be approximated by linear combinations of three basis functions, as in Eq. (B.1) (with N = 3). Furthermore, let spectral reflectances be approximated by linear combinations of three basis functions $r_k(\lambda)$ with coefficients q_k:

$$R(\lambda) = \Sigma_i q_i r_i(\lambda). \tag{C.1}$$

Then the tristimulus values of a reflectance R under illuminant S are

$$\langle SRx_j \rangle = \Sigma_i q_i [\Sigma_k p_k \langle s_k r_i x_j \rangle], \tag{C.2}$$

and the tristimulus values of a white-card reflectance $W(\lambda)=1$ under the same light are

$$\langle Sx_j \rangle = \Sigma_k p_k \langle s_k x_j \rangle. \tag{C.3}$$

Now, it is assumed that, in Eqs. (C.2) and (C.3), the quantities in $\langle\rangle$ are either measured by the television camera (for the left-hand sides) or known a priori (for the right-hand sides). The right-hand-side $\langle\rangle$ quantities can be inferred ahead-of-time by principal-component analyses on illuminants (to determine the s_k) and on reflectances (to determine the r_i). Once all these quantities are known, the quantities p_k are inferred by inverting the linear equation system (C.3), and then these values are inserted into Eqs. (C.2) which are then solved for the quantities q_i. Color correction from illuminant S to illuminant S' by substituting quantities q into Eq. (C.1), retrieving an estimate for $R(\lambda)$, and then computing tristimulus values of R under the new illuminant S', using the functions $x_j(\lambda)$.