A COLOR SCIENTIST LOOKS AT VIDEO

Mark D. Fairchild

Munsell Color Science Laboratory, Chester F. Carlson Center for Imaging Science,
Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, NY 14623-5604 USA
mdf@cis.rit.edu

ABSTRACT

This paper represents a critical review of some of the color processing in the consumer TV processing chain. As such, a default processing chain is assumed as a starting point. The flow of color image information through this chain is described and critiqued. That is followed by development and description of some "clean slate" theoretical approaches to video processing with color accuracy and quality as the highest priority. These two approaches are compared and contrasted to provide some practical insight into how color science could be used in a practical sense to improve consumer video processing. Additionally, some examples how color and image appearance models might be used in the development of consumer video systems are described.

1. COLORIMETRY OF IMAGING SYSTEMS

Perhaps it is no coincidence that standardized methods of colorimetry have developed in parallel with various forms of commercially viable color imaging systems over the past century. As a reference point one could look at the establishment of the CIE 1931 Standard Colorimetric Observer only a few years before the invention of Kodachrome film. In fact, David Wright, one of the fathers of CIE colorimetry and an early researcher on color television, pointed out that had CIE colorimetry not existed prior to the development of television, it would have had to be invented.[1]

Color measurement in imaging systems can be organized into a three-level hierarchy from device-dependent metrics through device-independent techniques to the viewing-conditions-independent methods of color appearance models.[2] Device-dependent metrics are those that define the amount of color signal in a given imaging device without defining the meaning of those signals outside the particular device being used (*e.g.*, CMYK dye amounts or RGB digital counts). Device-independent metrics express the image information in terms of

colorimetric coordinates such as XYZ or CIELAB or device coordinates such as RGB or CMYK combined with colorimetric definitions of the meaning of those device coordinates (*i.e.*, a device definition, characterization, or profile). Viewing-condition-independent specifications recognize that the color appearance of any given scene or reproduction will depend on viewing conditions such as luminance level, surround, chromatic adaptation, *etc.* and attempt to specify final image appearance.

In consumer video applications, the framework is present for some forms of device-independent color imaging, but actual implementations are not controlled to the degree necessary and, unfortunately, much of the consumer video world remains in the domain of device-dependency.

2. OBJECTIVES IN COLOR REPRODUCTION

Hunt[3] defined six objectives in color reproduction and also reviewed them in his reference work on the reproduction of color.[4] These objectives are 1. spectral, 2. colorimetric, 3. exact, 4. equivalent, 5. corresponding, and 6. preferred color reproduction. While full description of these objectives is not possible in this short paper, they do provide the necessary theoretical framework for understanding the requirements for evolution of color reproduction from device-dependent to viewing-conditions-dependent systems. Consumer video does not presently achieve any of Hunt's objectives, but it might be said that it is aiming for preferred color reproduction. However, as Hunt[3] pointed out, the first five objectives "provide a framework which is a necessary preliminary to any discussion of deliberate distortions of colour reproduction."

Fairchild[2] has taken a slightly different approach to this question by defining five levels of color reproduction that cast Hunt's objectives into a different hierarchy. These levels are 1. color reproduction, 2. pleasing color reproduction, 3. colorimetric color reproduction, 4. color appearance reproduction, and 5. color preference reproduction. In this format, each successive level depends on the previous levels being achieved first. Consumer video systems are currently functioning at levels 1 and 2. The

following sections attempt to describe some possibilities for advancing to a higher level.

3. DISPLAY-CENTRIC VIDEO

Essentially from the moment of image capture, color in video is defined by some form of standardized display. For example, the encoded video signal (*e.g.* Y'C$_B$C$_R$) is defined by the display primaries, transfer function, and scaling.[5] Such a system can be sufficient to implement accurate device-independent color imaging for the displayed content. It could, but apparently is not, also be used for device-independent video capture if the cameras were colorimetrically characterized and the signal then encoded directly in display-centric characterized RGB (or a known transform thereof). However, the capture end of the system is rarely implemented in a colorimetrically-accurate method. Instead, the controls available to the videographer are used to capture "pleasing" images encoded for the chosen standard display rather than accurate measurements. For some color reproduction objectives this is perfectly adequate and appropriate, but for others accurate color information about the scene might well be lost at the very first step of the imaging chain. One is then left with processing and enhancing information for the display.

This display-centric (sometimes called output-referred) bias of video systems is not necessarily a flaw, but it needs to be recognized that the processing of video color is focused on the display rather than the original scene (a scene-centric, or scene-referred approach). Once the display-centric nature of video systems is recognized, the task becomes one of properly interpreting the colorimetric meaning of the video signals in order to do appropriate video processing (*e.g.*, differential processing of luminance and chromatic information or colorimetric transformations to displays that do not match the nominal standard display).

Unfortunately, notations such as YC$_B$C$_R$ are significantly overloaded and it is often difficult to determine just what the quantities represent (not to mention that there are other similar transformations with ill-defined names like YUV). Are the quantities linear or "gamma corrected"? If nonlinear, which transfer function was used? What primary set was used to define RGB? What coefficients were used to define the luminance transform? Is it even luminance? How are the quantities scaled and quantized? The questions seem endless and unless the image data encoding is accompanied with some definition of these variables, there is no chance for accurate colorimetry and meaningful perceptual processing of content. As Poynton[5] summarized, "the existence of several standard sets of primary chromaticities, the introduction of new coefficients, and continuing confusion between luminance and luma all beg for a notation to distinguish among the many possible combinations."

Given the potential for confusion (and the reality), how is it possible that acceptable results are obtained at all? The answer probably lies in the reality that very few, if any, displays match a given encoding standard, they are adjusted differently by users, and the viewing conditions have a significant impact on image appearance. Again, the best that can be done is to be aware of the potential for confusion, minimize it, and process the color information appropriately given what is available.

4. VIDEO PROCESSING

There are many steps in a typical video processing chain that can have an impact on the ultimate color quality. These include decoding, conversion of subsampling, color space conversion, artifact removal deinterlacing, frame-rate conversion, color/contrast/sharpness enhancement, and scaling (with some processes repeated multiple times along the pipeline). Interestingly, most discussions of video processing[6] do not address color issues at all beyond the selection of which video component to use. Clearly there is room for improvement by more careful use or manipulation of the components. A few issues worth consideration include linearity of the data (luminance vs. luma), orthogonality of the color space (luminance-chrominance crosstalk), and the treatment of colorimetric data (matching vs. appearance).

4.1. Linear vs. Nonlinear
The Rec. 709 transformation from linear RGB to relative luminance is given by Eq. 1.

$$Y = 0.2126R + 0.7152G + 0.0722B \qquad (1)$$

If one had a display with Rec. 709 primaries, Eq. 1 would provide an accurate measure of luminance when RGB are expressed as linear scalars (amounts of light). The important feature is that this is a linear combination of linear amounts of light. In video encoding, linear RGB values are rarely used. Instead a transfer function is applied to compensate ("gamma correct") for the nonlinear characteristics of historical display technologies (CRTs). When Eq. (1) is applied to nonlinear RGB signals, the computed quantity is no longer luminance (or even luminance with the transfer function applied) since a linear combination of nonlinear quantities does not produce the same result as applying the nonlinear transfer function after the linear combination of RGB. Sometimes, Y' is used to denote this quantity and it is referred to as luma to denote that it is different from luminance. Equation (2) represents the transform of nonlinear RGB, R'G'B', and is presented to stress the fact that the computed quantities are meaningfully different.

$$Y' = 0.2126R' + 0.7152G' + 0.0722B' \qquad (2)$$

The important issue for color accuracy is that the relationship between Y and Y' is not the same as the relationship between R and R' (for example) except in the special case of display neutrals for which R=G=B. The error caused by assuming that luma is equivalent to luminance is color dependent. It is worth examining this color dependence with a simple numerical example (with quantities scaled between 0-1 for simplicity). Take a neutral with the RGB triplet (0.21, 0.21, 0.21) and a full red with RGB of (1.0, 0.0, 0.0). Computing luminance using Eq. (1) with these linear RGB signals shows that Y=0.21 for both the neutral and the red. These stimuli are equiluminant. However, if the computation is done on nonlinear quantities (a simply power function with exponent 0.45 in this case), then a different result is obtained. For the neutral the R'G'B' triplet becomes (0.50, 0.50, 0.50) and the luma computed using Eq. (2) is Y'=0.50. This result illustrates that the nonlinearity can be applied before or after the computation for neutrals. However for the full red, the R'G'B' triplet remains (1.0, 0.0, 0.0) and the computed luma is Y'=0.21. Thus, the two colors that are of equal luminance have a luma difference that exceeds a factor of two! If one were to develop or implement a video processing algorithm that operates on luma under the assumption that it represents luminance, the errors or artifacts introduced can be very significant and will also be color dependent.

The problems caused by applying the luminance coefficients to nonlinear quantities are not an issue if it was done only as a convenience for encoding the video signals. However, as soon as any differential processing of the luma and color difference signals (sometimes referred to as chroma, although they do not meet the definition of chroma, just as luma is not luminance) will result in luminance and chrominance errors in the decoded result. Such processing includes partial gamma correction and subsampling of the color-difference signals. Thus, such errors are almost always present and certainly present in consumer video. Hunt[4] has nicely illustrated the color errors introduced by this encoding in video.

4.2. Orthogonal vs. Nonorthogonal
The same simple numerical example can be used to examine another issue in video encoding, crosstalk between the nominal luminance and chrominance channels. It is clear, from the example in the previous section, that there is some chromatic information in the luma signal since the departure from true luminance is color dependent. The other potential problem is the presence of luminance information in the color difference signals. In other words, are the color difference signals at constant luminance when luma is held constant? The answer to this question is known, but not widely-acknowledged, to be "no". For

example, take the nominal C_B and C_R computations (again with simple 0-1 scaling) given in Eqs. (3) and (4).

$$C_B = B' - Y'$$ (3)
$$C_R = R' - Y'$$ (4)

These quantities are sometimes referred to as "chroma" or "chrominance" with an implicit assumption that they include no luminance information. Examination of the equations reveals that the quantities will be correlated with luminance in a way that depends on the color. The importance of this in video processing again depends on both how the luma and color difference signals are differentially treated and on the magnitudes of the color differences being treated (as in noise reduction).

Again, some numerical examples are helpful. Take a pair of neutral colors with RGB triplets of (0.21, 0.21, 0.21) and (0.11, 0.11, 0.11). With linear processing they result in YC_BC_R triplets of (0.21, 0.0, 0.0) and (0.11, 0.0, 0.0) respectively. There is only a luminance difference between these two colors and that is properly reflected in the constant and zero color-difference signals. If the encoding is computed with nonlinear signals, the R'G'B' triplets become (0.50, 0.50, 0.50) and (0.37, 0.37, 0.37) and the respective $Y'C_BC_R$ triplets are (0.50, 0.0, 0.0) and (0.37, 0.0, 0.0). Again there are no crosstalk problems for a luminance change between two neutral colors.

The story is different for a chromatic stimulus however. Take the full red with RGB of (1.0, 0.0, 0.0) from the previous example and version of lower luminance with RGB of (0.50, 0.0, 0.0). The computed linear YC_BC_R triplets are (0.21, -0.21, 0.79) and (0.11, -0.11, 0.39) respectively. Similar crosstalk also happens with nonlinear processing. The R'G'B' triplets become (1.0, 0.0, 0.0) and (0.73, 0.0, 0.0) and the respective $Y'C_BC_R$ triplets are (0.21, -0.21, 0.79) and (0.16, -0.16, 0.58). Not only does the luminance signal "leak" into the color difference signals, but the leakage is 100%! The C_B signal is a perfect copy of the luminance signal (with a negative sign). Note that there is also proportional leakage into the C_R signal. Thus, the luminance crosstalk ranges from none for neutrals to 100% for the R and B primaries back to zero (interestingly) for the G primary. The result of this is that differential processing of the color difference signals (such as subsampling) will have an impact on the perceived luminance signal (e.g., noise or sharpness) in a way that depends on color. This explains why noise in these color difference signals often appears more visible than one would suspect if they were truly constant luminance signals. This issue occurs for both linear and nonlinear signals.

Video processing algorithms could be made more effective if they were implemented in perceptual color spaces in which the luminance and chromatic information is accurately separated. For example, subsampling or compression in chromatic channels that truly had no

luminance information would produce significantly fewer visible artifacts. Figures 1-3 provide a simple example that illustrates the potential improvement.

Figure 1. A full resolution (4:4:4 sampling) original image.

Figure 2. 4:1:0 chroma subsampling with no luminance crosstalk. This image appears nearly as sharp as the original due to the insensitivity of the human visual system to detail in the chromatic channels. (The severe 4:1:0 subsampling was chosen over the more common 4:2:0 subsampling to assure that the demonstration would be clearly visible.)

Figure 3. 4:1:0 "chroma" subsampling with 25% luminance crosstalk into the chromatic channels. This image appears significantly less sharp than Figs. 1 and 2 because the luminance information that "leaked" into the chromatic channels has also been subsampled.

Figure 1 is an original RGB image with full sampling in each of the three color components. In video terminology, this could be considered an image with 4:4:4 sampling. In Fig. 2, the image has been processed into an approximate perceptual space that separates luminance and chroma information, subsampled at 1/4 the original rate (both horizontal and vertical) in the chromatic channels and then reconstructed into an RGB with the original pixel sampling. This is approximately equivalent to 4:1:0 chroma subsampling, but in this case it has been applied to true chroma and luminance information (usually not the case in video processing). Figure 2 illustrates the fact that human observers are significantly more sensitive to spatial detail in the luminance dimension than they are in the chromatic dimensions, a property taken advantage of in video encoding, compression, and processing algorithms to some degree. Lastly, Fig. 3 illustrates the same 4:1:0 "chroma" subsampling in a space that allows 25% of the image luminance information to "leak" into the chromatic signals and also be subsampled. This level of leakage is representative of what might occur for a color one third of the way between the display white (or G primary) and the R or B primaries with current typical $Y'C_BC_R$ encoding and processing. The main point of this demonstration is that Fig. 2 illustrates how much better differential processing of luminance and chromatic information could be if performed in appropriate dimensions (where luminance was computed linearly and there is better orthogonality).

Song *et al.*[7] have examined the issue of orthogonality of color space for measuring and modeling the perception of noise in images and proposed a linear transformation of CIE XYZ tristimulus values for such applications. Of note is that the luminance dimension in their space almost perfectly matches the CIE Y tristimulus value (true luminance) so any space with the chromatic dimensions containing no luminance (Y) information would be a good approximation to their result. For example, CIE XYZ could be effective.

4.2. Appearance vs. Tristimulus Values

Even if all of the video encoding and processing were performed in a perfectly calibrated and characterized color space that matched the ultimate display colorimetry, there would be cases when the color reproduction would be incorrect. This is due to the difference between color matching specifications (tristimulus values such as CIE XYZ or calibrated RGB) and color appearance specification (reproduction of perceived lightness, chroma, and hue) and represents the transition between levels 3 and 4 of the levels of color reproduction described in section 2.

Tristimulus and appearance matching are identical when viewing conditions are the same for the original (or reference) and the reproduction. If the viewing conditions change, then tristimulus matches will no longer be appearance matches. Variables that impact color appearance include absolute display luminance, chromatic adaptation to the white point and surroundings, relative luminance level and color of the surround as well as more complex variables such as display size, viewing distance, and image noise levels.[2] Many of these variables are treated in more advanced color and image appearance models that could be applied to video processing algorithm development and testing once some of the lower-level colorimetric issues are addressed. To begin to move to such a level of color reproduction, questions of display characterization and adjustment need to be addressed.

5. DISPLAY ADJUSTMENT & CHARACTERIZATION

Consumer video has been blessed (or perhaps cursed) with a wide array of display technologies with significantly disparate capabilities (luminance, dynamic range, color gamut, *etc.*) While this provides an opportunity for significant improvements in video quality, it also poses a difficulty in consistency. To the degree no individual adjustments are made in the displays, the same color signals will produce different colors. However, when adjustments are made to attempt to adjust for differences, artifacts that are invisible on one display might become readily perceptible on another.

5.1. Colorimetric vs. Preferred

In the world of graphic arts and digital photography, display differences are accommodated through colorimetric characterization of the displays and definition of the image data in colorimetric terms. Such an approach is quite effective, but does leave issues of gamut mapping that are automatically treated (although in a poor way) with device-dependent imaging (if RGB aren't defined, then they always fit in the gamut of any RGB display!).

This type of color management based on display characterization is essentially nonexistent in video applications for two apparent reasons. One is that the video signals themselves are ill-defined colorimetrically (or ambiguous) making the utility of an accurate characterization questionable. The second is that the display manufacturers want to retain significant differences in display appearance to help differentiate their products from competitors and many consumers will make custom adjustments in their home environment as well.

Adjustments available to viewers are becoming quite sophisticated (such as independent adjustment of various color regions). There is potential for the accuracy of both the initial settings and the utility of user adjustments to be enhanced by more accurate colorimetric specification of the video signals and characterization of the displays. However, this potential might be small given the lack of control of viewing conditions and the unpredictable preferences of various users.

5.2. Gamuts: Chromaticity vs. Appearance

Display color gamuts have only been addressed indirectly in this paper in that they impact the computation of $Y'C_BC_R$ signals due to differences in standard display primaries and white points. While this is often the extent to which gamuts are considered in video applications (additive-mixture triangles on a chromaticity diagram), there are other aspects of the display that significantly impact the actual, three-dimensional appearance gamut. These include the display luminance, dynamic range (contrast ratio), and relationships between the display capabilities and signal encoding (*e.g.*, what percent of the display maximum is used for display white).

Perhaps the most important contribution color science can make to video is the representation and interpretation of color gamuts in three-dimensional appearance spaces (lightness-chroma-hue, or brightness-colorfulness-hue). For example, a display with a color gamut that appears to be a smaller triangle on a chromaticity diagram might actually look more colorful if the luminance level is higher. Heckaman *et al.* have shown that perceived color gamuts for projection displays correlate well with color appearance metrics[8] and that the perceived color gamut of a display can be varied dramatically by manipulating the relationship between the diffuse white point and the display maximum. [9] In fact with an extended-dynamic range display (11 bits

with 8 bits for diffuse white to black and 3 bits reserved above that level) it is possible to create a display with a color gamut that appears more chromatic than the spectrum locus of a chromaticity diagram!

6. SUMMARY

It is possible to imagine a whole new video capture, processing, and display system completely based on accurate color science principles. While such a "clean slate" approach might make the overall video systems simpler and more efficient, it is reasonable to doubt that end users would notice a meaningful difference. Beyond that, the replacement of current infrastructure is difficult at best as witnessed by the long, drawn out transition from analog to digital (and high definition) television that is yet to be complete. A more practical approach is to use the knowledge available in the field of color science to take advantage of the processing and display capabilities available to improve color accuracy and image quality.

6.1. Reality
The reality of consumer video processes is that images are captured and immediately encoded in a display-centric manner that loses color information about the original scene and confuses luminance and chromatic information in the encoding. This confusion becomes irreversible as soon as any differential processing is applied to the luma and color difference signals. Then a wide range of processing, compression, decompression, enhancement, transformation, and display manipulations are performed on these ill-defined color signals. Lastly, the signals are used to derive a largely unknown display in essentially random viewing conditions. Given that pessimistic summary, it is amazing that it works at all. It works because every step of the way has been tweaked for "pleasing color reproduction" and there are some standards in the definitions of the video signals that allow for consistency.

6.2. Ideal Theory
An ideal system, from a color science perspective, would capture and encode high-dynamic-range colorimetric information about the scene (or rendered subject matter), process that information in perceptually meaningful color spaces such as linear colorimetric spaces and nonlinear color appearance spaces (without ever confusing the two), and then display the results on an accurately calibrated and characterized display with the viewing conditions appropriately accounted for. Such a system is theoretically possible and some attributes of this have been implemented in the domain of digital cinema (and have certainly been implemented in color management systems used in the graphic arts).

6.3. Possible Practical Improvements
Given that the world of video can't start over from scratch, how can some color science principles be put to effective use? At the capture end there is probably little that can be done. However it might well be possible to make more accurate cameras, characterize them, and then record the camera settings along with the video information to allow scene colorimetry to be recovered more easily when appropriate. Assuming the encoding standards are not changed, the best that can be done is to be aware of the limitations of current encoding schemes and perform video processing in more perceptually meaningful color spaces. In other words, transform out of $Y'C_BC_R$ and into colorimetric spaces to do processing that is not directly linked to the display. Clearly this can not be done for every step in the chain, but it might improve some results. Lastly, at the display end, careful colorimetric characterizations might allow for more consistent display across technologies (perhaps not desired by some) and for more intuitive user adjustments and setup. Color appearance algorithms could also be used to meaningfully adjust the display for changes in the ambient viewing conditions. Finally color and image appearance models can be used to better quantify image quality at every step along the way.

7. REFERENCES

[1] W.D. Wright, "50 years of the CIE standard observer for colorimetry," *AIC Color 81, Paper A3* (1981).

[2] M.D. Fairchild, *Color Appearance Models, 2nd Ed.*, Wiley-IS&T Series in Imaging Science and Technology, Chichester, UK (2005).

[3] R.W.G. Hunt, Objectives in colour reproduction," *J. Phot. Sci.* **18**, 205-215 (1970).

[4] R.W.G. Hunt, *The Reproduction of Colour, 6th Ed.*, Wiley-IS&T Series in Imaging Science and Technology, Chichester, UK (2004).

[5] C. Poynton, *Digital Video and HDTV: Algorithms and Interfaces*, Morgan Kaufman, Amsterdam (2003).

[6] G. de Haan, *Video processing for multimedia systems,* ISBN: 90-9014015-8, Eindhoven (2000).

[7] X. Song, G.M. Johnson and M.D. Fairchild, "Minimizing the perception of chromatic noise in digital images," *IS&T/SID 12th Color Imaging Conference*, Scottsdale, 340-346 (2004).

[8] R.L. Heckaman, M.D. Fairchild and D.R. Wyble, The effect of DLP projector white channel on perceptual gamut, IS&T/SID 13th Color Imaging Conference, Scottsdale, 205-210 (2005).

[9] R.L. Heckaman and M.D. Fairchild, "Expanding display color gamut beyond the spectrum locus," *Color Res. Appl.* **31**, 475-482 (2006).