Making Color "Plug and Play"

Todd Newman
Silicon Graphics, Inc.
Mt. View, CA

Abstract

The ICC members have determined that they want a single, consistent, well-defined model for color management. This paper first examines some of the issues that such a model addresses. It then discusses problems with the ICC specification that make it impossible to address these issues and the changes needed to solve those problems. Only by the adoption of these changes can the ICC meet our users' desire for "plug and play" color.

Introduction

About two years ago, it became clear that the members of the International Color Consortium had strongly divergent ideas about the goals of the organization. These differences were causing us to work at cross purposes, so we decided to spend some time coming to explicit agreement on the goals of the organization. The key agreement was "that users want to 'plug and play.' They want to pick devices, platforms, software, etc., from different vendors and have it provided the expected result. In short, they want color to 'just work.'''[1] This paper discusses what it means for color to "just work" and what the ICC will have change to make this happen.

There are two markets that may be treated as the litmus tests of whether we have achieved "plug and play" functionality. The first would be the film animation market, discussed below. That market is probably the most demanding of image quality and the farthest from traditional color management markets, so it can test whether we have addressed the full range of needs.

The other market is one I have called "distributed color,"[2] which uses the Internet or private networks for delivery of color images. An obvious example is the delivery of color images over the World Wide Web. Here the author of an image is not in physical contact with the person using the browser. There is no way for the author to reject an inadequate reproduction and demand a reprint. Also, the author cannot guarantee what tools will be available to the image consumer. The only recourse is through well-defined and broadly supported standards.

This paper first lays out a framework for considering the purpose and operation of color management systems. Then it discusses the problems with the ICC profile specification that prevent us from reaching the goal of "plug and play" color. Finally it presents some approaches that could let us solve these problems.

The Function of Color Management

The primary task of a color management system is to enable us to make a reproduction image where the colors look like the original image. Ideally, this similarity can be achieved despite changes in media, illuminant white point, illuminance level, image surround, or image background. Further, we would like to preserve this similarity even when the output medium has a different dynamic range than the input medium.

Color management systems achieve consistency in reproduction by performing three operations: colorant matching, appearance modeling, and gamut adjustment. Of course, these operations may be combined into one step. Colorant matching is the determination of the proper combination of colorants to match a required tristimulus value. Appearance modeling is the adjustment of tristimulus values to account for differences in the viewing conditions. Finally, gamut adjustment is the intentional distortion of local color appearance to accommodate all or most image colors in the destination device's limited gamut. There are no standards for any of these operations. There are several common algorithms for colorant matching; Tony Johnson[3] presents an excellent summary. Only recently, with the work of CIE TC1-34,[4] has there been the outlines for an international standard for color appearance modeling. Gamut adjustment remains a proprietary and highly secret art.

These three operations could be performed in many different ways. The ICC has adopted a general model in which device information is keep in a file called a profile.[5] A software module called a Color Manipulation Model, or CMM maps pixel data from source device space to destination device space. (The ICC specification also supports abstract profiles that do not represent any physical device. Their behavior is beyond the scope of this paper.) The CMM is primarily an interpolation engine, all the intelligence for colorant matching, appearance modeling, and gamut adjustment is contained within the device profiles. For most profiles there is no requirement to store any measurement data, either of device response or viewing environment, in the profile. Instead, the profile contains a multi-dimensional color lookup table that embodies a precomputed transformation from device color space into something called the profile connection space (PCS). Each profile provides a way to map both ways between device coordinate space and the PCS. Thus, any two device profiles may be linked together via the PCS. As will be discussed below, putting all the intelligence in the profiles and not standardizing on algorithms can lead to problems.

Problems in The ICC Specification

In a sense, all the problems that arise for the ICC in traditional media are the result of companies desiring that the ICC provide a framework for a solution rather than a solution proper. When the ICC was first formed there was no standard method for appearance modeling. We could have decide to share our algorithms and do the best we could in the absence of outside leadership. Instead, software vendors saw this as an opportunity to differentiate their products, rather than a limitation to the ICC’s effectiveness. So the ICC standardized on a profile specification that was a container for information, but did not standardize on algorithms for colorant matching, appearance modeling, or gamut adjustment.

Private tags

The ICC specification allows for private tags. These are data added to an ICC profile with no public specification for their use. Only those who know the semantics of the private tags can use them. An example might be measurement data of color swatches outside the gamut of the IT8 target. Such measurements would be needed to make a mapping of out of gamut colors. As long as all necessary information is provided in the standard tags, private tags are tolerable. However, any time private tags are used, we can expect inconsistent results from the CMM that knows how to use the tag. Assuming that the private tags were added to improve image quality, this means that image quality is left to suffer on open systems.

For many of my company’s customers, predictability of results is at least as important as quality. If they can predict the behavior of the system, they can work around any deficiencies. However, if some CMMs use private tags and others do not, our customers can no longer predict their results. For them, this problem far outweighs any gain in image quality that they may see on an individual device.

Ambiguity in the specification

The next major causes of inconsistency are differing interpretations of the specification. It is very hard to write a specification. Details that seem perfectly clear and unambiguous when writing the specification prove to be muddled and vague when one actually sits down to start programming. For example, the tag for the media white point "is referenced to the profile connection space (PCS) so that the media white point as represented in the PCS is equivalent to this tag." What does that mean? Consider the media white point for a D65 monitor. When one views the monitor, one's eye is adapted to that white point. In the PCS, one's eye would be adapted to D50 as white. Should the media white point for a monitor should always set to be D50? One profile vendor believes that viewers are always completely adapted to the monitor white point and so sets the media white point tag in all monitor profiles to D50.

After careful study, I believe the definition is useless. It says that the value of the media white point tag should be set equal to the media white point, but does not tell us ***how*** to represent the media white point in the PCS in the first place. A mathematical model would have revealed this gap immediately, but it is easily lost in a welter of words.

Intents

People have differing goals for color reproduction. The ICC has tried to capture those different goals with a concept called "rendering intents." One might say that the different rendering intents are designed to producing pleasing color reproductions by emphasizing the preservation of different aspects of the original image. Depending on what you want to do with the image, one or the other of these aspects will be of paramount importance. Other aspects of the image quality might not be preserved as faithfully. As you might expect by now, these intents are not well defined. Because the ICC has no standards for appearance matching and gamut adjustment, it is not possible to formally define how the selection of different intents alters color reproduction.

The four intents optimize for: colorimetry ignoring viewing conditions (Absolute), colorimetry relative to media white point (Relative) relative saturation (Saturation), and color balance (Perceptual). Only for the first two intents is there anything resembling a mathematical definition – and that definition does not extend to colors that are out of the destination device’s gamut. But the Absolute intent ignores all factors in the viewing environment. Areas in the source and reproduction may have the same measurements on a tristimulus device, but they may well not be the same color. The Relative intent factors in the background color, but no other parameters of the viewing conditions.

Neither the Absolute nor the Relative colorimetric intents use a sophisticated color appearance model. Profiles built using the Perceptual intent may incorporate a sophisticated appearance model, but users cannot be sure whether they will or which they might choose. If a profile seems to work well with the Perceptual intent, users can not be sure anyone else can get the same results unless they use the same profile. They are forced to choose between consistency and quality.

No Definition for the CMM

Not only did the ICC choose not to standardize on methods of profile construction, it did not standardize on the behavior of the CMMs either. Again, this was considered to be a valid area for vendor differentiation. Usually, this should not produce noticeable problems for the users, but in one circumstance it may.

The main purpose of the CMM is to use the color lookup tables in the device profiles to map device values into and out of the PCS. Color lookup tables are sparse; they do not provide an explicit mapping for every possible pixel value. CMMs must interpolate between the defined values. The ICC has not defined a standard method of interpolation. However, studies[6] indicate that the most common interpolation algorithms all produce results that are very similar: error values are within one CIELAB delta E. This would suggest that it is not vitally important for the ICC to standardize on a single interpolation algorithm. Given the same profiles, results will not differ significantly. However, I see no good reason not to standardize the interpolation algorithm.

ICC monitor profiles need not provide multi-dimensional color lookup tables. Instead, they may characterize the monitor by indicating the monitor’s white point, phosphor chromaticities, and tone response curves. This leaves the CMM with the jobs of colorant matching, appearance modeling, and gamut mapping. Essentially, the CMM builds an implicit profile. Once again, the ICC has not defined how this should be done. So there is no reason to think that two CMMs will produce the same results given the same monitor and monitor profile.

Gamut Stretching

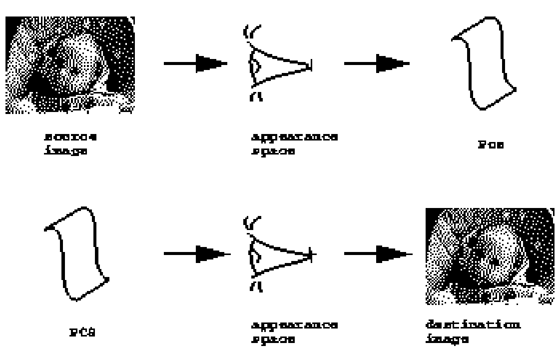
One of the more interesting color problems I have been involved with is film animation. Several studios are trying to take advantage of computer technology without sacrificing the hand drawn qualities they love. Backgrounds are hand painted. Characters are hand-drawn, the drawings are scanned in, and then ink and paint is done on a computer screen. These must be composited and then output to a film recorder. Image sources come from different media, canvas, paper, and monitor. The viewing conditions range from lighting booths for viewing the painted materials, to darkened offices for design, to movie theaters for viewing the film. While motion picture film is not a new medium, the attempt to apply color management techniques to film production is a recent phenomenon. Let us examine to problems that arise. We must solve these problems if we are to extend the “plug and play” capabilities to high end markets.

In the animation world, characters are “inked and painted’ on the CRT. If one colorimetrically reproduces the output on film, the result looks flat and lifeless. We have not used the full gamut of the film; the monitor just does not have the dynamic range of film. Traditional color management is always dealing with situations where the source image medium (be it slide film, paper, or monitor) always has at least as much dynamic range as the output medium. The traditional color management problem is: how do we compress the input images gamut to fit into the gamut of the output device? There are no standard algorithms for gamut stretching. It is an interesting question whether a color appearance algorithm that preserved colorfulness would suffice to solve this problem.

Appearance Modeling

Despite its central importance, there has never been extensive discussion within the ICC of the "ideal reflection print" model. The rationale behind the choice is not intuitively obvious, and this has led to great confusion for those trying to implement the ICC specification. The model stems from the lack of a viewing condition independent appearance model. The CIEXYZ space does not specify how a color appears, only the tristimulus values at the image surface. CIELAB does describe appearance, but the appearance is specific to the viewing conditions in which the measurement was made. The CIELAB algorithm does not factor degree of adaptation, adopted white point, background chrominance and luminance or surround into the appearance model. So a CIELAB value becomes inaccurate when any of these parameters change. Thus, when the ICC members wished to link two device profiles together using CIELAB, they had to define a reference viewing environment to which colors were normalized.

Figure 1: Current PCS



I believe that the reference viewing conditions were defined as they were because these conditions are very familiar to the graphic arts industry and hence to the traditional color management industry. They are the conditions under which graphic artists traditionally have assessed the quality of color reproduction. Also reproduction on paper has been the main application of color management systems, so it was a reasonable choice.

Remember that the PCS is defined as colorimetry in these viewing conditions. It is difficult to comprehend colorimetry – measurements – without there being something to be measured. So the ICC specification introduced the notion of an ideal reflection print. This notion is responsible for a great deal of confusion among experts debating the nature of the ideal medium. It is a reflection print because that is the only medium that can sensibly be viewed in the chosen reference viewing environment. Projected media such a slide film would not be viewed in a light booth, nor would a television. Since the medium is hypothesized only as a vehicle for the colorimetry, it must not restrict the possible colorimetric values. Hence the comment in Annex E that the medium uses “colorants having a large dynamic range and color gamut.” The exact details of the medium and colorants are not specified and do not need to be. Nor need we know the tone reproduction curve of the medium. Once you have the colorimetry (in a defined viewing context), these factors are irrelevant.

Similarly, several authors have said that there is flare in the PCS. (Dispoto and Stokes[7] propose one percent; Setchell and Giorgianni[8] at between 0.5 and 1 percent.) To the contrary, Annex E declares that measurements are done in a way that is flareless, as if any actual measurements had to be made. Any flare would limit the possible colorimetric values, making it impossible to reach a luminance of 0.0. The PCS viewing conditions must allow encodings for all conceivable appearances.

Unfortunately, the PCS as defined does not allow the encoding of all conceivable appearances. Imagine the somewhat unlikely scenario of someone standing in a field holding a MgO disk (a nearly perfect diffuse reflector) in one hand and a signal flare in the other. If it helps, you can imagine that he is a color scientist. The MgO disk is going to be brighter and whiter than nearly everything in the scene. But the signal flare is going to be a lot brighter than the MgO disk. The specification states that “since the PCS represents an ideal reflection print, and the media is a perfect diffuser, the largest valid XYZ values are those of the PCS illuminant.” But the PCS illuminant is only 2000 lux. There is no way to represent any light source more intense than that. There is no way to represent the brightness of the signal flare colorimetrically in the ICC encoding.

One could, of course, compress the luminance range while encoding into the PCS. One would represent the signal flare in the PCS with a luminance of 1.0 and represent the MgO disk with a lower luminance. By applying a "toe and shoulder" tone reproduction curve, one could produce a pleasing reproduction. At least one ICC profile vendor does just this. This approach conflates the appearance modeling role of the PCS with gamut adjustment, and it does so going into the PCS. Once we have thrown away that information, it cannot be recovered. If the output medium is one like motion picture film that can produce a wider range of luminances, there is no way to recover the original device information and reproduce it faithfully. This would be most unacceptable if the original image came from a motion picture film scanner. The current PCS is unable to represent these "whiter than white" values.

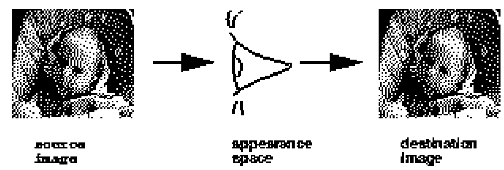
Simplifying the PCS

The ideal reflection print model begs the very question it is designed to answer. Most profiles are not designed to be used in a viewing booth, the reference viewing environment for the ICC. Profiles made for any other environment will need some way to simulate how colors would appear in the PCS viewing environment. For that, we need a viewing condition independent color appearance model. As Dispoto and Stokes have shown[9], makers of input device profiles must perform a two-stage process. First, determine how the color appears in the measurement conditions. Second determine what tristimulus value would be required to reproduce this appearance in the PCS viewing environment. Similarly, output device profile makers must first determine how an image in the PCS viewing environment with the specified tristimulus values would appear, and then how to reproduce such an appearance on the output device. Far from escaping the need for a viewing condition independent appearance model, we now require its use twice in every device profile. Profile vendors are required to choose their own model, with no guarantee that others will pick a compatible model.

Ironically, if we select a standard color appearance model, so that all profile vendors will build consistent profiles, we no longer need a viewing context dependent PCS. We can just use the appearance model's output values as the PCS. Profiles would just provide a mapping to the appearance space and could be linked in the appearance space.

The new CIECAM97 model includes predictors of lightness and colorfulness that should make it possible to define a workable mapping for the Saturation intent.

Figure 2: Simplified PCS



Gamut Adjustment

Sophisticated, standardized appearance matching will not solve all the problems discussed above. As Poe observes, "The most satisfactory renderings on different media will, in general, not have the same color appearance, because they will be 'tuned' to the characteristics of the individual media, adjusting gracefully to their limitations and making effective use of their capabilities."[10] There are no standard gamut adjustment algorithms. This has also been one of the major areas of vendor differentiation and the consequent lack of consistency in rendering.

Experts have found that the optimal method of gamut mapping is often image dependent.[11] Nevertheless, the ICC could define some standard algorithms. Ideally, we could then define a technique for determining which algorithm to use based on the image content. Then, for the most color critical applications, we could have the option of using the best algorithm and still support distributed color management. Falling short of that, profiles could indicate which of several standard gamut mapping techniques were used. Sophisticated users could visually inspect an image and specify that a profile using the appropriate technique be used for reproduction.

Colorant Matching

As mentioned above, this area has been discussed extensively in the literature. There is no evidence that the selection of different techniques in this area has lead to problem in color matching. It is possible that these problems exist but are hidden by more significant problems caused by differences in appearance matching and gamut adjustment. After we gain experience with the quality of results achieved when color management systems share a common color appearance model and gamut matching techniques, we can determine if this is a remaining issue.

A Recipe for Success

We need a standard color appearance model. If the appearance model works properly, it will provide enough information to solve the saturation intent, gamut stretching, and "whiter than white" problems. The recent work of CIE Committee TC1-34 provides a basis on which we can build. We need to test the model extensively. If it does not meet our needs, we should work with the Committee to refine the model until it does.

We need a reference implementation of the ICC specification. There are other problems in the ICC specification beside the profile connection space. Some sections are ambiguous and others are insufficiently specified. Because the work of the ICC is split between the profiles and the color manipulation modules (CMMs), I believe we need reference implementations of both profile builders and the CMMs. A reference implementation solves two problems: it forces us to address all the ambiguities in the specification; second, it provides a baseline set of choices in areas like interpolation models, color appearance models, and gamut mapping techniques. I believe this is the only way that the ICC will be sure that we have addressed all these issues; and only by addressing all of them will we be able to assure consistent results.

The reference implementation does not supplant the specification. A written specification makes it clear the way that vendors can improve upon the reference implementation and still be in conformance with the specification.

Finally, I think that the ICC has not done a satisfactory job defining the rendering intents. I believe that most users are actually looking for two intents that have not yet been defined. The first combines a sophisticated color appearance model with a standard gamut adjustment algorithm. This intent, call it the Picture intent, would be used for most image reproduction purposes. A second new intent, the Spot intent, would provide appearance modeling but not gamut adjustment. Colors out of gamut for the destination device would just be mapped to the nearest in gamut color. This intent would assure that in gamut spot colors would be reproduced accurately, and out of gamut colors as best as possible. These two intents would allow people to use a “distribute and print” model and still get predictable results without having to specify which CMM and profiles were to be used.

I have outlined some of the problems with the current ICC specification. These problems can be fixed by: altering the PCS to use a single, sophisticated, well-defined appearance model; defining standard gamut mapping algorithms; providing a baseline mathematical model for the ICC specification; and by providing a reference implementation. Only when we have done this, will it be possible for users to get the “plug and play” behavior they want.

---

[1] “Minutes of the International Color Consortium,” July 15, 1996 meeting.
[2] Newman, T., “Improved Color for the World Wide Web: A Case Study in Color Management for Distributed Digital Color,” *TAGA Proceedings*, pp 772, 1995.
[3] Johnson, T., *Colour management in graphic arts and publishing*, Pira International, 1996.
[4] Hunt, R.W.G., and M. R. Luo, “The Structure of the CIE 1997 Color Appearance Model (CIECAM97)”, unpublished?
[5] ICC Profile Format Specification, Version 3.3, 1996.
[6] Kasson, J. M., “Tetrahedral interpolation algorithm accuracy,” *SPIE volume 2170*, p. 24, 1994. and Kang, H. R. “Comparison of Three-Diemensional Interpolation Techniques by Simulation, *SPIE Volume 2414*, p 104, 1993.
[7] Dispoto, G. and M. Stokes, “Limitations in Communicating Color Appearance with the ICC Profile Format,” *Proc. IS&T/SID 1995 Color Imaging Conference*, p. 155, 1995.
8 Setchell, J. S., Jr. And E. J. Giorgianni, “Colour Encoding for Colour Management,**”** *Proc. IS&T/SID 1995 Color Imaging Conference*, p 114, 1995.
[9] Dispoto and Stokes, op. Cit. , p 157.
[10] Poe, R. F. “Aesthetic Considerations in Tone and Color Management,*” Proc. IS&T/SID 1995 Color Imaging Conference*, p. 166, 1995.
[11] MacDonald, L., “Colour Management and Display Calibration,” *Proc. CIE Expert Symposium ’96 Colour Standards for Image Technology*, p. 63, 1996. See also Johnson, op. Cit.