Color Image Quality in Digital Cameras

Paul M. Hubel
Hewlett-Packard Laboratories
Palo Alto, California, USA

Abstract

This paper addresses the challenge we face when attempting to judge the quality of color reproduction obtained from digital still cameras. We will consider possible methods and metrics and show how traditional colorimetric analysis (as used in scanning technology) has limited use as a tool for judging rendered digital camera imagery. We will talk about metrics that can be used to rate colorimetric accuracy of scene analysis of a digital imaging device and the constrained situations in which they are useful.

Digital still camera can be classified into two types: those that allow access to unrendered un-color-corrected data, and those that do not. The first type, most of which are professional cameras, allow someone testing the camera to measure the linearity and the spectral sensitivity of the camera system and determine metrics that indicate how well the camera can see color. This type of analysis can also be applied directly to the imager (such as a charged-coupled device) with or without IR filters. Although this analysis does not consider the resultant rendered image, it is the only reliable method yet proposed to compare colorimetric capabilities of digital cameras. The second camera type (and the first type after rendering is applied to the image data) requires a subjective analysis of a range of scenes under extremely different circumstances.

A discussion and several examples of what is involved in white-point balancing of digital images justifies our claim that color metrics are unsuited for judging rendered color quality in digital cameras. We describe how color appearance models fall short in their determination of the adapted white point of a scene and are thus unreliable as metrics for color image quality in digital cameras. The color transformations and the tonal rendering that are applied to camera images for display are also described briefly in terms of their effect on color image quality.

We present a recommendation for types of scenes that could give efficient testing of color quality when used in a subjective analysis.

Introduction

In digital camera systems we have a unique opportunity to explore a vast number of theories and techniques for color image reproduction. By having control of the color and intensity at every point in a scene (within the constraints of the imaging devices) we can explore the wonders of the human visual system and our attempts to understand color.

Conventional photographic systems have evolved over one hundred years to achieve an amazing ability to reproduce color scenes when used in the hands of a skilled professional (these systems can also achieve acceptable results when used in some automated systems). The complex chemical system combined with spectral sensitivity and dye selection result in non-linearity, much of which can only be explored by chemical manipulations (complicated by reciprocity, inter-image effects, cross-coupling, etc.). Digital image capture and processing is an inherently linear system which can be well controlled and can be perhaps too easily manipulated.

Although digital imaging systems bring us new tools to explore color vision, we are still at the early stages of our understanding. This incomplete understanding renders us incapable of devising a physical metric for rendered color image quality in digital cameras and forces us to continue to rely on subjective psychophysical analysis (as has been used in the photographic industry for decades).

Taking a photograph with a digital camera can be broken down into three stages: image capture, image processing, and image output rendering. These stages are analogous to latent image formation, development, and printing in chemical based systems.

Some physical metrics have been recently proposed which relate to the quality of the image capture stage[1]. These metrics attempt to describe how well the camera could analyze the scene colorimetry. I shall describe the limited cases where this kind of metric us useful and the assumptions that must be considered when using this type of metric relative to image quality. In the case of image processing and output rendering, physical metrics have been proposed[2], but I will give examples which show how they are not reliable metrics for color image quality.

Image Capture

The action of taking a photograph with a digital camera produces electrical signals from a sensor (such as a charged-coupled device) which have a well defined physical relationship to the light from the scene incident on the optics and the sensor. Through careful characterization of the system, one can describe this relationship in terms of the opto-electronic conversion function (OECF)[2], the lens MTF[3], and the complete spectral sensitivity[1]. The OECF describes the relationship between the radiant flux at a given wavelength, and the sensor's output signal. From the standard deviation of this measurement we can determine the dynamic range in which the sensor operates. Using the OECF characterization we can then measure the spectral sensitivity of the device which relates the device signal as a function of the wavelength of the incident light. At this stage it is vital to measure all wavelength components to which the sensor gives a measurable signal (not restricted to visible radiation). We give an example below that shows why this is so important.

The above measurements depend on access to raw signals off the camera sensor electronics. Any manipulations of the signals that add noise or uncertainty to the data preclude any further meaningful physical analysis of the color analysis capabilities of the camera. If any manipulation is made, this must be completely reversible without loss of accuracy. Several professional digital camera (such as the Kodak 200, 420, & 460, MegaVision digital backs, Leaf Lumina, and others) do give clean enough data for this kind of analysis[4]. The overwhelming majority of digital cameras on the market, however, only output data after irreversible color manipulation (this includes professional cameras such as the new Kodak 520 & 560, as well as all consumer digital cameras).

Assuming accurate spectral sensitivities of the camera can be obtained, we are now faced with the difficult task of relating these functions to a set of color matching functions. If the camera sensitivities are a linear transformation of the CIE color matching functions (XYZ) then the camera can be considered a colorimeter and will give excellent colorimetric analysis of the scene. This ideal case is not practical (requires expensive filter fabrication) nor desirable (the filters will have lower signal-to-noise than other combinations[5]).

A fundamental assumption of colorimetry is that there are a set of functions which are linearly related to the human visual system's spectral sensitivities. The functions are derived from observers matching colors under constrained conditions and have been shown to be approximately a linear transform from measured microspectrophotometry of cone pigments (two cases are compared in reference[6]). Even these functions, upon which the entire field of colorimetry is based, have been shown to be unreliable and incorrect in some circumstances[7]. Nevertheless, if we assume the color matching functions are an accurate enough representation our color system we would like to compare our camera's sensitivities to them.

The ISO committee TC 42. has written a draft standard[1] which proposes such a metric: the ISO Camera Metamerism Index. This metric is computed in the following manner:

1. Camera spectral sensitivities determined.
2. Sensitivity functions fit to a set of color matching functions.
3. Error metric calculated corresponding to the deviation of these functions from the color matching functions used in step 2.

There are three assumptions that must be considered in this proposal: the color matching functions chosen (ISO RGB) are appropriate; the transformation method to these functions is appropriate; and the error metric is appropriate. When the camera sensitivities are a linear transform of the color matching functions then the metamerism index will be 0 (there is no error in the camera's colorimetric analysis of the scene) as the camera spectra deviates from these ideal curves, the color analysis obtained by the camera deteriorates. The color matching functions chosen for the metric (ISO RGB) have the advantage that they have primaries based on the most common display and communication color spaces (such as sRGB[8] and compared with others in reference[9]). Thus the transformation matrix used for this metric will be similar or the same as will be used to render the data for display. It could be possible, however, that one set of sensitivities could produce better results using a different transformation technique than the one proposed in the standard. The error metric of the curve fit is the most difficult assumption to relate to general image quality – an example is shown below.

Although there are flaws in the proposed metamerism index, the index is useful for relating and comparing imaging sensors; there does seem to be a correspondence between the index and the camera's ability to estimate the colorimetry of a scene.

It should be stressed that this metric only judges how well the camera sees color and says nothing about what happens to the data after capture.

A camera with a favorable metamerism index has the potential of producing an image with high color image quality, but by no means assures this. A camera with a poor metamism index will be unlikely to produce high image quality over a large range of scenes.

It is impossible to relate the many complex interactions between spectra and color with a single quantity. Spectral mismatch between the transformed sensors and the ideal functions will have very different effects depending on their wavelength, and these effects will depend on the scene. Applying a single error metric to this spectral interaction is dangerous. For example, even slight discrepancies between the transformed camera spectra and the ideal curves in the near IR region can cause dramatic color shifts when IR reflecting surfaces are viewed under tungsten illumination. When the same object is illuminated by fluorescent illumination, however, the colors are accurate.

IR color shift example:

The images shown in supplemental figure 1 & 2 on the CD-ROM were taken under tungsten illumination. These two images were taken with two types of IR filtration, giving a slight change in the IR signal seen by the camera sensor. The image in supplemental figure 3 on the

CD-ROM has the same IR filtration as image 1, but the illumination was changed to warm fluorescent. Color measurements show that the color errors on the ColorChecker in all three scenes are small. The dress, however, appears purple in the first image, purplish-blue in the second image, and navy blue in the third image. This example shows that perhaps more weight should be given to the metric for errors in the IR region. The weighting of the error metric for different spectral regions based on this kind of effect would be obviously nontrivial. The ISO does give a minimum qualification of a camera's IR response when determining the OECF and Metamerism Index, but this requirement is not strict enough to compensate for these large color shifts. The device used for the first image fails the ISO IR qualification, but the device used for the second and third images (the first device using an additional IR filter) does meet the ISO requirement yet still gives large color shifts due to IR light.

Image Processing and Rendering

All digital cameras (or accompanying software) must transform the raw camera data into an image data file that is fit for display by some output media (a crt, projector, print, etc.). This type of operation is extremely complex having many important components to consider. The three main components are:

1. White balance / color balance
2. Transformation of camera data into a color space.
3. Rendering for output media.

We shall describe how each of these relates to color image quality and why subjective analysis is necessary for determining the effectiveness of these algorithms.

When considering image processing and output rendering, there are large differences between cameras and other color imaging devices (such as scanners or color copiers). The light source in a scanner is known, the color transformations can be optimized for the input media, and the dynamic range and gamut of the input media are small and similar to that of the desired output rendering. All of these factors are similar to the assumptions of colorimetry and color appearance, and thus metrics based on these fields can be successfully correlated to color image quality[10]. In the case of a digital camera viewing a scene, however, these assumptions fail and we must rely on subjective analysis.

White Balance / Color Balance

Color perception relies on a phenomena known as 'color constancy' or 'chromatic adaptation' which allows the visual system to adapt to widely different viewing environments. A white surface appears white under a range of illumination conditions: from reddish tungsten light, to bluish daylight, to greenish fluorescent light. A digital camera before white balance will record the color of the illumination as a strong color cast over the scene. Conventional photography corrects for this by film that is tailored to the illumination condition or by the use of special filters. Video cameras and digital cameras can compensate for illumination color by analog or digital gain adjustments or by digital calculations.

In order for a digital camera image to have high or even acceptable image quality the white balance and neutral scale must be reproduced as perceived by an observer viewing the original scene. When an observer looks at a scene, the colors have a relationship anchored by the observer's *adapted white point*. The term *adapted white point* refers to the color which an observer deems to be white in a given circumstance. In color appearance models (the extension of colorimetry that includes chromatic adaptation and other environmental conditions[11]) and indeed, in most situations, the *adapted white point* of the observer is probably the color of the scene illumination. Color appearance models employ some kind of chromatic adaptation transform which factors out this illumination color. The actual white value that is used by these color appearance models has been recently referred to as the *adopted white-point*[12]. The distinction here is that the *adopted white point* is what the model is using and the *adapted white-point* is what the visual system is using. Some of these models[13] will even account for subtle color shifts beyond chromatic adaptation – such as tungsten illumination giving warmer tones than daylight illumination. These models rely on many careful measurements of the scene and viewing conditions including a spectrophotometric measurement of the scene illumination and many other artificial constraints (single scene illumination, equal color gamuts and dynamic range, etc).

The white balance algorithm in a digital camera does a similar type of chromatic adaptation transform using some estimate of the scene illumination color and using this estimate as the adopted white point for the color model used for the camera image processing. The camera's white point estimation algorithm must obtain the scene illumination color from the image data, a series of previous images, or from an additional sensor on the camera. Some cameras also allow a calibration step which requires the photographer to place a white or gray card in a preceding exposure.

One might ask the question why we cannot derive metrics using a color appearance model to describe camera color image quality? Under the constraints of the intended use of color appearance models this is practical and could lead to interesting results. Under these constraints, the *adopted white point* measured, calculated, and used by the color appearance model will be a good estimate of an observer's *adapted white point* under the same conditions. I will argue with an example, however, that this is far from the use model for digital cameras and can lead to misleading and even contradictory results. The example explores the complexities of the adapted and adopted white points.

Color Appearance for Camera Image Quality Example:

Consider the case where a digital camera is used to photograph a MacBeth ColorChecker Chart under three different viewing conditions. The resultant images from the camera can be compared with the original scene using a metric based on a color appearance model such as CIECAM97s[12] – such as the Reproduction Index proposed by Pointer.[14]

In the first scene (supplemental figure 4 on the CD-ROM) the ColorChecker and surroundings are illuminated by tungsten light. The camera produces an image balanced for tungsten light which compares well with color appearance percepts which are calculated using a measurement of the tungsten illumination as the adopted white point. Here the metric will correlate well with high color image quality.

In the second scene (supplemental figure 5 on the CD-ROM) the ColorChecker is illuminated by a D50 daylight simulator. Again there is good correlation between the color appearance metric and high color image quality.

The third and most interesting case is when a ColorChecker is illuminated by sunlight late in the day when the sun is just setting (supplemental figure 6 on the CD-ROM). This is the familiar light that photographers strive for when objects impart a warm glow. If the digital camera were to balance the ColorChecker for the color of the illumination – and balance the ColorChecker much as it did for the first two cases, then this would again give a high ranking in our color appearance metric. The color image quality, however, will be poor and unacceptable (see supplemental figure 7 on the CD-ROM). The warm glow of the sunset lit ColorChecker will be gone – the image will have a cold stark white balanced chart and cold blue surroundings.

In the third example, the adopted white point used to calculate the color appearance model metric and used to balance the digital camera image (the color of the illumination) did not match the adapted white point of an observer. Since these kinds of scenes are very common and familiar to everyone, a person viewing the digital camera output will recognize the image as having poor color image quality without having to be present in the original scene. This example shows how this metric is unacceptable and misleading. The relationship between a scene, an observer's *adapted white point,* and a model's *adopted white point* is simply not well understood; we have yet to find any way to measure this *adapted white point* other than a subjective psychophysical experiment on observers (and even this will simply give us an acceptable *adopted white point* and not necessarily an *adapted white point*).

The physiology of color constancy is far from understood and thus our evaluation of our models must be accomplished by subjective testing. The image shown in supplemental figure 7 on the CD-ROM has been balanced by a new algorithm that produces a version of the image in supplemental figure 6 on the CD-ROM resulting in much higher color image quality. This new algorithm is based on our recent work on color constancy[6, 15] and will be described further in future publications. Our conclusion that this image has higher color image quality that that shown in CD figure 6, comes solely from subjective evaluation by both observers present at the scene and those who are viewing the image for the first time (such as the audience at this paper presentation). We have no physical metric to predict this behavior, and metrics based on colorimetry and/or color appearance will be contradictory to this result. [The image in CD figure 6 has smaller colorimetric error than the image in CD figure 7, but clearly the image in CD figure 7 has higher image quality].

Transformation of Camera Data Into a Color Space

In a recent publication[16] we described several methods that could be used to transform camera data into a color space. The task addressed was to compute a 3x3 transformation matrix that would take camera RGB data to a standard RGB color space (a similar task as used in the proposed metamerism index[17]). In this study, we compared four methods for determining the linear transform matrix. The first three methods relied on measured data from a set of reference color surfaces (the MacBeth ColorChecker Chart) using a least-squares fit, a white-point preserving least squares fit, and a weighted white-point preserving least squares fit. The forth method was based on a matrix computed from just the camera spectral sensitivities and did not rely on any reference surfaces nor scene statistics (the so-called maximum ignorance method).

One important conclusion from this study was that neither the CIELAB delta E nor the CMC deltaE metrics are reliable predictors of camera color image quality. In the rendered samples used, reproductions with high deltaE errors were preferred over those with significantly lower deltaE errors (where the only difference was the linear transformation matrix). Within the constraints of this study there was no correlation between color image quality and these colorimetric metrics.

Rendering of Output

How a digital camera maps the dynamic range recorded on the image sensor to the output image file will have profound impact on the color image quality. Selection of the luminance scale of the adopted white-point[18] and the black level determined in the scene often results in extreme clipping of highlight and/or shadow regions. In conventional photographic systems a soft roll-off is employed, but the shape and position of this curve must be tailored to each image for dependable results.

In the case of clipping of highlight regions, digital cameras often wash out large important parts of images – such as the sky. Part of this effect is caused by frequent over exposure which leads to excessive saturation. Unlike conventional photography, this type of over exposure cannot be corrected after exposure – once the pixels saturate, the image information is lost.

Suggested Test Scenes

The following is a list of some suggested scenes that could be used as part of a subjective evaluation of digital cameras. These should be a part of a much larger selection of scenes that fit the use model for the given camera. Additional scenes that include objects that test the camera's resolution, sharpness, and spatial image processing (such as textures, bicycle wheels, etc.) should be used for testing image quality that is not directly related to color. The following scenes represent some corner case examples that can be used to test certain aspects of the camera's color image quality performance. Where applicable, pictures should be taken with available light, and also with fill and full flash.

1. Typical indoor scenes with tungsten illumination, no flash, and containing materials that are known to reflect IR (many fabrics are designed to be cool).
2. Scenes used in 1 using fluorescent illumination.
3. Known reference colors (Munsell, Pantone, etc.).
4. Outdoor scenes at all phases of daylight.
5. Sunsets and sunset lit scenes.
6. Scenes which are comprised of a limited number of colors (where the background has only one saturated color, for example a portrait where the background is blue sea and blue sky).
7. Hi range scenes – Outdoor scenes with important shadow detail. Indoor/outdoor scenes (an indoor scene with a view out a window).
8. Low range scenes – view out an airplane window.
9. Hi key scenes – white dress, white background.
10. Low key scenes – black tux on black background.
11. Snow scenes.
12. Beach scenes.
13. Night scenes, with and without moon light.
14. Candle lit scenes.
15. Scenes with multiple light sources.
16. Scenes with foliage and saturated flowers.
17. Scenes with small specular highlights.

Conclusions

Although digital cameras have the potential to achieve extremely high color image quality (higher than silver halide camera systems), the flexibility and ease of manipulation have resulted in generally inferior results thus far. Silver halide systems have evolved using decades of subjective psychophysical evaluation. Digital cameras, for the most part, have come from video and electronics based systems which have given less attention to color image quality

Digital cameras generally give cold, flat, desaturated images with variable preservation of highlights and shadows. In some cases, the algorithms used to adjust the images to the adapted white point, produce worse results than if they were not used at all. In the example used in this paper, a sunset lit scene was corrected for the illumination resulting in a cold reproduction with poor color image quality. If photographic film were used for the same scene (and not balanced for the illumination), the reproduction would have high color image quality.

References

1. ISO 17321 (WD3), *Graphic Technology and Photography - Colour characterisation of digital still cameras (DSCs) using colour targets and spectral illumination.*
2. ISO 14524 (FDIS), *Photography - Electronic still picture cameras - Methods for measuring opto-electronic conversion functions (OECFs).*
3. ISO 12233 (FDIS), *Photography - Electronic still picture cameras - Resolution measurements.*

note: references 1-3 can be obtained on the internet from: http://www.pima.net/standards/iso/tc42/wg18.htm

4. G. D. Finlayson, S. Hordley, and P. M. Hubel, "Recovering device sensitivities with quadratic programming", *Proceedings, 6th IS&T/SID Color Imaging Conference*, Scottsdale, AZ, 90-95, 1998.
5. A.K. Juenger, "Color sensitivity selection for electronic still cameras based on noise considerations in photographic speed maximization" *Proceedings, 1st IS&T PICS Conference*, Portland, OR, 79-83, 1998
6. P.M. Hubel, J. Holm, & G.D. Finlayson: "Illuminant Estimation and Color Correction," *Proceedings, Colour Imaging in Multimedia - CIM 98*, 97-105, Derby, England, 1998. Expanded into a book chapter of the same name for Colour Imaging in Multimedia, Lindsay MacDonald, editor, Wiley, 1999.
7. W.A. Thornton, "Spectral sensitivities of the normal human visual system, color-matching functions and their principles, and how and why the two sets should coincide," *Color Res. Appl.*, In Press.
8. IEC 61966-2-1 (CDV), *Colour measurement and management in multimedia systems and equipment - Part 2-1: Default RGB colour space – sRGB*
9. M.H. Brill, G.D. Finlayson, P.M. Hubel, and W.A. Thornton, "Prime colors and color imaging," *Proceedings, 6th IS&T/SID Color Imaging Conference*, Scottsdale, AZ, 33-42, 1998.
10. R.G. Gann, Desktop Scanners: Image Quality Evaluation, Prentice Hall, Upper Saddle River, NJ, 1998.
11. M.D. Fairchild, Color Appearance Models, Addison-Wesley, Reading, MA, 1998.
12. R.W.G. Hunt & M.R. Luo, "The Structure of the CIE 1997 Colour Appearance Model (CIECAM97)," manuscript handed out at the AIC Conference, 1997.
13. R. W. G. Hunt, "Revised color appearance model for related and unrelated colours", *Color Res. Appl.* **16**, 146-165, 1991.
14. M.R. Pointer, "Measuring Colour Reproduction", *J. Photogr. Sci.*, **34**, 81, 1986.
15. G. D. Finlayson, P. M. Hubel, and S. Hordley, "Colour by Correlation", *Proceedings, 5th IS&T/SID Color Imaging Conference*, Scottsdale, AZ, 6-11, 1997.
16. P.M. Hubel, J. Holm, G.D. Finlayson, & M.S. Drew: "Matrix Calculations for Digital Photography," *Proceedings, 5th IS&T/SID Color Imaging Conference,* 105-111, Scottsdale, AZ, 1997.
17. G.D. Finlayson & M.S. Drew, "The Maximum Ignorance Assumption with Positivity," *Proceedings, 4th IS&T/SID Color Imaging Conference*, 202-205, 1996.
18. J. Holm, "Adjusting for the Scene Adopted White," *Proceedings, 2nd IS&T PICS Conference,* (this proceedings), Savannah, GA, 1999.