CFS-243

Maintaining Color Integrity in Digital Photography

David Dunthorn
C F Systems
February 29, 2004

Maintaining Color Integrity in Digital Photography

CFS-243

February 29, 2004

This document is part of a series of three documents from early 2004 investigating some aspects of digital photography. The companion documents are CFS-242, *Film Gamma Versus Video Gamma* and CFS-244, *Negative to Positive.* Refer to those documents, available on our web site, for the further elaboration of the symbols, equations, and concepts used here. See also our web page on Dunthorn Calibration (http://www.c-f-systems.com/DunthornCalibration.html) for a method of testing your photographic image input system for color integrity and what to do to correct any problems which you find.

This document is viewable only. We believe this will be adequate for people who do not intend to study it. Please contact us through our web site if you need a printable version. We are aware that the no-print can be defeated, but again we ask that you contact us instead. We really need to know if and how people are finding these documents useful, and this seems one of the few ways we have to encourage feedback.

David Dunthorn
www.c-f-systems.com

The problem we describe here is at the very core of digital photography. The effects of it are visible in many of the digital images we currently see all about us and are the unintentional result of widely accepted methods of dealing with digital images. This may mean that the reader will have to re-evaluate some basic concepts, as we had to when we launched this investigation. It is necessary to have a clear understanding of the basics to see how the problem has come about and why it has persisted. For that reason some of the following material is basic and we suggest following along even when the material being covered is familiar.

Color Integrity in Digital Imaging

By *color integrity* we mean that when brought into color balance, a photographic image reproduces the colors in an original scene as best it can within the constraints of the primary color system being used. We must make clear at the start that color integrity is not an absolutely necessary requirement for every successful photographic image. However, it should be the starting point. Departures from color integrity should be a matter of *choice*, and not the unknowing result of working procedures.

The Key Role of Intensity Ratios

To understand color integrity in photographic images it is necessary to understand the key role light intensity ratios play in vision, imaging, and the physics of light. We will first deal with black and white images so we can better understand the foundation of color images. Light intensity, I, is the amount of light playing on a specific area, typically measured in nits (candelas/m^2). As light plays upon a scene, objects in the scene reflect some of the light, and when a camera or eye resolves the light reflected from various point in a scene, an image is formed. The amount of light reflected from a uniform, small object , I_o, is a specific fraction (percentage) of the light shining on the object, I_{illum}. That is, $I_o/I_{illum} = \alpha_o$, where α_o is a constant specific to the object "o." As the light level I_{illum} increases so does I_o and so does the amount of light reflected from each object in a scene, all in concert with one another. If the overall illumination decreases by 20%, the light reflected by object "o" also reduces by 20%, *and so does the light from every other object in the scene*. For those who like to "nit" pick, of course there are unusual exceptions to this, but the rule is generally true. We leave exceptions for the future.

The light coming from a scene operates according to *intensity ratios.*

It can be easily seen that this also means that the ratio between the intensities of any two arbitrary points in a scene or image will remain constant regardless of the absolute intensity of either point. Consider two points in a scene with intensity I_{o1} and I_{o2} operating under I_{illum}. Change the illumination of the scene to I'_{illum}. Because $(I'_{o1}/I'_{illum}) = (I_{o1}/I_{illum})$ and $(I'_{o2}/I'_{illum}) = (I_{o2}/I_{illum})$, the two points now have intensities of $I'_{o1} = I'_{illum}\,(I_{o1}/I_{illum})$ and $I'_{o2} = I'_{illum}\,(I_{o2}/I_{illum})$. Thus

$$\frac{I'_{o1}}{I'_{o2}} = \frac{I'_{illum}\left(\dfrac{I_{o1}}{I_{illum}}\right)}{I'_{illum}\left(\dfrac{I_{o2}}{I_{illum}}\right)} = \frac{I_{o1}}{I_{o2}}\,.$$

The ratio of the light intensities coming from any two objects in a scene stays the same even if the overall illumination level changes. That is, all the intensities float together.

The iris of the eye and the lens opening of a camera each attenuate the amount of light passing from a scene to the light sensitive elements of the eye or camera. Again, the amount of light passing through the lens opening or iris, I_i, will be a specific fraction of the light reflected to the lens from the object, I_o. When we pass the light coming from a scene, I_o, through the iris of a camera or an eye so that I_i emerges, we find that $I_i / I_o = \alpha_i$, and again the effect is a *ratio* applied to all the light intensities. We note that this is principally the effect of changing the fraction of the lens area that is transparent versus the fraction that is opaque.

Changing the iris or lens opening operates as an intensity ratio and has *exactly* the same effect as changing the overall illumination level, reducing or increasing *all* intensities in a scene by the same percentage.

We can also place a transparent but partially darkened medium somewhere in the path of the light to attenuate it. Examples of such media are photographic filters (it is easiest to think of neutral density filters as we still are considering black and white here) and even the images on transparencies. Both filters and the points making up an image on a transparency are characterized by their photographic *density*. The density, d, of either a filter or a transparency is defined in terms of the light incident on the filter, I_{incident}, and the light transmitted through the filter, $I_{\text{transmitted}}$, so that $d = \log_{10}\left(\frac{I_{incident}}{I_{transmitted}}\right)$. A filter is specified according to its constant density, so the log term is constant, meaning that the ratio of $\frac{I_{incident}}{I_{transmitted}}$ is also constant.

A filter also operates to produce a constant ratio of illuminations and so has *precisely* the same effect as changing the overall illumination level or changing the lens opening. All three produce effects in terms of constant ratios and so all three are interchangeable in terms of effect.

In the normal darkening and lightening of an image, the intensity ratios representing points in an image all are lowered or raised in concert; that is, by the same percentage, independent of the cause.

The response of the human eye is in accord with all this - how could it be otherwise? The CIE (*Commission Internationale de l'Éclairage*) has been a primary source of color standards for most of the past century. When they set up CIELUV and CIELAB (called Lab in Photoshop) in 1976, they defined a "lightness" function based on data and studies from much earlier in that century:

$$L^* = \begin{cases} 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16; & 0.008856 \geq \frac{Y}{Y_n} \\ 903.3\left(\frac{Y}{Y_n}\right); & 0.008856 < \frac{Y}{Y_n} \end{cases}$$

$$\frac{Y}{Y_n} = \begin{cases} \left[(L^*+16)/116\right]^3; & L^* \geq 8 \\ L^*/903.3; & L^* < 8 \end{cases}$$

Here L^* is the defined dimensionless quantity "lightness," and corresponds closely to the human eye's response to light in images, at least under controlled conditions. Y is a measure of light intensity at some image point, analogous to I above, and is spectrally weighted to correspond to the response of the human eye. The eye is less sensitive to the blue and red toward the ends of the spectrum than it is to green at the center of the spectrum, so an increase in green intensity will count more in Y than a similar increase in either red or blue. The spectral weighting is done in accord with measurements of the light response of the eye, from zero weighting at the extremes of the spectrum extending smoothly to a maximum in the green area of the spectrum. Y_n is the Y intensity of the dimmest bright white that the eye sees as pure white and is generally thought of as the brightest area in a scene, so that (Y/Y_n) is the fraction of the maximum distinguishable light intensity that is reflected by or projected through a specific point in a scene or image. L^* has been arranged to range from 0 for perceptually totally black to 100 for perceptually totally white, and the 100 steps of lightness are arranged so that each successive gray (50, 51, 52,…) is at the threshold of distinguishability (approximately) from the preceding under controlled viewing conditions.

The important physical principle we learn from this is that what the eye sees tends to respond to the intensity ratios of a scene, Y/Y_n, and that the perceived "lightness" intensities scale down from the brightest area of a scene or image (Y_n). As the illumination source for an image brightens or darkens, the entire scene or image tends to follow, each point reflecting the same fraction (same percentage) of the brightest point, regardless of how bright or dim that point is.

Rather than in terms of absolute intensities, the eye actually sees images in terms of *ratios* of intensities. In this way the response of the eye depends upon the same ratio effect that governs the way images naturally change in response to changing illumination levels and the same ratio effect that governs the action of photographic filters.

Storage of Digital Images

Images are most naturally expressed as collections of intensity ratios. Digital images can be stored as collections of pixels representing discrete points within the image and which contain an intensity ratio (or several ratios for color). The denominator of all ratios is the same and typically represents some governing maximum intensity in the image, so that the ratio is also normalized: $0 \leq J \equiv I/I_{max} \leq 1$. We will often use the shorthand J for such ratios. The natural lightening or darkening of an image raises or lowers all intensity ratios in pixels in concert; that is, each pixel is changed by the same percentage. Thus, such a stored image can be applied to any actual maximum intensity without internal change. An important side-note is that although J is conceptually fractional, it is often stored as the numerator of a ratio in which the denominator is understood. For example, in 8-bit notation, the numerator is an integer $0 \leq n \leq 255$ and the denominator is understood to be 255. Often this concept is "simplified" by saying that the pixels take on values of 0 to 255. It is worth remembering that is not strictly correct; particularly since the "simplified" version is not normalized.

An important complication arises in common current practice in digital imaging where pixels are actually stored as "gamma-adjusted" intensity ratios. That is, in place of J, the quantity J^{γ}, a mathematical proxy of the actual intensity ratio, is stored. We will deal with the importance of this later, but for the moment we will continue to use intensity ratios directly as pixel values. Since the gamma adjustment is reversible (ignoring integer truncation problems), the actual intensity ratios are accessible from such gamma-adjusted images.

And Now Color

As nearly universally implemented, color images are sensed as three individual images called R = red, G = green, and B = blue and that will certainly be sufficient for our purposes. Each of these images is obtained as a spectrally weighted intensity. Just as the CIE *Y* above is spectrally weighted to match the overall sensitivity of the human eye, R is spectrally weighted to strongly favor the red end of the spectrum, B is weighted for the blue end, and G is weighted for the central part of the spectrum, which contains green. It has been found that using the R, G, B intensities sensed in that manner to generate red, green, and blue light having matching properly weighted intensities will recreate the color of the sensed point, nearly perfectly for normal colors and usually acceptably even for very saturated colors.

An Aside Concerning Primary Colors

The way in which primary colors work is usually glossed over and even treated as obvious. It is not at all obvious by any stretch of the imagination. The spectral sensitivities of the red, green, and blue sensors are typically quite wide and overlapping, whether electronic or film, and are very different from the spectral sensitivities of the three different color sensors in the eye, which overlap even more. One would think that the spectral integration conflicts generally known as metamerism would be a constant and serious problem. Yet a resulting realistic-looking image may be produced from red, green, and blue lights each having very narrow spectra, as in a CRT, or very broad, as in dye film images. Rather than intuition, the use and practice of primary colors relies on a very broad base of experience and many years of experiments. It has taken a lot of work by many people to know that this works and works quite well.

While the results of the red, green, and blue sensors of an electronic camera can be recorded directly to a computer file, the "sensors" in a film image must be recorded in subtractive primary colors, magenta (red-blue), cyan (blue-green), and yellow (red-green). A dot resulting from the red-sensitive film layer eventually will be expressed as magenta and yellow in metrically equal quantities. However, the red sensor in a scanner will also respond to light passed through magenta and through yellow metrically equally, so the scanned R in the R, G, B image will be a reasonably accurate reading of the original red-sensitive film layer.

Returning to Photographic Imaging in Color

Examine what happens when we form a color image. For each color, using red as an example, the amount of red light reflected from a uniform, small object , I_{Ro}, is a specific fraction (percentage) of the red light shining on the object, I_{Rillum}. That is, $I_{Ro}/I_{Rillum} = \alpha_{Ro}$, where α_{Ro} is a constant specific to the object "o." The behavior is thus exactly the same as with a black and white image. When the red light level I_{Rillum} increases so does I_{Ro} and so does the amount of red light reflected from each object in a scene, all in concert with one another; all red intensities increase by the same percentage. It is easy to see that when the general illumination level changes - with no change in its spectral quality - so does the red intensity level I_{Rillum}, and also I_{Gillum} and I_{Billum} so that the entire color image changes in concert with the light level. After adaptation the image will look the same to the eye over a wide range of illumination.

In digital image terms, a black and white image is stored so that each pixel is the ratio $J \equiv I / I_{max}$, where I is the intensity at the pixel point, I_{max} is a maximum intensity prevailing for the entire image and J is a convenient shorthand notation. We know from the above that this ratio form will work even if applied to an I_{max} that is markedly different than the original scene, so that a paper print of a bright ocean scene looks natural when displayed in a comparatively dim indoor gallery.

A color image has each of its individual small areas (pixels) specified in terms of three intensity ratios, $J_R \equiv I_R/I_{Rmax}$, $J_G \equiv I_G/I_{Gmax}$, and $J_B \equiv I_B/I_{Bmax}$. These three intensity ratios are treated as though they are three independent black and white images - note that they even have separate maximum intensity references. Overall the maximum intensities are tied to one another according to some representative maximum intensity in the image, but the *maximum intensities selected to represent each of the three colors must properly produce the color balance for the image*. This is the way in which color balance works.

A combination of eye and mind automatically adjusts to both the prevailing illumination and to context to produce color balance in vision. Photographic film requires a specific illumination quality for proper color balance and filters are used, both when the image is taken and when it is printed, to achieve this. This use of filters is exactly the equivalent of changing the intensities of each of the R, G, and B components of the lighting for the original scene so they are the correct percentage relative to one another; that is, the specific quality required by the film. Digital cameras often are able to make color quality measurements of the scene to estimate the balance between the R, G, and B components, adjusting the sensitivities of the R, G, and B sensors to match the measurements. Normally, this again should operate just as filters do for film and can be quite effective depending upon how the estimate is drawn and the nature of the scene.

Color Balance

Color balance of a scene or image is thus achieved by adjusting the intensity of illumination produced in different parts of the spectrum so that it matches a reference spectrum in an integral sense. The principles of primary colors allow us to separate out three parts of the spectrum, Red, Green, and Blue, and to adjust the spectral intensities in these three parts to match the reference spectrum. We have shown that the use of color correction filters and the changing of illumination intensities are equivalent in action, and that changing the sensitivity of properly designed and implemented sensors is also equivalent in action. Thus color balance may be achieved equally well using any of these means. In each case the result is that the pixel intensity values for any one primary color, say Red, are changed by the same fraction or percentage throughout the image. Each of the other two primaries either remains unchanged or changed throughout the image in a similar way by a (different) fraction or percentage.

Mathematically, consider two pixels in the same image with intensity ratios: (J_{R1}, J_{G1}, J_{B1}) and (J_{R2}, J_{G2}, J_{B2}). We will adjust the color balance of the image by applying a fractional change β_R to the red channel and a fractional change β_G to the green channel. Thus the color balanced pixels will then be ($\beta_R J_{R1}$, $\beta_G J_{G1}$, J_{B1}) and ($\beta_R J_{R2}$, $\beta_G J_{G2}$, J_{B2}). Now take the ratios of the pixel values in each channel: $\frac{\beta_R J_{R2}}{\beta_R J_{R1}} = \frac{J_{R2}}{J_{R1}}$, $\frac{\beta_G J_{G2}}{\beta_G J_{G1}} = \frac{J_{G2}}{J_{G1}}$, and $\frac{J_{B2}}{J_{B1}}$. Note that these ratios remain the same as they were before the color balancing operation. This leads us to:

The Cardinal Rule of Color Balance

Given any two arbitrarily chosen pixels in an image, a correctly performed color balance will leave unchanged the ratio of intensity values for each color channel. Expressed in another way, any correctly performing color balance operation will multiply all the R channel intensities in an image by the same factor β_R, all the G channel intensities in an image by the same factor β_G, and all the B channel intensities in an image by the same factor β_B. β_R, β_G, and β_G will typically be different values. This is a key diagnostic for proper color balancing, and *any operation which does not meet these requirements introduces a loss of color integrity*.

Gamma Matching and Gamma Adjustment

Before we can explore the adjustments commonly used in a program like Photoshop, it is necessary to understand the video gamma adjustment, where it came from and how it is used. The document *CFS-242, Film Gamma Versus Video Gamma,* defines the video gamma as:

$$\left(\frac{I_2}{I_{max}}\right) = \left(\frac{I_1}{I_{max}}\right)^{\gamma}$$

In the following we will use the notation $J \equiv (I / I_{max})$, the normalized intensity ratio and so in general $0 \leq J \leq 1$. Thus the gamma adjustment becomes simply $J_2 = J_1^{\gamma}$.

CFS-242 has a more thorough treatment of video gamma as it is used in programs such as Photoshop, where it appears as the Levels tool middle gray adjustment, among other places. We would be remiss if we did not at least mention that in its original context, the video gamma equation has a difference. It is typically introduced as $L = V^{\gamma}$ where L is normalized luminance of the video display and V is normalized voltage driving the display and in this form it is the same as the above equation. However, in actual use this is modified to $L = (V + \varepsilon)^{\gamma}$ where ε is called the black level voltage. We do not intend to minimize the importance of the black level, which is critical to correct CRT display operation, but in the computer treatment of digital images, the equivalent of black level adjustment is done differently and will be discussed later.

Gamma adjustments were originally intended for *gamma matching*. CRTs (and printers) behave non-linearly in response to a control voltage or signal, and the non-linearity has been found to follow the gamma form $L_V = V^{\gamma}$ reasonably well (less well for printers). If voltages are encoded directly as luminances from an original scene, $V = L_S$, then we will see on the CRT display $L_V = L_S^{\gamma}$, and unless $\gamma = 1$, the display image will not look like the original scene. So, practice has been to *gamma-encode* the voltages according to $V = L_S^{1/\gamma}$. Then we find that $L_V = (L_S^{1/\gamma})^{\gamma} = L_S$ so that the luminance of the display and the luminance of the scene match. In practice other factors come into play, but this general concept prevails. In a convention which dates back to the earliest days of television, video cameras are still designed to deliver gamma-encoded images, prepared in advance for delivery to a CRT display.

Color television uses three primary colors. Consider the three primary colors as each color goes through this same process. Each primary will display with the same luminance it had in the scene and thus for any pixel, the ratio of any pair of primary luminances will remain constant, as required for maintaining color integrity. Note that this is true even if the gammas are different for the three primaries, as long as each gamma/gamma-encode is properly matched.

This same gamma matching process has been used with film, in particular where color negative film is made to have γ_n, normally less than one to widen film latitude, and the gamma-matched print paper has a gamma of $1/\gamma_n$ to compensate. Again, the product of the gammas is 1 so the luminances match and so the ratio of any pair of primary luminances will remain constant, as required for maintaining color integrity. Again, it is possible to have different gammas for the different primaries as long as each is separately matched. See *CFS-242, Film Gamma Versus Video Gamma* for a further discussion of this general topic.

While gamma matching preserves color integrity when used properly, gamma adjustments which are not matched can lead to serious losses in color integrity.

Currently it is widespread - almost universal - practice to store digital images in gamma-adjusted form. Rather than being stored as directly as intensity ratios, J, a gamma-adjusted ratio $J^{1/\gamma}$ is stored in its place. γ is usually 2.2, gamma-matched for a CRT, or 1.8, gamma-matched for a long-past Macintosh printer. Initially this convention arose from convenience. Television cameras were a widely used source for digital images and as the signal from standard television cameras already is encoded as $J^{1/2.2}$, it would actually have to be *decoded* to be stored as intensity. There was the additional advantage that if left encoded, the image could be sent (via DACs) directly to a CRT display.

Then it was noticed that the standard 18% reflectance gray card ($J = 0.18$) encoded as $0.18^{1/2.2} = 0.46$. Since the 18% gray card perceptually is "middle gray," one might regard 46% as more appropriate for it than 18%. Taking this one step further, it was observed that the $J^{1/2.2}$ encoding was not very different than the L^* luminance encoding used in CIELAB, so that very approximately $L^* \approx 100J^{1/2.2}$. This "natural" coincidence sealed the deal, as it completely solved what could have been a serious problem in digital imaging. The intensity ratio J does not satisfactorily represent a continuous gray scale in the 256 levels of the commonly used 8-bit pixel. Only 0.18*256 = 46 counts of this represent all shades up to middle gray, while from L^* we know a nominal 50 counts are required, and the problem gets worse for grays darker than middle gray. With the gamma-adjusted proxy $J^{1/\gamma}$ nearly matching L^* in its spacing along the entire gray scale, we have 256 counts to cover the 100 counts needed to produce a continuous appearance. It is not generally realized that the argument is considerable weaker when applied to color pixels[1].

Both gamma matching and the use of gamma-encoding to enhance the ability of small integers (bytes) to represent continuous colors are thus useful, completely valid concepts. Problems arise only from the misuse of gamma. Misuse occurs in two basic ways, 1) failing to compensate for gamma-encoding when necessary, a pervasive problem when using programs such as Photoshop and 2) use of the gamma adjustment as a general tool.

As we have previously indicated, the gamma adjustment, $J_2 = J_1^{\gamma}$, appears most visibly in Photoshop as the Levels tool middle gray adjustment (but see Note on Photoshop Gamma at the end of this document). This adjustment has several attributes which are very

[1] The argument probably is not valid for photographic color images (as opposed to generated color images or monochrome images), when applied to color pixels in which each of the three primaries is represented in a separate 8-bit byte. It is difficult to mathematically evaluate the effect of crosstalk between color channels, but in practice it appears to be either sufficient or nearly sufficient to compensate for the problems seen in monochrome intensity ratio images with a single 8-bit pixel. With a single 8-bit number, there is only one step between, say, 10 and 11 on the gray scale. With color, the equivalent gray step is (r,g,b) = (10,10,10) to (11,11,11). In practice, however, there also are the intermediates (10, 10, 11), (10, 11, 10), (11, 10, 10), (10, 11, 11), (11, 10, 11), and (11, 11, 10). In photographic images, these generally appear systematically sprinkled in any transition from (10,10,10) to (11,11,11) so that a photographic gray scale will appear continuous while a computer-generated one might not.

seductive indeed. Foremost is the fact that highlight and shadow detail are both completely retained. Following closely behind is the fact that in any main channel (RGB) gamma adjustment, all grays will remain precisely gray, no matter how large the adjustment. In practice, these two attributes are compromised somewhat by low precision truncation but still are largely true. We also know that gamma adjustments to the color channels will have very little effect on highlights (or deep shadows), so one can adjust for proper color in highlights and then gamma adjust the color channels to control the mid-tones. In fact, this is often how "color balance" is done. Finally, less well-known is the fact that if colors are misadjusted using color gamma adjustments, the result will almost always be muddier, less intense color; bad color which may more easily sneak by.

Now let us examine several adjustments and how they actually behave in practice.

What All This Means in Photoshop

Overall Intensity •••
Photoshop Levels Tool - Highlights Adjustment

Suppose that in a stored image we find that $0 \le J < m \le 1$. That is, the highlights generally appear too dark and there are no bright highlights. We wish to expand the values J by effectively reducing the maximum for J to what appears to be its proper value, m. Then $J_{adj} = J/m$, and we will perform this operation on the entire image. To see how this affects the image, we will take the ratio of two arbitrary values within the image, J_1/J_2 and compare the ratio of the adjusted values of those same two points $J_{adj1}/J_{adj2} = (J_1/m)/(J_2/m) = J_1/J_2$. The ratio is not changed by this transformation. This operation is exactly what Photoshop does in the Levels tool, when operating on Highlights.

In the case of a gamma encoded image[2], the values to be operated upon are J^γ and the corresponding adjustment would be $J^\gamma_{adj} = J^\gamma/m^\gamma$. So, $J^\gamma_{adj1}/J^\gamma_{adj2} = (J^\gamma_1/m^\gamma)/(J^\gamma_2/m^\gamma) = J^\gamma_1/J^\gamma_2$. Since $(J_{adj1}/J_{adj2})^\gamma = (J_1/J_2)^\gamma$, then $J_{adj1}/J_{adj2} = J_1/J_2$. Even when working with gamma adjusted intensities, the highlight adjustment preserves the ratios between intensities.

Note that this analysis also applies when $m > 1$. This situation occurs when something has occurred to chop off the brightest highlights, either overexposure or something in previous processing (or that there were no bright highlights). Since values of J greater than one are not retained, it is impossible to regain what the highlights should have been, but the tones below the highest highlights may be restored to their proper values.

The Photoshop Levels Tool Highlight Adjustment operates exactly like a filter or adjustment of overall illumination, whether the image is gamma-encoded or not, and is in full accord with the requirements for color integrity. This is true overall and also for each of the color channels. The Highlights adjustment must be the *primary* tool for color balance. We recommend that in general, the Levels tool "highlight" adjustment actually be used to properly balance or match a near-neutral tone in the middle to lighter gray range, with simultaneous observation of the color channel histograms, with adjustment if necessary to insure highlights are not lost or grayed. The overall image should appear to be in color balance if the match choice was proper and the image had color integrity to start with.

---

[2] A gamma encoded image is really $J^{1/\gamma}$, but in the following series of analyses we will use J^γ in order to greatly simplify the appearance of the equations.

Black Level •••
Photoshop Levels Tool Shadows Adjustment

Now suppose we have the situation complimentary to the above, where the darkest blacks are missing: $0 \le b < J \le 1$. Here we wish to expand the values of J by moving b to 0 so that $J_{adj} = (J - b)/(1 - b)$. In this case, $J_{adj1}/J_{adj2} = [(J_1 - b)/(1 - b)]/[(J_2 - b)/(1 - b)] = (J_1 - b)/(J_2 - b)$. Obviously J_{adj1}/J_{adj2} is not the same as J_1/J_2 unless $b = 0$, and the question is how much it differs from that. To investigate, we define a multiplier α such that $J_{adj1}/J_{adj2} = \alpha(J_1/J_2)$. Thus $\alpha(J_1/J_2) = (J_1 - b)/(J_2 - b)$. This simplifies to:

$$\alpha = \frac{1 - b/J_1}{1 - b/J_2}$$

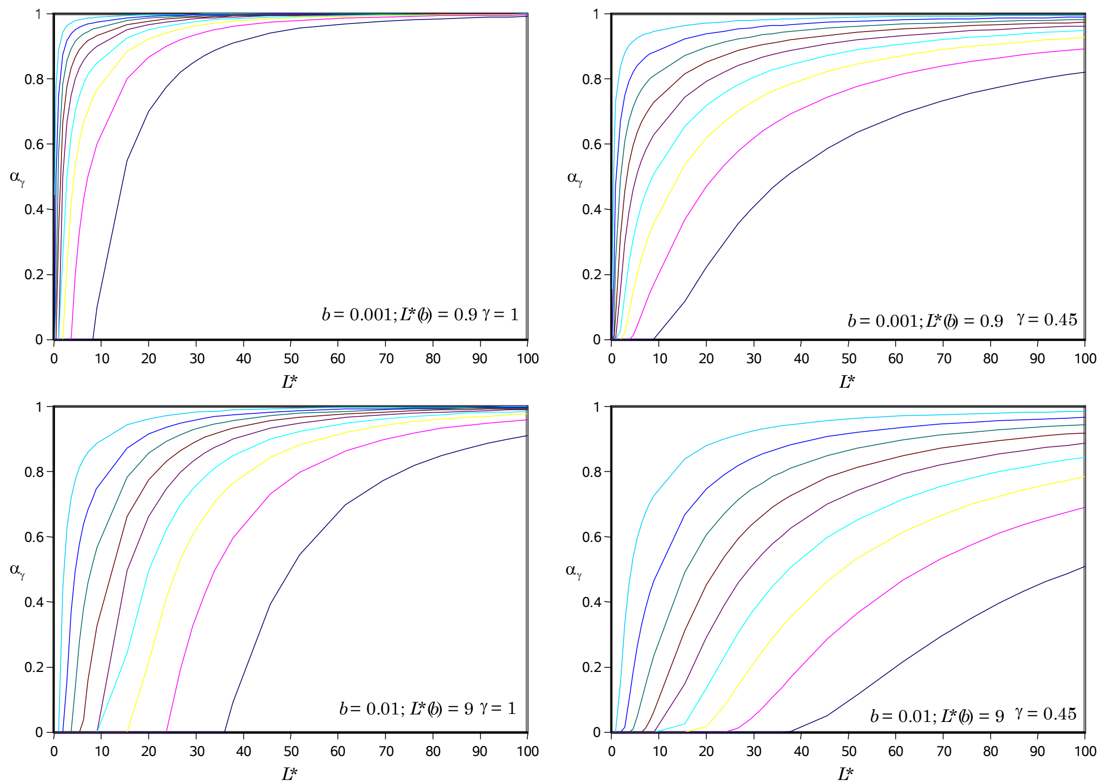
So that α is nearly 1 except where J_1 and J_2 are unequal and one of them approaches b, which is to say, for very dark shades of color.

For the situation in which we work with gamma adjusted values, we will have $J^{\gamma}_{adj} = (J^{\gamma} - b^{\gamma})/(1 - b^{\gamma})$, where we recognize that for direct comparison, the b value that we wish to trim to will be stored as b^{γ}. Thus $J^{\gamma}_{adj} = (J^{\gamma} - b^{\gamma})/(1 - b^{\gamma})$ and $J^{\gamma}_{adj1}/J^{\gamma}_{adj2} = [(J^{\gamma}_1 - b^{\gamma})/(1 - b^{\gamma})]/[(J^{\gamma}_2 - b^{\gamma})/(1 - b^{\gamma})] = (J^{\gamma}_1 - b^{\gamma})/(J^{\gamma}_2 - b^{\gamma})$. Thus $J_{adj1}/J_{adj2} = [(J^{\gamma}_1 - b^{\gamma})/(J^{\gamma}_2 - b^{\gamma})]^{1/\gamma}$, and this also is not the same as J_1/J_2 unless $b = 0$. We still want $J_{adj1}/J_{adj2} = \alpha_{\gamma}(J_1/J_2)$ where the γ subscript denotes the gamma adjusted case. So, $\alpha_{\gamma}(J_1/J_2) = [(J^{\gamma}_1 - b^{\gamma})/(J^{\gamma}_2 - b^{\gamma})]^{1/\gamma}$, and this simplifies to:

$$\alpha_{\gamma} = \left(\frac{1 - (b/J_1)^{\gamma}}{1 - (b/J_2)^{\gamma}} \right)^{1/\gamma}$$

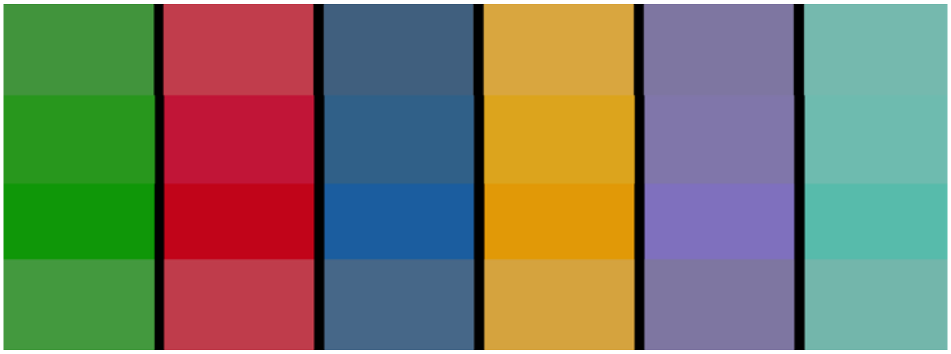
This is, of course, a generalized form of the previous α equation, reducing to it for $\gamma = 1$. Again, it is quite possible to have a situation with $b < 0$, representing shadows that have been blocked by underexposure or prior calculations.

In order for the cardinal rule of color balance to be followed, it would be necessary for $\alpha_{\gamma} = 1$. Given that this can only be approximated, it is next useful to know how much of the color map is covered when α_{γ} is still acceptably close to 1. That can be visualized in the following plots.



Effect of Shadow / Black Point Adjustments

Shadow adjustments at unity gamma are shown on the left and the exact same cases applied to the common gamma encoding of $\gamma = 0.45$ are shown on the right. Curves for J_1/J_2 values of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 are shown, with $J_1/J_2 = 0.1$ the lowermost curve. J_2 is always brighter than J_1 and so α_γ values are plotted versus the value of L^* that corresponds to J_2 *after* it has been revised by the shadow adjustment, to give a better idea of where the effects are placed on the adjusted visual range. The top pair of graphs correspond to the shadow adjustment raised to $b = 0.001$, which corresponds to $L^* = 0.9$ (measured *before* the shadow adjustment). This is a minor adjustment, corresponding to about 10 or 11 (of 255) on the gamma encoded image. The closer the curves are to the $\alpha_\gamma = 1$ axis, the better, and they represent successive error bands of a sort. From these graphs, the correction done in unity gamma is not a problem except in the darkest tones where J_1 approaches b, while the gamma encoded correction shows definite effects through more of the visual range. The lower pair of graphs correspond to the shadow adjustment raised to $b = 0.01$, which corresponds to $L^* = 9$ (measured *before* the shadow adjustment). This is a larger adjustment, corresponding to about 32 (of 255) on the gamma encoded image. Here the effects have become important even for unity gamma, which looks much like the gamma encoded case at $b = 0.001$. The gamma encoded case at $b = 0.01$ showing pronounced effects even for values of J_1/J_2 near 1, meaning that even the near-grays are starting to be affected. Please note that the shadow has been trimmed at exactly the same level of gray in both the unity gamma and the gamma encoded cases. The difference here is entirely due to applying the adjustment formula directly to gamma encoded data rather than to intensity ratios.



Shadow Adjustment Comparison Color Chart

The comparison chart above has four approximately equal-sized horizontal bands. An original color band image was shadow adjusted 10 counts (of 255) at unity gamma and then adjusted to 0.45 gamma for display as the second band from the top. The same starting color band image, gamma encoded to 0.45 gamma, was then shadow adjusted 59 [$= 255(10/255)^{0.45}$] counts of 255, to form the third band from the top. The original color band image was then placed in the top row. For each color cell in the top row the overall percent intensity was adjusted until a visual best match with the color cell below was achieved. This method eliminated the sometimes considerable difference in visual density, allowing direct observation of the color effects. As we have shown, a percentage adjustment such as we applied to the top band preserves its color integrity. The same procedure was done for the bottom band, matching it to the one above it. It can be seen that the third band from the top is definitely the worst, while the second is just becoming visible. The important thing to realize here is that *no adjustment to the separate color channels was made*. These color changes are *entirely* due to a shadow level adjustment to the overall RGB = gray scale!

Unity Gamma, Red Shadow Adjustment
Trim Below $b_R = 8/255$

Gamma Encoded 0.45, Red Shadow Adjustment
Trim Below $b_R = 8/255$

The color bars above illustrate the effect of shadow/black adjustment of a color channel. In each bar, the upper 1/3 is the original image, the center 1/3 is shadow/black point adjusted, and the lower 1/3 is the original image adjusted by color correction filter (percentage adjustment) so that the next to the last color bar (magenta-blue) exactly matches the shadow adjusted version. In both cases the adjustment was to red, trimming below $b_R = 8/255$, which gamma encodes to 54/255. The effect is larger for the gamma encoded image, and has a definite effect on color integrity.

The shadow/black level adjustment is a good deal less straightforward than the highlight adjustment. It is an important adjustment which should usually be done first. We will suggest how this be done, but there are still loose ends under study which could make us revise our recommendations later. First, we again recommend that the "shadow" adjustment actually be used to properly balance or match a near-neutral tone in the darker gray range, with simultaneous observation of the color channel histograms, with adjustment if necessary to insure shadow detail is not lost or grayed. Second, based on the analysis above, we would like to recommend that if the adjustment is other than minor, first convert to intensities (gamma adjust by $\gamma = 2.2$ or 1.8), correct the black level, then return to gamma-encoded (gamma adjust by $\gamma = 0.45$ or 0.56). We would *like* to recommend that, but we cannot. Photoshop does something strange in place of the expected black level adjustment when the adjustment is over some trigger value. We have not figured out exactly what, why, or precisely where it cuts in. Consequently we are not sure exactly what to recommend. With color negatives, we currently recommend doing black level adjustments prior to inverting, using the highlight tool, but that only becomes satisfactory in comparison to the alternatives.

Probably the best compromise is to try to keep black level corrections to a minimum, and concentrate on getting a good color balance in the dark grays. Then if the shadows are left too light, trim it up later with a curves adjustment using the method described later in this document.

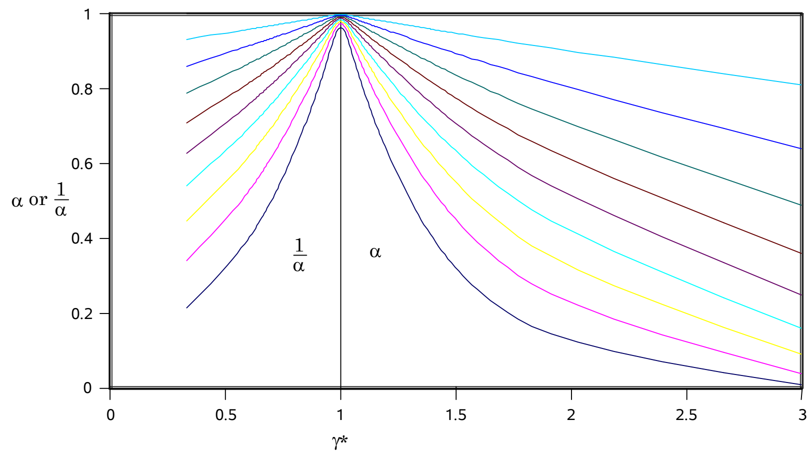
Middle Gray Level •••

Photoshop Levels Tool Middle Gray Adjustment

As stated before, the Photoshop Levels tool middle gray adjustment is simply a gamma adjustment $J_{adj} = J^{\gamma^*}$ (but see Note on Photoshop Gamma at the end of this document). We use γ^* to distinguish this from the gamma-encoding γ commonly applied to J in image files. If we analyze this adjustment similarly to the above, we find the ratio of $J_{adj1}/J_{adj2} = J_1^{\gamma^*}/J_2^{\gamma^*} = (J_1/J_2)^{\gamma^*}$. Thus the requirement that $J_{adj1}/J_{adj2} = \alpha(J_1/J_2)$, means that $\alpha(J_1/J_2) = (J_1/J_2)^{\gamma^*}$ or:

$$\alpha = \frac{\left(\frac{J_1}{J_2}\right)^{\gamma^*}}{\frac{J_1}{J_2}} = \left(\frac{J_1}{J_2}\right)^{\gamma^*-1}$$

Now consider the case in which J is already gamma-encoded: $J^{\gamma}_{adj} = (J^{\gamma})^{\gamma*}$. We need to examine $J^{\gamma}_{adj1}/J^{\gamma}_{adj2} = (J^{\gamma}_1/J^{\gamma}_2)^{\gamma*}$, so that $(J_{adj1}/J_{adj2})^{\gamma} = ((J_1/J_2)^{\gamma})^{\gamma*} = ((J_1/J_2)^{\gamma*})^{\gamma}$, and $(J_{adj1}/J_{adj2}) = (J_1/J_2)^{\gamma*}$, so we find that for this adjustment there is no difference in effect between operating on intensities and operating on gamma-encoded intensities.



Effect of Gamma (Middle Gray) Adjustment

Curves for J_1/J_2 values of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 are shown, with $J_1/J_2 = 0.1$ the lowermost curve. Note that this graph applies to all values of J_1 and J_2, that is, to both light and dark. α is plotted for $\gamma^* > 1$ and $1/\alpha$ plotted for $\gamma^* < 1$ to better compare the effects on either side of unity gamma. From this graph it can be seen that any change in gamma will affect color balance, with the effect increasing as gamma departs from unity.

Red Gamma 1.5 Adjustment (Top) Compared to
Color Compensation Filter Adjustment
Matched on Third (Blue)

Red Gamma 0.667 Adjustment (Top) Compared to
Color Compensation Filter Adjustment
Matched on Fourth (Yellow)

The above illustrates the sort of color distortion introduced by gamma adjustments. The chosen examples are not exceptional in either direction. There is considerable distortion and loss of color integrity. They illustrate the visual effect of muddying the colors that is so commonly seen. When this muddying effect is noticed, often the Hue/Saturation control is use to "correct" it, leading to even stranger distortions.

The following pair of images is used to illustrate how deceptive the problem can be:

The original image for the above has reasonably good color integrity and is in fact not far from being in color balance. The image on the left was "color balanced" from this original using the standard method of setting the black point, setting the white point, and adjusting the middle gray a considerable distance from the original. The result should be acceptable, or nearly so, to most people. The image on the right was produced from the same original by using color correction filters (percentage adjustments as described above) so that its mid-tones exactly match the same mid-tones in the image on the left. That the mid-tones in the left image are really that far off may be hard to believe, and we suggest that the skeptical engage in some personal checking. For more in-depth checking we suggest exploring the more practical summary of this document that is found on our web page www.c-f-systems.com.

The Levels tool middle gray adjustment, gamma, should ***never*** be used as an element of color balance correction where color integrity is to be preserved. This is the case with both individual color channel middle gray (gamma) adjustments and with overall (RGB) middle gray (gamma) adjustments. All of these play havoc with the Cardinal Rule of Color Balance. The middle gray (gamma) adjustment can be constructively used for minor *tweaking* as the color integrity loss is normally invisible over small changes. It also can be a powerful tool, used in combination with others, when it is clear that color integrity of an image has already been seriously compromised. While it generally will not restore color integrity, it may aid in coercing the color into a visually acceptable compromise.

Preserving Color Integrity During Curves and Levels Middle Gray Adjustments. •••

Even if only applied to the overall (RGB) channel, both the Levels tool middle gray adjustment (gamma) and the Curves tool play havoc with the Cardinal Rule of Color Balance. After these operations the R, G, and B in individual pixels will not remain in proportion as required. Even experienced photographers can be trapped into thinking they might because they both these tools *do* preserve grays perfectly, and for many years photographers have been conditioned that if you care for the grays (neutrals) then everything else will follow. The problem with these tools is that the other colors *do not* follow, and as colors get away from neutral gray they can be warped quite far from where they started.

Fortunately there is an easy and effective trick available. Convert the image to CIELAB (Using Image→Mode→Lab color). Then use either the Curves tool or the Levels tool middle gray adjustment (gamma) and operate on the Lightness channel. This will leave the color balance largely unaffected. Convert back to RGB for further work if Lab mode is not comfortable.

While this method will preserve color integrity when using curves and gamma corrections to adjust the tonal scale, *do not* use it for gamma adjustments that are intended for gamma matching. For instance, to convert a unity gamma image and by gamma encoding it to the typical $\gamma = 0.45$, use the gamma correction directly on the RGB image. In this case the color distortions must be present, as they will be corrected by matching opposite distortions before the image is viewed or printed.

There are cases in which it is more difficult to determine whether gamma encoding is appropriate or not. For example, it is well known images intended to be projected in darkened rooms need higher contrast to look right. Is it more appropriate for that extra contrast to be generated by gamma encoding or by means such as the above that will preserve color integrity? There are numerous other specific situations that systematically call for more or less contrast. We claim no definitive answer for this at the moment, but we believe that since it is a visual effect, it probably would more closely be matched by a gamma encoding.

At this time we recommend trying both ways when there is any doubt.

Continue on to Note on Photoshop Gamma just below for important information on performing gamma adjustments in Photoshop.

Note on Photoshop Gamma

We refer to the Photoshop Levels tool middle gray adjustment as being a gamma adjustment. Like so many things in Photoshop, that is only *nearly* true. With a few exceptions it is near enough to being true that we have decided to continue calling it a gamma adjustment in our web pages and CFS documents, qualifying it when most important. Note that *Adobe* does not openly advertise the adjustment as "gamma," but there are places where they mention the (near) fact that it is gamma. We have not found that they mention the disparities, however.

The first problem is really a non-problem once one is aware of it. The "middle gray" parameter in Levels is a decimal number centering about 1.00, just as would be the case for gamma. Unfortunately, instead of gamma, γ, the number shown for middle gray corresponds to $1/\gamma$. Thus, if one wants to apply a gamma of 0.5 to the image, one must dial in 2.00 as the middle gray adjustment.

The second problem can be fairly serious in some cases. It appears that the gamma equation is incorrectly coded or has been modified for some inexplicable reason for $1/\gamma > 1.0$. The variation is small enough to be of little concern below about $1/\gamma = 1.4$. Above 1.4, the effect gradually increases, peaking somewhere around $1/\gamma = 2.5$ or 3.0 and again becoming less at higher gammas. It occurs only in the darkest part of the scale, affecting values about 25 and less (out of 255) at the peak, with minor effect at 25 increasing to a fairly large effect at 10, and then lesser effects back to 0. The effect is always to make the value less than it should be, by as much as about 25% at its worst. Or perhaps its worst is that it will change 0 values, to 1, 2, even above 100 for very high gammas! This is true for Photoshop Versions 7 and CS(8) and presumably is baggage left over from the earliest versions.

In most cases, this is not of major concern. If gamma adjustments even as high as 1.4 are being used, concern for color integrity has already been compromised. However, it *is* of concern when a gamma shift is being made for the purpose of gamma matching. For example, if one has been working with a negative (or even positive) image scanned at $\gamma = 1$ and has reached the stage where it needs to be gamma encoded to match with current conventions, then a gamma encode of $\gamma = 0.45$ should be applied to it. This means that middle gray will have to be adjusted to $1/\gamma = 2.2$ (or 1.8), right in the range of the worst problem. We currently suggest that *three* gamma adjustments be applied in succession, each with $1/\gamma = 1.3$. As $1.3^3 = 2.2$, the overall adjustment will be correct. (For $1/\gamma = 1.8$, use *two* gamma adjustments of $1/\gamma = 1.34$.) All such adjustments should be carried out in 16 bits/channel mode, of course.

For the curious, it is easy to demonstrate this peculiarity of Photoshop (yet another). Create a new image 1000 pixels high by 300 wide, RGB, 16 bit, white background. Set the active colors to pure white (255,255,255) and pure black (0,0,0). Draw a gradient, straight as possible, top to bottom of the new image. Then duplicate the layer (Layer→Duplicate Layer). Apply a Levels middle gray correction of 0.5 (Image→Adjustment→Levels, set middle number to 0.5 and exit). Apply a Levels middle gray correction of 2.0 (Image→Adjustment→Levels, set middle number to 2.0 and exit). Invert the image (Image→Adjustment→Invert). All of these adjustments have been to the background copy. Now on the Layers tab of the Layers-Channels-Paths box, carefully adjust the Opacity to 50% for the background copy layer. The image should gray out to fairly uniform. Now flatten the image (Layer→Flatten Image). If the mouse cursor is hovered over this image, the lower portion should show a fairly uniform (127,127,127) in the Info tab of the Navigator-Info box. In the upper portion of the image (the end that started as black), the numbers will rise to as much as (136,136,136). To see the pattern, Image→Adjust→Equalize. This can be done in 16-bits/channel mode as described above only in Photoshop CS(8). However, the particular test also works reasonably well in 8 bits/channel and can be done in that mode in earlier versions of Photoshop.

The adjustment to $1/\gamma = 0.5$ should have been completely matched by the subsequent adjustment to $1/\gamma = 2.0$, so the two layers should have ended up as identical. When the duplicate was inverted and set to 50% opacity, it should have completely neutralized the original to a uniform middle gray (127 or 128) over the entire image. For any who remain curious, the above demonstration can be repeated, but in place of the single adjustment of $1/\gamma = 2.0$, do three successive adjustments of $1/\gamma = 1.26$. Since $1.26^3 = 2.0$, this should be equivalent, but instead it will be found that the maximum excursion is to (130,130,130). The 8 bits/channel version naturally will show considerable more scatter and banding in the results.

In these tests, the effect will be visible over most of the upper half of the gradient, and that may seem to conflict with the comment above that it mostly affects values below 25. However, the initial adjustment to $1/\gamma = 0.5$ will convert 80 to 25, so the primary effect will cover the part of the gradient from 0 to 80 in the original.