



Negative to Positive

CFS-244

April 19, 2004

David Dunthorn
www.c-f-systems.com

Personal Note

This report has a lot of algebra and basic physics of photography in it, but it is *not* "rocket science." Nevertheless, it *is* quite tricky to wade through and we have not seen this material treated in a similar way elsewhere. We *have* seen a great deal of basically incorrect material on the topic, both on the web and in print. For a simple example, what Photoshop calls "Invert," that is, Image→Adjust→Invert, sounds like it should convert a negative to a positive, and the result *looks* like that may be happening, but that process - which also appears in other imaging programs and scanner software to an unknown extent - is far removed from what actually needs to be done to correctly convert a scanned negative image to a positive. It is not necessarily easy to tell from results whether scanner software is based on the principles in this report because it is possible to *approximate* what we present here by "profiling," that is, by having a test case for which the results are known and adjusting a series of arbitrary correction factors to make the results match. With a basic principles approach, the equations have physical meaning rather than being arbitrary curve fit forms and thus are more likely to respond better to variations in processing, manufacture, and storage.

If any reader is aware of this material being published and available elsewhere, we would appreciate knowing about it so we can compare notes. Please contact us through our web site. 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.

In this document we make frequent references to the companion documents CFS-242, *Film Gamma Versus Video Gamma* and CFS-243 *Maintaining Color Integrity in Digital Photography*. Refer to those documents, available on our web site, for the further explanation of the symbols, equations, and concepts used here.

Much of the material and its treatment in this document is original with us and this document is

Copyright © 2004 by C F Systems

If you plan to use our original material in any published form, please contact us at www.c-f-systems.com for terms and conditions.

Negative, Positive, and the S-Curve

In CFS-242, *Film Gamma Versus Video Gamma*, we dealt with the S-curve, used to describe traditional or silver-based images, expressing the sensitivity of film to light. This S-curve relationship is physically measured for each type of film under specific processing:

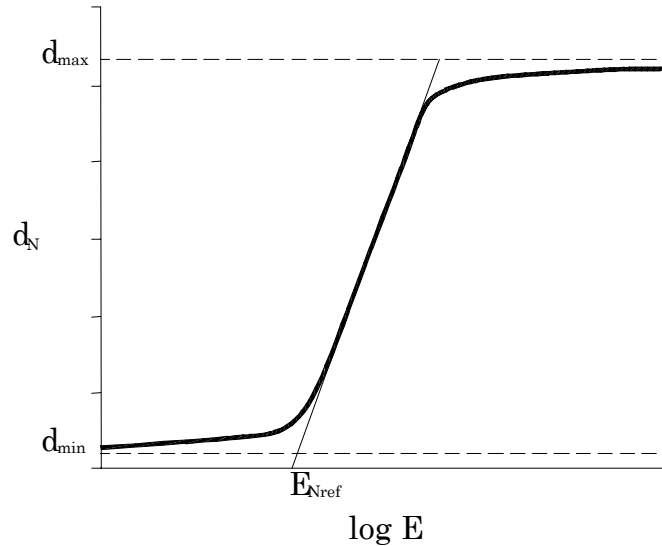


Figure 1
S-Curve for a Negative Photographic Film

Where the straight line portion of the "S" has the equation:

$$d_N = \gamma_N \log_{10}(E/E_{N\text{ref}})$$

Here d is the photographic density, with $d = 0$ meaning transparent film and as d becomes larger, transparency decreases (the film becomes darker). E is the light exposure the film receives, measured in candela-sec/m² or similar units. γ_N is the slope of the straight-line portion of the S-curve. $E_{N\text{ref}}$ is the exposure for which $d = 0$ when the straight-line portion is extrapolated through the $\log E$ axis.

Photographic density is the base 10 logarithm of an intensity ratio. In this case, it is the ratio between the intensity of the light impinging on a the negative to the light passing through a the negative. That is:

$$d = \log_{10}\left(\frac{I_{\text{max}}}{I}\right)$$

Such intensity ratios turn up very often in digital photography and are usually the inverse of the one above, whence they become *normalized*, confined between 0 and 1 in value.

We use the notation $J = I/I_{\max}$ as a shorthand for such normalized intensity ratios, and thus $d = -\log_{10}(J)$.

In CFS-242, *Film Gamma Versus Video Gamma*, we also showed that a positive silver-based image has a similar form to the negative, but with the "S" mirror reversed, and with the straight-line portion having the equation:

$$d_P = -\gamma_P \log_{10}(E/E_{\text{Pref}})$$

We showed that this is true whether the positive image is produced by reversal processing, such as is done for slides, or is printed as a positive from a negative.

Tracing the Path from Negative through to Positive

The process under investigation involves producing a positive image from a scene by first creating a negative film image. There will be separate exposures for the positive and negative images and we use E_S as the exposure to the scene for the negative and E_N as the exposure to the illuminated negative to produce the positive:

$$\begin{aligned} d_N &= \gamma_N \log_{10}(E_S/E_{N\text{ref}}) \\ d_P &= \gamma_P \log_{10}(E_N/E_{\text{Pref}}) \end{aligned}$$

To find what exposure the positive media will get from light passing through the negative media, we apply the definition of density for the negative and use the J notation for the normalized intensity ratio:

$$d_N \equiv \log_{10}(I_{N\max}/I_N) = -\log_{10}(I_N/I_{N\max}) = -\log_{10}(J_N)$$

The exposure to produce this density will be proportional to the scene intensity applied for an exposure time t_S through a f -aperture: $E_S/E_{N\text{ref}} = k_N J_S$, so that

$$d_N = \gamma_N \log_{10}(k_N J_S)$$

The k_N is a single constant factor relating scene intensity ratios to negative densities. One can think of k_N as being determined by taking a light meter reading of the scene. Make k_N larger and the negative of the scene will be darkened, with k_N smaller, the negative will be thin. In any case, we will be selecting some specific $J_S = J_{SR}$ that we want to result in some specific density in the negative, $d_N = d_{NR}$, or equivalently $J_N = J_{NR}$ so:

$$\begin{aligned} d_{NR} &= \gamma_N \log_{10}(k J_{SR}) = \gamma_N \log_{10}(k_N) + \gamma_N \log_{10}(J_{SR}) \\ -\log_{10}(J_{NR}) &= \gamma_N \log_{10}(k J_{SR}) = \gamma_N \log_{10}(k_N) + \gamma_N \log_{10}(J_{SR}) \end{aligned}$$

So, since

$$\begin{aligned} d_N &= \gamma_N \log_{10}(k J_S) \\ \log_{10}(J_N) &= -\gamma_N \log_{10}(J_S) - \gamma_N \log_{10}(k) = -\gamma_N \log_{10}(J_S) + \log_{10}(J_{NR}) + \gamma_N \log_{10}(J_{SR}) \end{aligned}$$

$$J_N = \frac{J_{SR}^{\gamma_N} J_{NR}}{J_S^{\gamma_N}}$$

or

$$\frac{J_N}{J_{NR}} = \left(\frac{J_{SR}}{J_S} \right)^{\gamma_N}$$

The exposure for the positive will be this intensity applied for an exposure time t_N through a f -aperture so that $E_N/E_{\text{Pref}} = k_P J_N$ where k_P accounts for both the exposure time and f -aperture applied to the light passing through the negative:

$$d_P = \gamma_P \log_{10}(k_P J_N)$$

as before, we meter an exposure which will match a specific density in the positive d_{PR} , or equivalently J_{PR} , to a specific intensity ratio in the negative J_{NN} (which may be different than J_{NR}):

$$d_{PR} = \gamma_P \log_{10}(K J_{NN}) = \gamma_P \log_{10}(k_P) + \gamma_P \log_{10}(J_{NN})$$

$$-\log_{10}(J_{PR}) = \gamma_P \log_{10}(K J_{NN}) = \gamma_P \log_{10}(k_P) + \gamma_P \log_{10}(J_{NN})$$

$$\log_{10}(J_P) = -\gamma_P \log_{10}(J_N) - \gamma_P \log_{10}(K) = -\gamma_P \log_{10}(J_N) + \log_{10}(J_{PR}) + \gamma_P \log_{10}(J_{NN})$$

$$J_P = \frac{J_{NN}^{\gamma_P} J_{PR}}{J_N^{\gamma_P}}$$

or

$$\frac{J_P}{J_{PR}} = \left(\frac{J_{NN}}{J_N} \right)^{\gamma_P}$$

The Connection between Exposure Parameters

The base requirement is that the intensity ratios of the positive image match the intensity ratios of the scene, that is, $J_P = J_S$:

$$J_N = \frac{J_{SR}^{\gamma_N} J_{NR}}{J_S^{\gamma_N}}$$

so that

$$J_P = \frac{J_{NN}^{\gamma_P} J_{PR}}{J_N^{\gamma_P}} = \frac{J_{NN}^{\gamma_P} J_{PR}}{J_{SR}^{\gamma_N \gamma_P} J_{NR}^{\gamma_P}} J_S^{\gamma_N \gamma_P}$$

First, it is obvious that to have $J_P = J_S$ requires *matched gammas*, that is, $\gamma_N \gamma_P = 1$, so that

$$J_P = \frac{J_{NN}^{\gamma_P} J_{PR}}{J_{SR} J_{NR}^{\gamma_P}} J_S$$

and further that

$$\frac{J_{NN}^{\gamma_P} J_{PR}}{J_{SR} J_{NR}^{\gamma_P}} = 1$$

Thus, if we wish the positive image to match the scene, $J_P = J_S$, we must have matched gammas and also require that the exposure of the negative and the positive be related:

$$J_{NN}^{\gamma_P} J_{PR} = J_{SR} J_{NR}^{\gamma_P}$$

or

$$\frac{J_{PR}}{J_{SR}} = \left(\frac{J_{NR}}{J_{NN}} \right)^{\gamma_P}$$

Numerical Examples of the Scene to Negative to Positive Trace

The CIE L^* "lightness" function is designed to mimic the behavior of human vision (as explained in CFS-243, *Maintaining Color Integrity in Digital Photography*) and runs from $L^* = 0$, pure black, to $L^* = 100$, the brightest discernable area in a scene. Each increment of 1 in L^* approximates a barely visible step in a gray scale from black to white. Each value of L^* corresponds to an intensity ratio Y/Y_n , through a somewhat complicated formula. Y/Y_n corresponds to our J_S . In Tables 1 and 2 below we use the L^* scale to trace the resolution of a scene J_S , as a positive image, J_P , going through an intermediate negative, J_N . In this way the rows in the table are visually evenly spaced along a gray scale, making it easier to understand what is happening. In preparing Table 1 we chose our exposure so that $J_{SR} = 1$ and $d_{NR} = d_{Nmax}$, that is, the maximum density d_{Nmax} occurs at the maximum intensity ratio in the source. While that appears mathematically sensible and is useful for illustration, note that one does not normally tie exposure to the brightest highlights in real situations.

The requirement for the positive is

$$J_{NN}^{\gamma_P} J_{PR} = J_{SR} J_{NR}^{\gamma_P}$$

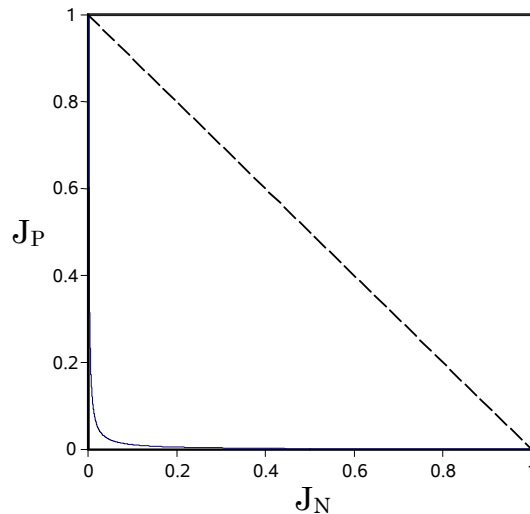
If we expose the positive using $J_{SR} = J_{NN} = 1$, and use $\gamma_N = \gamma_P = 1$, then $J_{NR} = J_{PR}$. We have used $J_{NR} = J_{PR} = 0.001107$ which a density range of about 2.95, going from 1 to 100 on the L^* scale.

Table 1
 From Scene through Negative to Positive
 For $\gamma_N = \gamma_P = 1$ and $J_{NR} = J_{PR} = 0.001107$

L_S^*	$J_S = [Y/Y_n]_S$	d_N	J_N	L_N^*	d_P	$J_P = [Y/Y_n]_P$	L_P^*
100	1	2.955832	0.001107	1	0	1	100
96	0.900078	2.910112	0.001230	1.111015	0.045720	0.900078	96
92	0.807044	2.862729	0.001372	1.239090	0.093103	0.807044	92
88	0.720653	2.813558	0.001536	1.387631	0.142274	0.720653	88
84	0.640658	2.762458	0.001728	1.560896	0.193374	0.640658	84
80	0.566813	2.709272	0.001953	1.764251	0.246560	0.566813	80
76	0.498872	2.653822	0.002219	2.004520	0.302010	0.498872	76
72	0.436590	2.595906	0.002536	2.290477	0.359926	0.436590	72
68	0.379720	2.535296	0.002915	2.633517	0.420536	0.379720	68
64	0.328017	2.471728	0.003375	3.048625	0.484104	0.328017	64
60	0.281233	2.404899	0.003936	3.555766	0.550933	0.281233	60
56	0.239124	2.334456	0.004630	4.181927	0.621376	0.239124	56
52	0.201443	2.259985	0.005496	4.964177	0.695847	0.201443	52
48	0.167945	2.180998	0.006592	5.954346	0.774834	0.167945	48
44	0.138382	2.096912	0.008000	7.226370	0.858920	0.138382	44
40	0.112510	2.007022	0.009840	8.857109	0.948810	0.112510	40
36	0.090082	1.910468	0.012289	10.76919	1.045364	0.090082	36
32	0.070852	1.806182	0.015625	12.99996	1.149650	0.070852	32
28	0.054574	1.692816	0.020285	15.63632	1.263016	0.054574	28
24	0.041002	1.568638	0.027000	18.79995	1.387194	0.041002	24
20	0.029891	1.431366	0.037037	22.66661	1.524466	0.029891	20
16	0.020993	1.277908	0.052734	27.49994	1.677924	0.020993	16
12	0.014064	1.103932	0.078717	33.71422	1.851900	0.014064	12
8	0.008856	0.903092	0.124999	41.99992	2.052740	0.008856	8
4	0.004428	0.602060	0.250000	57.07542	2.353772	0.004428	4
1	0.001107	0	1	100	2.955832	0.001107	1

It can immediately be seen that $L_S^* = L_P^*$ and $J_S = J_P$ for the entire table, meaning that the positive image is a match for the scene, which is our base requirement. Closer inspection shows some surprising results. On the negative, J_N , the entire range of intensities from 0.25 to 1, that is the part of the negative that is about middle gray in the negative (for middle gray, $J = 0.18$ and $L^* = 50$) through clear, corresponds to only the very darkest grays in the original scene (and in the positive image). In the negative image the very darkest areas, mostly well below middle gray in the negative, are responsible for most of the tones in the final positive image, while all the lighter parts of the negative, from clear through middle gray, form just the first 4 darkest steps out of the 100 steps in the gray scale in the positive image.

The Connection between Negative and Positive is Extreme



Positive Intensity Ratio versus Negative Intensity Ratio

For $\gamma_N = \gamma_P = 1$ and $J_{NR} = J_{PR} = 0.001107$

The graph above shows how extreme this relationship really is. The curve barely breaks away from the axes even near the (0, 0) point in the lower left corner. Most of the intensity ratios in the positive image are represented by just the lowest intensity ratios in the negative, and vice versa. It should be noted that what Photoshop calls "Invert," that is, Image→Adjust→Invert, corresponds to the diagonal dashed line in the above graph, far removed from the actual inversion of a negative to a positive.

Another way of describing the Photoshop Invert is that it *subtracts* the negative image from a totally white image. The belief that this subtraction will actually invert a (film) negative image to an accurate positive is fairly widespread, whether the function is in Photoshop or in other image editing or scanner software. In photographic media, however, it is not intensity ratios, J 's, which are additive, but *densities*, $d = \log_{10} J$, that are additive. See *Mathematical Proof of Photographic Gamma Equation for Positives* in CFS-242, *Film Gamma Versus Video Gamma* - or just think of the way that color compensation filters are added by *density*. Negatives are just complex filters, in which the color of the filter changes for each "pixel." Adding and subtracting densities is the equivalent of multiplying and dividing intensity ratios.

A Numerical Example with Non-Unity Gammas

Table 2 shows the same trace for a matched gamma where both $\gamma_N = 0.5$ and $\gamma_P = 2$. (To be matched, $\gamma_N \gamma_P = 1$). Again we require that

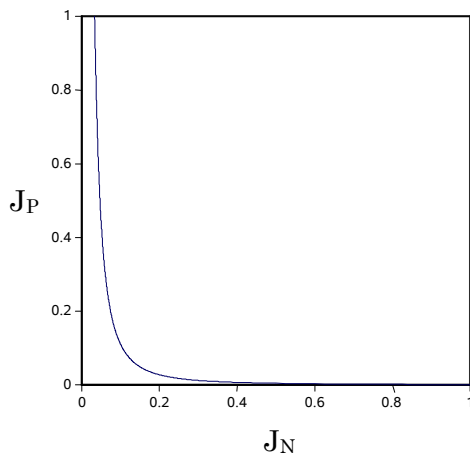
$$J_{NN}^{\gamma_P} J_{PR} = J_{SR} J_{NR}^{\gamma_P}$$

If we expose again using $J_{SR} = J_{NN} = 1$, this time $\gamma_P = 2$, and so and $J_{NR}^2 = J_{PR}$. We again have used $J_{PR} = 0.001107$ which corresponds to going from 1 to 100 on the L^* scale, which is what we wish for the positive image.

Table 2 represents a thinner negative, that is, one with a lower d_{\max} . If we allowed the negative to be exposed to the same d_{\max} as it was in Table 1, it would be possible to make a full-range positive image from a portion of the density range of the negative. There would still be (mathematically) an exact correspondence between the scene intensity ratios and the positive image intensity ratios, at least for the range of ratios covered by the positive image.

Table 2
 From Scene through Negative to Positive
 For $\gamma_N = 0.5$, $\gamma_P = 2$ and $J_{NR} = 0.033272$, $J_{PR} = 0.001107$

L_S^*	$J_S = [Y/Y_n]_S$	d_N	J_N	L_N^*	d_P	$J_P = [Y/Y_n]_P$	L_P^*
100	1	1.477916	0.033272	21.30949	0	1	100
96	0.900078	1.455056	0.035071	21.96989	0.045720	0.900078	96
92	0.807044	1.431365	0.037037	22.66664	0.093103	0.807044	92
88	0.720653	1.406779	0.039194	23.40321	0.142274	0.720653	88
84	0.640658	1.381229	0.041569	24.18355	0.193374	0.640658	84
80	0.566813	1.354636	0.044194	25.01217	0.246560	0.566813	80
76	0.498872	1.326911	0.047107	25.89425	0.302010	0.498872	76
72	0.436590	1.297953	0.050356	26.83581	0.359926	0.436590	72
68	0.379720	1.267648	0.053995	27.84385	0.420536	0.379720	68
64	0.328017	1.235864	0.058095	28.92658	0.484104	0.328017	64
60	0.281233	1.202449	0.062741	30.09369	0.550933	0.281233	60
56	0.239124	1.167228	0.068041	31.35677	0.621376	0.239124	56
52	0.201443	1.129992	0.074132	32.72971	0.695847	0.201443	52
48	0.167945	1.090499	0.081190	34.22944	0.774834	0.167945	48
44	0.138382	1.048456	0.089443	35.87674	0.858920	0.138382	44
40	0.112510	1.003511	0.099195	37.69753	0.948810	0.112510	40
36	0.090082	0.955234	0.110858	39.72456	1.045364	0.090082	36
32	0.070852	0.903091	0.125000	41.99996	1.149650	0.070852	32
28	0.054574	0.846408	0.142427	44.57898	1.263016	0.054574	28
24	0.041002	0.784319	0.164316	47.53577	1.387194	0.041002	24
20	0.029891	0.715683	0.192450	50.97259	1.524466	0.029891	20
16	0.020993	0.638954	0.229639	55.03515	1.677924	0.020993	16
12	0.014064	0.551966	0.280565	59.93977	1.851900	0.014064	12
8	0.008856	0.451546	0.353553	66.02433	2.052740	0.008856	8
4	0.004428	0.301030	0.500000	76.06926	2.353772	0.004428	4
1	0.001107	0	1	100	2.955832	0.001107	1



Positive Intensity Ratio versus Negative Intensity Ratio
 For $\gamma_N = 0.5$, $\gamma_P = 2$ and $J_{NR} = 0.033272$, $J_{PR} = 0.001107$

The graph shows that the relationship between corresponding positive and negative intensity ratios is still very extreme.

The practice of using a low-gamma negative has been widely used in manufacturing color negative films. Film (like most media) is largely limited by its d_{\max} . When negative film is made with a low gamma (low contrast), the same d_{\max} capability allows it to accurately represent a much wider range of scene intensity ratios, so that it has much wider exposure latitude. When the negative is printed on media with a matched gamma (higher contrast), intensity ratios from the scene will be correctly represented, but of course the positive print will represent only the range of intensity ratios that falls within its d_{\max} capability. A thin "underexposed" negative will print well but so will a dark "overexposed" negative. The quote marks are intended to mean that the negatives appear to be under- or over-exposed but are within the expanded latitude of the color negative film and so not actually under- or over-exposed.

Before we can effectively deal with real world inversion of negative to positive, we must first deal with the much maligned "orange mask."

The Infamous Orange Mask

To judge from advice given on the web, it is widely believed that the problems in dealing with color negatives are due mostly to the color mask, the pronounced orange, blue, or other color cast seen on most types of color negative. In actuality, the mask is equivalent to a uniform colored filter sandwiched with a negative color image that is more accurate than could be obtained otherwise and more accurate than any positive color film. All that is necessary to completely correct for the mask is to add a filter of complimentary color to the mask.

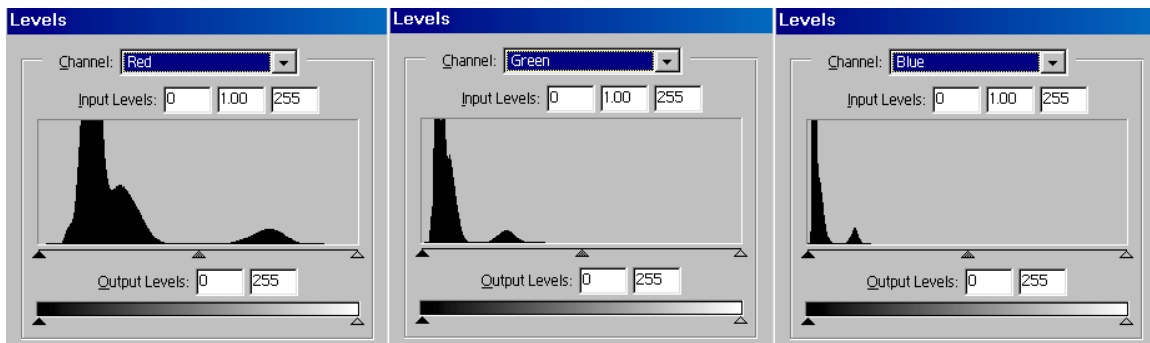
Why the Orange Mask?

The colored dye molecules used to form images in color film are chemically combined with "coupler" molecules so that the combined molecule is colorless and transparent. When the film is developed, light exposed grains of silver halide are converted to silver metal to form the image. Coupled dye molecules adjacent to the sites where silver grains are formed break apart, releasing the dye molecule in its colored form, and the dye molecules form the colored image. In color film the dyes are normally: magenta, which passes red and blue light freely but blocks green light; cyan, which passes green light and blue light freely but blocks red; and yellow, which passes red light and green light freely but blocks blue. The problem is that very few types of dyes can be made to work chemically as required to form the coupled compounds that properly break apart during film development. Consequently, the actual dyes that are used are not at all perfect. For instance, the magenta dye typically blocks some of the blue light it is supposed to pass freely, and the cyan typically blocks some of the blue and green light that it is supposed to pass freely.

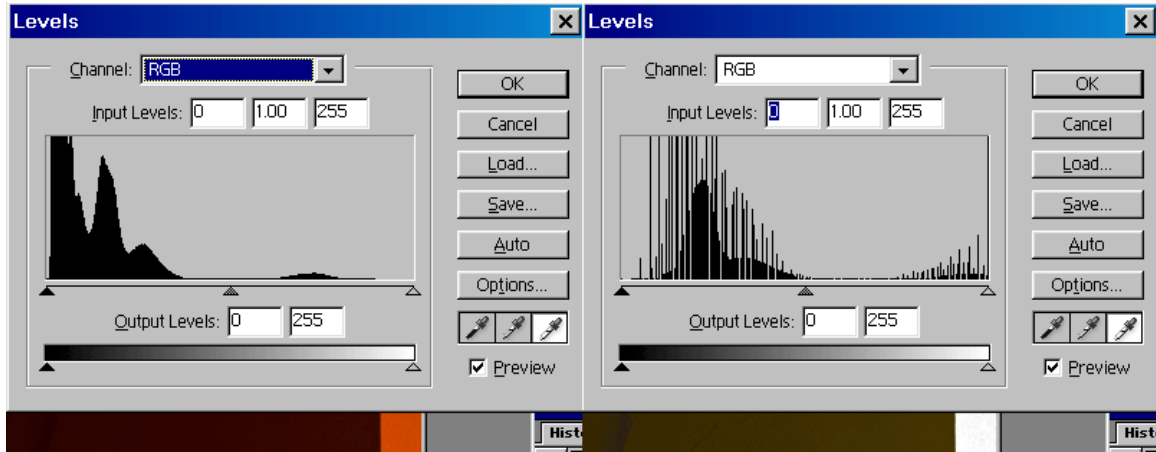
A clever way of improving the dye action was introduced very early in color negative history. Instead of combining the dye molecules with coupler molecules that made them colorless, the dye was combined with a coupler chosen so that the compound had exactly the same light absorption as the *unwanted* absorption of the dye molecule. So, the magenta dye would be combined with a coupler producing a yellow-colored compound so it would absorb some blue light just like the imperfect magenta dye, and the cyan dye would be combined with a coupler producing a pinkish compound that would absorb some blue and green light, just like the imperfect cyan dye. The pinkish and yellow colors together formed the orange mask. If during film development some silver halide was reduced to silver, causing a nearby magenta dye molecule to break loose of its coupler, that meant the yellow coupled dye molecule had been broken and the yellow was gone. So, throughout the image, *either* a magenta dye molecule would be present or the yellow coupled dye molecule remained. Either way, the same amount of blue light would be absorbed, so the effect is like placing a yellow filter over a clear – but more perfect – negative. The same action with the cyan dye places a pinkish filter over the whole image. So, the entire mask effect is a filter that can be cancelled by adding a filter complimentary to the mask color - and brightening the image to compensate for the darkening from the resulting "neutral density" filter.

Dealing with the Orange Mask

This is how "difficult" it is to deal with the infamous "orange mask" in Photoshop. Scan in a color negative, using "raw" (linear) mode, 16 bits/channel, and preferably as a positive, or "slide." The reason for this latter is that most scanners really will deliver a raw linear scan for a slide, but many will attempt their own "orange mask" correction on what they deliver as a "raw linear" scan of a negative. Arrange the negative so that some of the orange mask that is between frames shows at the edge of the scan - between frames rather than at the film edge to avoid possible film edge effects. The following shows the Photoshop Levels histograms for the Red, Green, and Blue channels for a color negative with quite a bit of edge showing:

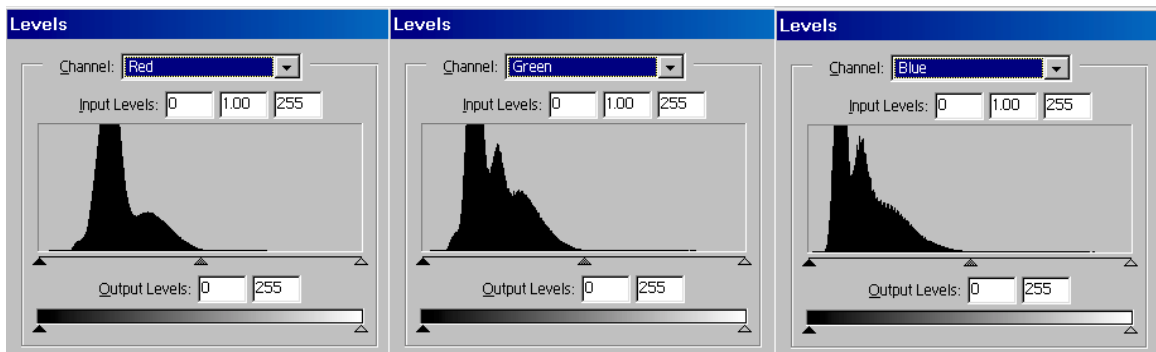


In each case, the smaller lump at the right of the histogram represents the orange mask. Since this mask is the result of placing a (fairly) uniform orange filter over the image, all that needs to be done is to make the mask "clear" or white.



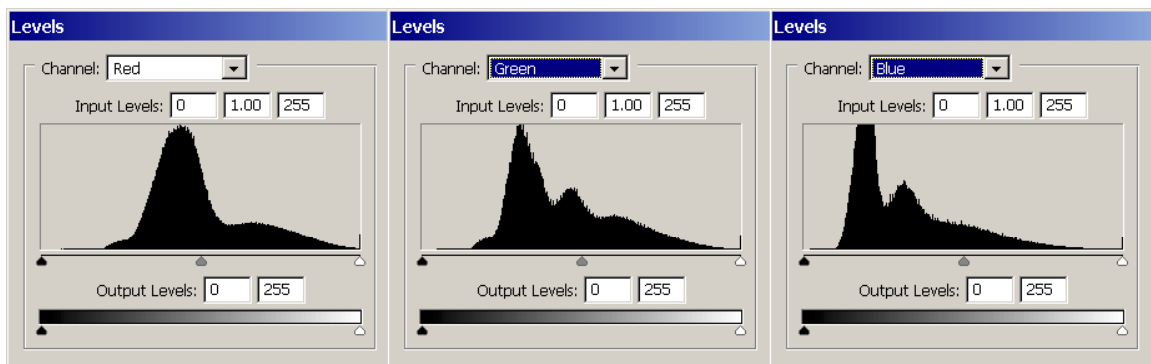
In the above illustration, at left is the Photoshop Levels showing RGB, and below the dialog box a small portion of the color negative image can be seen, with the orange mask at the edge. To correct the orange filter, we click the highlight button, the rightmost eyedropper¹. Click on the orange mask and get a reasonably stable setting. The mask will be mostly white, and the correction for the "orange mask" is completed. This is true whether the orange mask really is orange, or if it is bluish, reddish, or some other color, as is the case for some varieties of color negative. Very easily done. The problems in dealing with color negatives lie elsewhere. Incidentally, the "spiny" histogram on the right is due to the fact that Photoshop uses 8-Bits/Channel values for intermediate histogram calculation even when the image is 16-Bits/Channel. The histogram of the image is smooth when viewed after actual conversion, as shown below.

The result of removing the mask is shown below:



¹ To use the highlight eyedropper properly, be certain this button is set to pure white, (R,G,B) = (255,255,255), which is the default value. This may be checked in the dialog box obtained by double-clicking the highlight button. Also, prior to calling the Levels tool, be certain that the eyedropper selection tool is set to sample a 5x5 area rather than 3x3 or single point. This can be done by clicking the eyedropper in the "Tools" box and setting the "Sample Size:" The exact method varies with the version of Photoshop. Finally, prior to calling up the Levels tool it is wise to zoom in, select a portion of the edge and apply a gaussian filter to it, enough to make it fairly uniform (typically 5 pixels or so), then deselect. Be sure the selected area is well within the uniform edge mask so that non-mask areas do not get blurred into the sample. Sample from this more thoroughly mixed area when correcting for the mask. With very grainy negatives this blurring becomes more necessary.

We have cropped off the now clear mask before taking the Levels histograms - it would have shown up as a half-lump at the very right of each histogram. Note that there is very little on the right hand side of the histograms. This is exactly what should be expected. Earlier we learned that the lighter areas of the negative, to the extent of the rightmost 3/4ths, represent just the very darkest shadows of the positive image, the first four steps of gray in a 100 step uniform scale. Note, however, that this usually long tail is "gathered up" by the inverse relationship and collectively is very important to shadow detail. One problem we have noted with some scanner software we have worked with is that it seems to work on the theory that if there is very little in the negative histogram, it should be trimmed, so some scanners will return an image with the following histograms as a "linear raw" negative!



All we can say is that such scans will lack shadow detail, are not "raw," and may not be linear. Without knowing precisely what has been done, the user is left with an image that is impossible to accurately convert to a positive image. The best one can do with such a scan is to guess.

Real World Film Negative Inversion

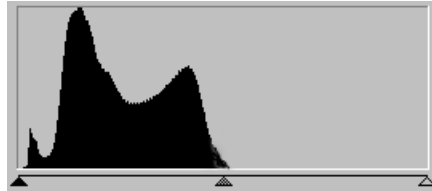
To better understand the properties of the negative-positive film system, reconsider the basic equation:

$$\frac{J_P}{J_{PR}} = \left(\frac{J_{NN}}{J_N} \right)^{\gamma_P}$$

For actual applications there is no need to separate out J_{PR} and J_{NN} , so this can be reduced to:

$$J_P = \frac{K}{J_N^{\gamma_P}}$$

This inverse relationship has a very interesting property relative to the action of filters. Suppose we have a *negative* which has a histogram like the following:

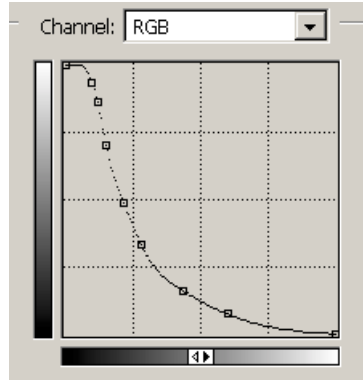


There appears to be nothing on the right hand side of this histogram, so let us move over the right hand Highlights slider until it meets the data. As we have learned previously, the Highlights slider is equivalent to a filter and/or a change in overall illumination and merely multiplies all the J_N 's by a constant factor. Moving that slider is exactly what happened above to the individual color channels to remove the orange mask filter. Thus, the action of moving the highlights slider is exactly the same as changing K in the above equation. Now consider the situation where we are looking at the histogram of a positive image, intending to move the highlights slider to apply a filter and/or a change in overall illumination. This will multiply all the J_P 's by a constant factor. So again the action of moving the highlights slider is to *change the K* in the above equation. Whether we are working with the negative or the positive image, the action of moving the highlights slider is to change K , so filters can be applied to either the positive or the negative as is convenient for the required numerical operations.

Now we are equipped to experiment with a real photographic negative in Photoshop. The image for this experiment should contain a gray scale and the same gray scale should be available for direct comparison. For the purpose of illustration, we used a very old test negative which included the image of a Kodak test booklet which we still have:

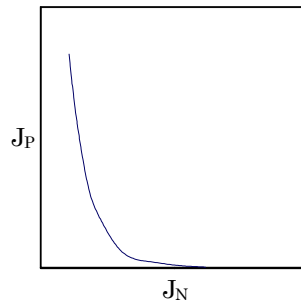


We scanned the negative as a slide in raw mode using a Minolta Dimage Scan Multi PRO film scanner and removed the orange mask as described above. We then used the Photoshop Curves tool to adjust the negative to match the positive. This was done using a visual comparison with the gray scale in the original book and also numerically against a positive scan of the book produced by a Hewlett Packard Scanjet 4670 flatbed scanner. The two methods produced very similar correction curves and the comparison with the scanned image is shown below:



This curve was determined using a PC computer which has an operating display gamma of 2.2. That is, what is stored as the computer image is "gamma adjusted"

$J_P^* = \left[\frac{K}{J_N^{\gamma_P}} \right]^{\gamma_c}$ where $\gamma_c = 0.45$. The real $J_p = [J_P^*]^{\frac{1}{\gamma_c}}$. Correcting the above curve for this gamma gives a curve that looks like:

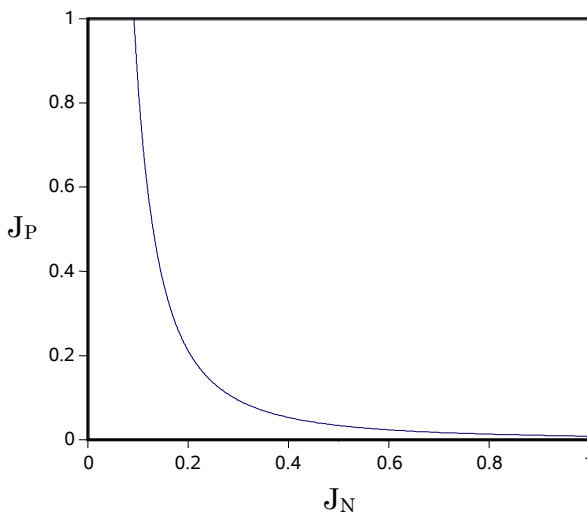


This curve can be closely matched with $\gamma_P = 2.0$ and $K = 0.0085$. Table 3 below represents this real negative to positive inversion in the same format that we used above, going from scene to negative to positive.

Table 3
 From Scene through Negative to Positive
 For $\gamma_N = 0.5$, $\gamma_P = 2$ and $K = 0.0085$

L_S^*	$J_S = [Y/Y_n]_S$	d_N	J_N	L_N^*	d_P	$J_P = [Y/Y_n]_P$	L_P^*
100	1	1.035291	0.092195	36.403602	0	1	100
96	0.900078	1.012431	0.097178	37.331172	0.045720	0.900078	96
92	0.807044	0.988739	0.102627	38.309807	0.093103	0.807044	92
88	0.720653	0.964154	0.108604	39.344373	0.142274	0.720653	88
84	0.640658	0.938604	0.115185	40.440407	0.193374	0.640658	84
80	0.566813	0.912010	0.122459	41.604249	0.246560	0.566813	80
76	0.498872	0.884285	0.130531	42.843192	0.302010	0.498872	76
72	0.436590	0.855328	0.139532	44.165676	0.359926	0.436590	72
68	0.379720	0.825022	0.149616	45.581533	0.420536	0.379720	68
64	0.328017	0.793239	0.160976	47.102294	0.484104	0.328017	64
60	0.281233	0.759824	0.173851	48.741587	0.550933	0.281233	60
56	0.239124	0.724602	0.188537	50.515658	0.621376	0.239124	56
52	0.201443	0.687367	0.205415	52.444047	0.695847	0.201443	52
48	0.167945	0.647874	0.224971	54.550509	0.774834	0.167945	48
44	0.138382	0.605830	0.247839	56.864252	0.858920	0.138382	44
40	0.112510	0.560886	0.274862	59.421667	0.948810	0.112510	40
36	0.090082	0.512609	0.307179	62.268762	1.045364	0.090082	36
32	0.070852	0.460465	0.346365	65.464711	1.149650	0.070852	32
28	0.054574	0.403783	0.394655	69.087115	1.263016	0.054574	28
24	0.041002	0.341694	0.455309	73.240119	1.387194	0.041002	24
20	0.029891	0.273057	0.533265	78.067345	1.524466	0.029891	20
16	0.020993	0.196329	0.636314	83.773487	1.677924	0.020993	16
12	0.014064	0.109341	0.777427	90.662344	1.851900	0.014064	12
8	0.008856	0.008920	0.979669	99.208499	2.052740	0.008856	8
4	0.004428	-0.141600	1.385465	113.31726	2.353772	0.004428	4
1	0.001107	-0.442630	2.770929	146.92953	2.955832	0.001107	1

There is one area in which Table 3 differs from the previous theoretical tables. Full representation of the deepest shadows in the scene requires negative densities that are less than zero, or J_N 's that are greater than one. These are impossible requirements and thus this negative theoretically can not represent the deepest shadow detail - it is underexposed.



Positive Intensity Ratio versus Negative Intensity Ratio
 For $\gamma_N = 0.5$, $\gamma_P = 2$ and $K = 0.0085$

These tests above were done measuring and judging a gray scale in general. It is possible, of course, to do the same thing to the individual primaries, red, green, and blue, finding a separate value of γ_p and K for each. In fact, there sometimes is an intentional manufacturing difference in the γ_p 's for each color. This can be measured if there are gray scale images for the particular film type, and it can be visually estimated if several negatives of the film type are available for test, and it is found that, for example, the shadows run too blue in all of them while the highlights run too yellow. The K values will naturally tend to differ for the three colors, as the relative adjustment of the K 's is the adjustment of overall color balance.

The Limits on K

The above example shows that there are limits on the value of K . When K is outside these limits the negative image is either under- or overexposed and the resulting positive image will lack either shadow or highlight detail. These limits depend upon d_{Pmax} , the maximum density range expressible in the positive image, and d_{Nmax} , the maximum density range of which the negative material is capable.

The operating relationship is:

$$J_P = \frac{K}{J_N^{\gamma_P}}$$

To represent the complete density range of the positive image, at one limit the minimum expressible $J_{Pmin} = 10^{-d_{Pmax}}$ will correspond to a density in the negative of zero, or $J_N = 1$. Thus:

$$K_{max} = J_{Pmin} = 10^{-d_{Pmax}}$$

At the other extreme, the positive density of zero, or $J_P = 1$, must occur at or above a negative density $J_{Nmin} = 10^{-d_{Nmax}}$, so:

$$1 = \frac{K_{min}}{J_{Nmin}^{\gamma_P}}$$

$$K_{min} = J_{Nmin}^{\gamma_P} = 10^{-\gamma_P d_{Nmax}}$$

So for full range coverage

$$10^{-\gamma_P d_{Nmax}} = J_{Nmin}^{\gamma_P} \leq K \leq J_{Pmin} = 10^{-d_{Pmax}}$$

Note that if $\gamma_P = 1$ and $d_{Pmax} = d_{Nmax}$ there is only one value of K that works to give full range coverage. Normally $\gamma_P > 1$, and enough greater than one that even though typically $d_{Pmax} > d_{Nmax}$, there remains a substantial range of values over which K can range and still

give full dynamic range coverage. For example, with $\gamma_P = 2$, if we assume that $d_{P_{\max}} = d_{N_{\max}} = 3$, the approximate range represented by the CIE L^* "lightness" function the mimics the behavior of the human eye, we find that the range is:

$$0.000001 \leq K \leq 0.001$$

which is a substantial range of values.

For some additional perspective on this, the K for both Table 1 and Table 2 was 0.001107, the high end of the range. In those tables the assumed $d_{N_{\max}}$ was 2.95, just slightly shy of the 3 assumed in the above paragraph.

Digital Limitations on K

The negative to positive conversion involves taking a power, taking an inverse of that, and then taking that result to a power. While such operations are trivial and quite precise on the double precision floating point digital arithmetic commonly used on computers, these same operations are problematical when using fixed point arithmetic of fairly low precision, as is commonly the case in working with digital images. Some problems remain even if the actual calculations are done precisely and the low precision fixed point is used only for image storage.

The complete transformation between negative and positive images is:

$$J_P^* = \left[\frac{K}{J_N^{\gamma_P}} \right]^{\gamma_c}$$

where γ_c , the "gamma adjustment," is usually 0.45 for PCs and 0.56 for Macs. When working with negatives, the J 's are typically stored in 16-bit unsigned integers, thus limited to integer values between 0 and $2^{16} - 1$ (= 65535), where the integer is understood to be divided by 65535 to obtain J . Eventually, the J 's are usually converted to 8-bit integers, limited to values between 0 and $2^8 - 1$ (= 255) where the integer is understood to be divided by 255 to obtain J . Two different problems can occur, for values near $J_P^* = 1$ and for values near $J_P^* = 0$. For the purposes here we will define j as the integer part of 65535 J .

First, near $J_P^* = 1$ ($j_P^* = 65535$) the possible values thin out. That is, two successive integer values j_N and $j_N + 1$ will produce widely spaced values j_P^* and $j_P^* - m$. To produce an image without visible "stair steps" in its tonal range, it is desirable that $m \leq 256$ so that in the 8-bit version of J_P^* , all values 0 to 255 will be possible. To determine this, start with

$$J_P^* = \left[\frac{K}{J_N^{\gamma_P}} \right]^{\gamma_c}$$

$$J_P^* = K^{\gamma_c} J_N^{-\gamma_c \gamma_P}$$

The problem will occur where $J_P^* = 1$:

$$1 = K^{\gamma_c} J_N^{-\gamma_c \gamma_P}$$

$$J_N = K^{1/\gamma_P}$$

Find the rate of change in J_P^* with respect to J_N :

$$\frac{dJ_P^*}{dJ_N} = -\gamma_c \gamma_P K^{\gamma_c} J_N^{-\gamma_c \gamma_P - 1}$$

substitute the value of J_N where $J_P^* = 1$, and the absolute value of that derivative must be less than 256; that is, 1 step in 65535 for J_N must lead to no more than 1 step in 255 in J_P^* (which is 256 steps in 65535 in J_P^*):

$$256 > \gamma_c \gamma_P K^{\gamma_c + \frac{-\gamma_c \gamma_P - 1}{\gamma_P}} = \gamma_c \gamma_P K^{-\frac{1}{\gamma_P}}$$

$$\frac{256}{\gamma_c \gamma_P} > K^{-\frac{1}{\gamma_P}}$$

$$\left[\frac{256}{\gamma_c \gamma_P} \right]^{-1} < K^{\frac{1}{\gamma_P}}$$

$$K > \left[\frac{256}{\gamma_c \gamma_P} \right]^{-\gamma_P}$$

For the values used above, $\gamma_P = 2$ and $\gamma_c = 0.45$, this gives $K > 0.0000124$, so that there might be stair-stepping in the highlights for a conversion toward the lower end of the range of K we gave for $\gamma_P = 2$ example above. This situation should be rare, but might occur for a very heavy negative. There appears to be no generalized remedy when it does occur (beyond dithering, which should be a last resort).

Shadow Treatment, Low J_P^*

The second problem occurs near $J_P^* = 0$ ($j_P^* = 0$), where high values of J_N cannot possibly produce the lowest values of J_P^* .

$$J_P^* = \left[\frac{K}{J_N^{\gamma_P}} \right]^{\gamma_c}$$

The highest possible value of J_N is 1, so

$$J_{P\min}^* = K^{\gamma_c}$$

Obviously, that cannot be zero and to have $j_{P\min}^* = 1$ (or $J_{P\min}^* = 1/65535$) would require that for the example $K = (1/65535)^{2.2} = 2.5 \times 10^{-11}$, which is way out of the practical range. For practical purposes, as above, it would be desirable to have $j_{P\min}^* = 257$ (or $J_{P\min}^* = 1/255$). Then $K = (1/255)^{2.2} = 0.000005$, which is still so low that it excludes most of the practical range for the example case. In fact, we find that at the upper end of the range in the example case $K = 0.001$ which leads to $J_{P\min}^* = 11/255$, so that there can be no detail in the deep shadows represented by the first 10 counts in the 0 - 255 range of an 8-bit image with a conversion K that large. In our underexposed test case, we found that $K = 0.0085$, which leads to $J_{P\min}^* = 30/255$, even worse.

This can present a serious problem, sometimes making it impossible to get deep shadows from a negative. It should be pointed out in passing that most of this problem is due to the "gamma correction," that is, $J_p^* = J_p^{\gamma_c}$. In CFS-243, *Maintaining Color Integrity in Digital Photography* we mentioned that the "Levels" gamma correction in Photoshop is incorrect for gammas around the value of $\gamma_c = 0.45$ (or 2.2 in Photoshop convention), and we believe this error is really intended to be a makeshift correction for the problem we are dealing with here.

While there is no completely accurate solution to this problem when it occurs, there is a practical work-around. For the work-around, we will adjust J_p instead of J_p^* in order to lessen the distortion of the colors in the shadows. This color distortion effect is explained fully in CFS-243.

We can adjust J_p so that the minimum will be zero:

$$J_{P\text{adj}}^* = \left[\frac{J_p - J_{P\min}}{1 - J_{P\min}} \right]^{\gamma_c}$$

Note that to use this form we must use $J_{P\min}$ instead of $J_{P\min}^*$ in order to properly match J_p . Thus if we choose to take $J_{P\min}$ where $J_N = 1$, then $J_{P\min} = K$. In practical terms, it often makes more sense to find from histogram data the $J_{N\max}$ as the largest value for which there is a significant count of pixels. (Alternatively, above which the count of pixels is small enough to be ignored.) In doing this, it is well to remember that in this lightest area of the negative, a long range of brightnesses will represent just a short range of brightnesses in the darkest shadows of the positive, so a long span of nearly empty histogram cells is important to the shadow quality of the result. In any event, the relationship then becomes:

$$J_{P\min} = \frac{K}{J_{N\max}^{\gamma_c}}$$

The key to the success of this operation is to perform all the intermediate calculations in full double precision floating point arithmetic on the computer and never store an intermediate integer value of j_p . That is, to perform the calculation as:

$$J_{Padj}^* = \left[\frac{\frac{K}{J_N^{\gamma_p}} - \frac{K}{J_{Nmax}^{\gamma_p}}}{1 - \frac{K}{J_{Nmax}^{\gamma_p}}} \right]^{\gamma_c}$$

Note that in terms of computer operations, once K , γ_p , γ_c , and J_{Nmax} are known, a table corresponding to all 65536 possible values of J_N can be formed and filled with the double precision computed values of J_{Padj}^* , so that the actual conversion of an image file can be done using the resulting lookup table.

Alternatively, an adjustment can be made that may be more accurate in terms of color integrity (although preliminary testing suggests that the changes in color would be essentially invisible in most cases). First, for the largest component, "m" of R, G, B, (that is, the smallest J_{Nq} where "q" is the color index) calculate:

$$J_{Padj} = \left[\frac{\frac{K}{J_{Nm}^{\gamma_p}} - \frac{K}{J_{Nmmax}^{\gamma_p}}}{1 - \frac{K}{J_{Nmmax}^{\gamma_p}}} \right]$$

Note that this is *not* gamma adjusted. Then, also calculate the uncorrected values for each component:

$$J_{Pq} = \frac{K}{J_{Nq}^{\gamma_p}}$$

Note that although substitution tables are no longer possible, a set of four tables of double precision calculated values can be created and used. So, for the predominant primary m , and for color q , use:

$$J_{Pq}^* = \left[J_{Padj} \frac{J_{Pq}}{J_{Pm}} \right]^{\gamma_c}$$

This calculation preserves the ratios between the color components as they should be in the linear positive image, and thus is more accurate but slower to calculate. The entire calculation should be kept in double precision floating point until J_p^* for the final image is stored.

Highlight Treatment, Low J_N

On most of the J_P versus J_N curves we have shown, at low J_N , there is a considerable gap between the curve and the axis even when it reaches $J_P = 1$. It appears like this gap might represent considerable loss in highlight detail, but that is mostly an illusion. This area of the curve is where the negative is at its greatest density, and that will normally not be full-scale of the scanner. Thus no pixels will have intensities for the lowest J_N values, and the "missing" values will not be a problem at all. The best approach is to find $J_{N\min}$ from histogram data, the lowest J_N for which there is a significant pixel count. The value of K can then be set to that $J_P = 1$ at this J_N .

$$K = J_{N\min}^{\gamma_P}$$

This is the best starting value of K , and the minimum value which will retain all the shadow detail.

Final Summary

The infamous "orange mask" that is so often blamed as the main source of difficulty in dealing with film negatives is actually very easily handled.

The relationship between the positive and negative images is an inverse one:

$$J_P^* = \left[\frac{K}{J_N^{\gamma_P}} \right]^{\gamma_c}$$

Here γ_c is the "gamma adjustment" that is typically applied to images stored in PCs, and the star notation means that the positive image J_P^* has been gamma adjusted. The real problem in dealing with negatives is in determining γ_P and K . γ_P should be nearly the same for all properly processed color negatives of a particular film type, although γ_P may be different for each of the three primary colors. It can be visually estimated by examining test images of a negative inverted to a positive, particularly if several tests are made and compared, but is more accurately estimated by comparison of an inverted image to a known final image. The correct estimation of γ_P is important to image color integrity, since this γ_P should be "matched" to the γ_N of the negative to get the colors correct.

The K regulates the overall lightness/darkness of the image. The best estimation or selection of K varies from negative to negative, even of the same negative type. Color negatives typically are designed with $\gamma_N < 1$, which gives them wider exposure latitude, and K is a measure of exposure. A starting value for K is $K = J_{N\min}^{\gamma_P}$, and there are numerous constraints, adjustments, and calculation methods for K that arise from physical and numerical considerations.

Since color negatives have a wide latitude, it will sometimes be that case that final positive images are best made as composites of two images derived using different K values. An example would be a landscape in which the sky is bright enough to lack detail. The second image would be darker, showing the sky detail. This is similar to the traditional photographic technique of "burning in" the sky working with a color negative and an enlarger. Much more detail is potentially available than would be available from digitally burning in the image from a single K conversion.

For best results, separate values of γ_p should be obtained for each of red, blue, and green, for a particular film type, estimated from gray scale images if possible. Differing gammas for the separate colors can be visually estimated if several negatives of the film type are available for test, and it is found that, for example, the shadows run too blue in all of them while the highlights run too yellow. The K values for the three primaries will naturally tend to differ for the three colors, as the relative adjustment of the K 's is the adjustment of color balance.

The method of conversion of negative to positive that is described here, including the use of the correct gammas, is necessary to preserve the color integrity of the original negative (see CFS-243, *Maintaining Color Integrity in Digital Photography* for further elaboration of this problem). The result may have a tonal scale that needs improvement; often it will be too dark. A "curves" adjustment in 16-Bits/Channel mode could compress the highlights without losing detail. Since at this point the colors should be correct or at worst out of color balance in a way that can be adjusted by a color correction filter, the best method is to convert the image to CIE Lab color and adjust the Lightness curve in that mode to place the tonal values where they are wanted. The image can then be converted back to RGB for further adjustment, including final color balance.