CFSYSTEMS

Improving Highlight and Shadow Detail in Digital Imaging Systems

CFS-211b

May 22, 2002

Revised for web, December, 2002

Comments added January, May, and July, 2004

Personal Notes from the Author

12/2002

This is based somewhat on speculation because hard data on specific digital camera designs are not easy to come by. However, I noticed that when I saw a professional picture with blocked highlights and/or muddy shadows it had almost always been taken with a digital camera. That prompted the investigation. After recognizing the following, it has been easy to spot the defects when they occur and to identify the picture as taken with a digital camera, so I have reasonable confidence that my speculations are either correct or nearly so. Please note that the effect does *not* normally apply to pictures taken with silver-based media and scanned into a computer for digital treatment and representation. I apologize for the mathematics involved, but it is difficult to explain the concepts otherwise. Those of you who are mathematically inclined may initially believe that cameras with higher bit depths (12 or 16 bit pixels) automatically solve this problem. With deeper reflection you will realize that is not true; such cameras still have the problem, although the added bit depth will make them easier to fix.

To clarify my position on this, I am posting the information with no plan to patent this idea; it being my method, no one else has the right to patent it, so it will be freely available. However, in the past I have had patentable ideas stolen and patented by someone else (which is highly illegal). I have taken appropriate actions establishing my priority so that any such action will be detected, thwarted, and, I hope, prosecuted.

1/2004

It appears that Fuji with its "4th Generation Super CCD" may be the first to address the dynamic range problem described in this paper. This is the same device that has led to Fuji claiming a 3.1 megapixel camera that produces a 6 megapixel file. That statement conveys the idea that the camera produces definition similar to a 6 megapixel camera without their actually saying it does, and I do not believe that is accidental. For some reason Fuji seems not to have made much of the supposed increased dynamic range - I had difficulty locating anything about it other than in a brochure describing the CCD itself. Fuji claims this device has "four times greater dynamic range than offered by the same size of the preceding <CCD version>." Whatever that may mean; dynamic range is usually given in density units and a range of 12d instead of 3d is just a *little* too hard to believe. Density is a *log* quantity and in *log* terms, "four times" might mean going from 3d to 3.6d, less overwhelming but enough to help. From the description it appears that the device may have a response similar to the S-curve of Diagram 1 below, with the curve approximated by three straight line segments. I have not tested this device but tests I have seen seem to indicate that there is more substance to the dynamic range claim than to the increased definition pseudo-claim.

5/2004

I have bought a Fuji FinePIX f700 with the 4th Generation Super CCD to test. The results are mixed. As delivered, with the software provided, the camera delivers images with a modest improvement in highlight detail over other digital cameras I have tried. It delivers a 6MB image file that may be fairly compared in image quality to the 3 or 4 MB image files delivered by other digital cameras. It often is not as sharp or detailed as a 4 MB Olympus interpolated to 6 MB in Photoshop for comparison. This is complicated by the fact that the Fuji delivers jpeg files of only moderate jpeg quality. The tiff files converted from "raw" image files are sharper but also unnecessarily larger, and still often not as sharp as the 4 MB Olympus.

That is the bad news. The good news is that the Super CCD itself appears to have all the capability of an extended dynamic range of around four stops beyond a standard digital camera - approaching the range of color negative film. It also has the potential for much better resolution than Fuji delivers. All of the problems are in how Fuji treats the image as it is converted from the sensor readings to jpeg or tiff. I have directly examined and tested the content of Fuji raw files before they are converted to tiff and the dynamic range information is all present in the raw file. Further Fuji processing to tiff or jpeg simply throws it away! As to resolution, for those who are unaware, *all* digital cameras interpolate to get their claimed resolution. The differences in image quality and sharpness between different manufacturers of, say, 6 megapixel cameras are due almost entirely to how well the interpolation is done. Some manufacturers are better at interpolating than others and the difference can be quite visible. Due to the structure of the Super CCD, interpolation is more complicated than for standard sensors, but Fuji could do much better.

Because I am interested in the dynamic range and I have had considerable past experience in developing better mathematical treatments for problems similar to the interpolation, I will be writing my own raw file converter when I get a chance. Very likely, at some future date a version of it will be made available on this web site for other people to test.

7/2004

As part of ongoing research, I have tested the Photoshop CS "RAW" option for the Fuji f700. The one place in numerical treatment where Adobe consistently has really has first class code is in interpolation, and that really shows in this case. The images produced by Photoshop RAW are sharper, actually have somewhat better resolution, and have much less color fringing than the images Fuji software produces. Photoshop also appears to take a little more advantage of the dynamic range. I believe there is much more that can be done with the dynamic range and as good a job as Photoshop does, I think there may be a trick or two to gain still more in resolution.

My free programming time recently has been devoted to producing a photographic negative inverter as a Photoshop plug-in. For those who do not realize this is a lot more involved that the Photoshop Invert command (which does that job incorrectly), see CFS-244, Negative to Positive. It turns out that the curve can be very usefully applied in inverting negatives. That code is undergoing some final tests and improvements and will be posted on the web soon. It is likely that I will use that plug-in as a bootstrap to help implementing the Fuji code as a Photoshop plug-in also, instead of continuing to develop the stand-alone program which I had used for early testing.

David Dunthorn
www.c-f-systems.com

In silver-based photography the density, d, of an image, such as a photographic negative or print, is a measure of the blackness. In viewing a negative, for any particular spot d is defined in terms of I_i the intensity of the light incident on the negative and I_t, the light transmitted through that spot on the negative such that $d = \log_{10} \frac{I_i}{I_t}$. An exposure E is required to obtain a particular d at a spot on the negative. Exposure is typically I_s, an intensity of light coming through the camera lens from a point on the subject, times the exposure time, t, governed by the shutter speed of the camera. The response of a typical photographic negative material is shown in the following diagram, commonly called the S-curve or the H&D curve. Although the curve normally has an S-shape similar to the one shown, the exact curve and relationship is characteristic of and specific to the type of negative material when processed according to a specific method.

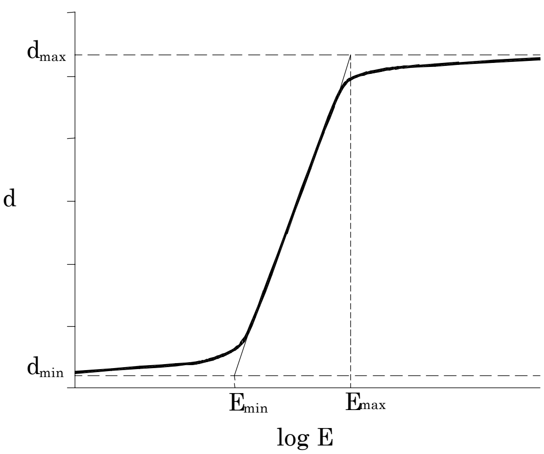

Diagram 1
S-Curve for a Negative Photographic Film

Note the long, nearly straight line section in the center of the S-curve. That portion of the curve has traditionally been used to characterize the behavior of the specific film, using the linear relationship:

$$d = \gamma \log_{10} \frac{E}{E_{\min}} \tag{1}$$

where γ (gamma) is the *slope* of the straight section – a measure of the steepness of the straight section – and is related to the contrast of the photographic image. E_{min} is called the *reference exposure* (often noted as E_{ref}), and represents the exposure at which the straight-line approximation of Equation (1) predicts a density of zero. For a real photographic film the density never goes to zero – no film is perfectly transparent – and in fact the density asymptotically approaches $d_{\min}$ as the exposure becomes small. Likewise, the photographic film has a maximum density, $d_{\max}$, which cannot be exceeded, and this value is approached asymptotically at very high exposures. (Technically, at extremely high exposures the density actually starts to decrease with increasing exposure, a phenomenon called *solarization*, but this condition is rarely encountered in normal photographic practice.)

The slope γ and the straight-line portion of the curve which it defines has been regarded as a fundamental property of silver-based photographic films, and for many films a goal has been to make the straight-line section as straight and long as possible. The physical nature of the film has made it impossible to actually achieve the straight line shown in Diagram 1 and the goal of achieving that straight line something of a Holy Grail of silver-based photography.

When digital cameras began to be designed, it was thought that the Holy Grail was suddenly achievable. It became just a matter of electronic design and programming. In general, digital cameras are considered to be *perfect*, or nearly so. Digital cameras typically produce positive rather than negative images, so Diagram 1 and Equation (1) must be revised to the positive equivalent for comparison, Diagram 2 and Equation (2):

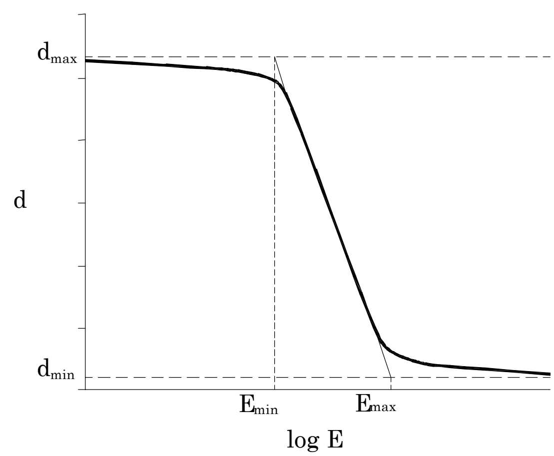

Diagram 2
S-Curve for a Positive Photographic Film

$$d = -\gamma \log_{10} \frac{E}{E_{max}} \qquad (2)$$

Digital cameras typically are designed to behave exactly according to the linear Equation (2) from E_{min} to E_{max}, or very nearly so. Below E_{min} density is set to d_{max} and above E_{max} density is set to zero.

What has not been appreciated is that outside the linear region this actually is far from *perfect* and in fact, is not even *good*. Articles in the journals for professional photographers often complain of the lack of shadow detail or the lack of highlight detail in images from digital cameras. Shadow and highlight detail are very necessary in high-quality photography.

Diagram 2 shows that for a real photographic film (the S-curve) in the darkest areas (shadows) and the brightest areas (highlights) the density is not rendered correctly, at least according to Equation (2). Exposures of somewhat more than $E_{\min}$ and below are rendered lighter than they should be according to Equation (2) and exposures of slightly less than $E_{\max}$ and above are rendered darker than they should be. However, it is very important to note that although the densities are not rendered correctly in these regions, the densities for exposures below $E_{\min}$ get continuously darker as the exposure is less and for exposures above $E_{\max}$, the densities get progressively lighter as the exposure is greater.

For the typical digital camera, all exposures below $E_{\min}$ yield $d_{\max}$. Note that this is not correct according to the linear Equation (2), which says that exposures below $E_{\min}$ should yield successively greater densities *above* $d_{\max}$. Similarly, all exposures above $E_{\max}$ yield $d = 0$, while Equation (2) calls for the density to become progressively more negative.

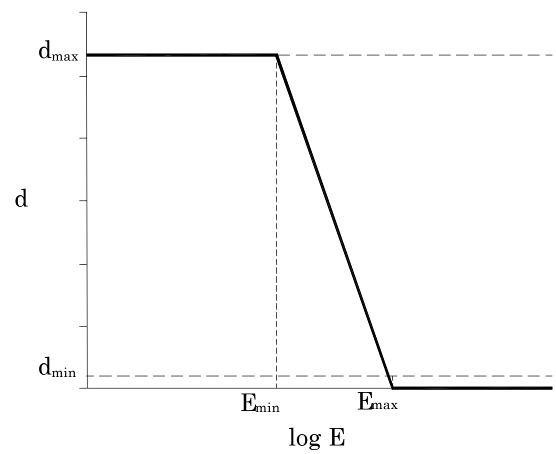

Diagram 3
"Curve" for a Typical Digital Camera

Diagram 3 shows the density – exposure relationship for a typical digital camera, as described above. The behavior between $E_{\min}$ and $E_{\max}$ is perfect, but this leads to complete failure below $E_{\min}$, where there is no difference in the density produced by successively lower exposures. This causes completely blocked shadows, with no detail in for any areas producing exposures below $E_{\min}$. Similarly, there is complete failure above $E_{\max}$, causing completely blocked highlights with no detail in any areas producing exposures above $E_{\max}$. Photographic film, with its natural S-curve, does distinguish exposures well below $E_{\min}$ and well above $E_{\max}$, producing a different density for each different exposure. The shadow detail and highlight detail are somewhat compressed, but *the detail is present.*

Patent Terminology

Putting this into the terminology required by the world of patents, this invention has been devised to improve the highlight and/or shadow detail in images produced by digital imaging systems. Such systems include, but are not limited to, digital cameras, digital camera backs, and digital scanners. To achieve this we use a digital imaging system in which the electronic circuitry and the programming of the image recording components have been altered to produce a d versus $\log E$ curve in which the densities d monotonically decrease for increasing exposure E over a wider range of E than the $E_{\min}$ to $E_{\max}$ range predicted from the mid-range γ in use by the imaging system, and in general over a wider range of exposure than is the case for current digital imaging systems for the specific γ being used. In one expression, the d versus $\log E$ curves may be similar to the S-curves for silver-based films now in use or similar to positive versions of those curves.

One means by which this can be implemented is to design a digital imaging device in which the imaging sensor and associated electronics are arranged to produce an image in which gamma is much smaller than is the current norm but which still has a response described approximately by Equation (2). This will necessarily make the range between $E_{\min}$ and $E_{\max}$ much larger than for a normal digital imaging system. The response can then be remapped to an S-curve similar to that shown in Diagram 2, where the central slope is selected to match a gamma in which will giving the mid-tones of the image an accurate or other desired representation. It is desirable to capture the original response in electronic and/or digital form of sufficient precision to allow the mapping to the S-curve to the full digital output precision of the imaging device.

The above is described in terms of a digital imaging device producing a positive image. The same method can be used for devices producing negative images by targeting a density monotonically increasing for increasing exposure.