Generating Scene-Referred Data in a Digital Still Camera

Lawrence R. Hanlon,
Hewlett Packard Laboratories,
Richard L. Baer,
Agilent Laboratories,
Jack Holm,
Paul M. Hubel,
Hewlett Packard Laboratories,
Palo Alto, CA 94304, USA

Abstract

Digital still cameras designed for the consumer market have traditionally employed image processing techniques derived from practices widely used in the video industry. In this approach, designers are constrained to manipulating the image data alone as they attempt to produce pleasing photographs. In contrast, the HP PhotoSmart 618 and 912 digital cameras and their Pentax counterparts employ a new approach to digital photographic reproduction. We use not only the image data, but also physical data from camera calibration measurements and from the statistics of natural scenes. In these new cameras, physically meaningful, scene-referred image data is generated, thus enabling the use of a new class of pictorial processing algorithms for creating the output-referred image data that is shared and printed. As a result, pleasing photographs can be produced from a broader range of photographic situations than the traditional video techniques allow. The calibration measurements and the process of using the measurement data to generate the scene-referred data will be described.

Introduction

From a technical perspective, digital photography requires implementation of a digital color reproduction process that produces pleasing pictures. The entire imaging chain from capture to the end uses of an image must be considered, as each step can impact the final result. End-to-end imaging system design, as opposed to concentration on just a single component of the system, is therefore fundamentally important. In developing the new HP Imaging Technology introduced in the HP PhotoSmart 618 and 912, we have not only started from a system-level perspective, but also introduced new component algorithms that together enable capabilities not previously seen in a digital camera. These include the ability to individually adjust the reproduction of each individual image, automatically, in the camera. This adjustment attempts to optimize the reproduction based on an analysis of the image data, consideration of camera characterization information and settings, and user preference selections. Consequently, the HP 618 and 912 cameras have the ability to gracefully reproduce high and low key scenes as well as high and low contrast ones; and the ability to achieve more accurate color balancing, especially under artificial lights.

These photographic benefits are the result of:
- comprehensive system design,
- extensive analysis of image sensors to discover their true behaviors,
- a new approach to pictorial digital color reproduction,
- new image processing algorithms, and
- the synthesis of the imaging solution into an efficient processing pipeline suitable for a battery-powered imaging appliance.

The approach to pictorial digital color reproduction embodied in these cameras has been presented previously [1]. It represents a major paradigm shift in that the image processing is no longer being applied to a "bag of bits". Knowledge of the actual physical significance of the image data enables using "physically aware" algorithms. This makes possible the use of more aggressive, adaptive, and even non-convergent algorithms. Experience in pictorial color reproduction points to the advantage of using such algorithms to produce preferred results, but implementation has been hindered by the lack of image data with a known physical meaning. Without such image data, these less constrained algorithms can produce inappropriate results. Therefore, this paper will concentrate on the issues involved in generating physically meaningful scene data in the camera.

Background

For consumer digital still cameras, it has been traditional to employ picture processing techniques very similar to those used in video products such as camcorders [2]. While this approach has the advantage of leveraging known solutions to a new class of products, it also brings performance limitations because of design tradeoffs previously made to optimize the solutions for video. It is well known that still photography is more demanding of intrinsic image quality than is video because the images are not moving, the predominant medium for output is a reflection print, and because of long-established customer expectations.

At HP we set out to produce a digital still camera pictorial processing solution that was intrinsically capable of taking full advantage of the power of *digital* color reproduction solutions, while meeting the constraints of a battery-powered portable imaging appliance. We also required that the processing solution implemented be extensible both to high-end applications as well as to low cost ones. In our approach, we chose to base the pictorial processing on physically meaningful image data in order to enable the use of powerful new component algorithms in the processing pipeline. Estimation of the physical parameters of the original scene is then required, as is calibration of the capture subsystem of the camera, thereby making it a calibrated transducer. While this requires additional effort during the product design phase, the resulting photographic benefits to the user include a high percentage of pleasing pictures created automatically from a wide range of photographic situations, as summarized above in the introduction. In the PhotoSmart 912, for example, 12 bit per channel scene data is color rendered to 8 bit output images. The rendering of each image can be uniquely tailored because the absolute scene radiances are known. Color balancing and rendering are accomplished using a collection of internally developed and licensed proprietary algorithms [3-5].

As mentioned above, our approach is therefore very different from the traditional one, in which the digital picture data is much like a “bag of bits” to which various image processing algorithms are applied to improve the pictures. Without a link to the physical significance of the bits, these algorithms are in a sense forced to operate blindly. While there are many clever techniques that have been developed, the range of photographic situations from which truly pleasing results are typically obtained is significantly more limited than the range of photographic situations commonly encountered. Indeed, even with professional digital cameras, the prevailing paradigm is that the captured picture data is just a starting point from which a human then makes modifications to produce the desired pictorial result. In contrast, our objective has been to have the camera automatically deliver pictorially pleasing data. Previous research indicates that, in principle, the potential of digital photographic capture exceeds that of film capture [6-10]

Calibration of the capture subsystem involves quantifying the behavior of the spectral sensitivities, the OECF (opto-electronic conversion function, ie the transfer function from photons to digital counts), and properly generating the color transforms from sensor RGB to a linear scene space. Since the sensors used in these cameras typically utilize a mosaic color filter pattern to reduce component costs, it is necessary to generate fully populated planes of color data from the spatially subsampled image data actually measured. We utilize a new reconstruction technique [11] which requires knowledge of the point spread function (PSF) and considers the noise present, so measurement of the PSF, line spread function (LSF), or spatial frequency response (SFR) is also part of the calibration process.

We next describe the calibration measurement procedures and techniques used to ensure that the capture subsystem delivers physically valid scene data to the pictorial processing pipeline.

Measurements

a) Measurement apparatus

The system that is used to characterize linearity and spectral sensitivity consists of a halogen lamp, a lamp power supply, a monochrometer, an integrating sphere and a power meter. Two filter wheels are mounted on the input side of the monochrometer, one of which contains ND filters and the other of which contains order-sorting filters. The monochrometer contains motorized input and output slits, a grating turret upon which a grating and a mirror are mounted, and an input shutter. The integrating sphere presents the camera with diffuse uniform illumination. A silicon diode detector located on the auxiliary port of the sphere is used to monitor the illumination level. The photon flux can be calculated from the detector current (measured by the power meter), the wavelength, and the known quantum efficiency of the detector.

These basic measurements cannot be performed unless raw data can be extracted from the camera. We utilized a special diagnostic program that extracts raw 12-bit image data through the USB port of the camera. The images were then decomposed into four separate color planes *(RGGB)*, and the average values of the pixels in the center of the image were computed. The program also controlled the monochrometer and interrogated the power meter through a GPIB interface.

In the case of PSF measurements, the camera was used to collect images of a special PSF test chart. The raw 12-bit data was saved on the camera’s compact flash card for subsequent analysis.

b) Point Spread Function

The point spread function (PSF) is determined by collecting an image of a slanted edge and analyzing it. The color channels are balanced to the same level and defective pixels are corrected before the analysis begins. The image can also be corrected for vertical and horizontal shading. The slant makes it possible to sample the response across the edge at a much higher resolution than the pixel pitch. The center of the edge is determined for each horizontal line by computing the centroid of the first derivative. The linear function of row number that best fits the set of centroid values is then used to re-register all the rows around the edge transition point. This super-resolution data is then filtered and differentiated to yield the PSF. We use a transmission target with a single black bar in a clear field, mounted on a white plastic diffuser that is illuminated from the rear by a tungsten lamp. This method is described in the ISO 12233 standard [12]. The signal-to-noise ratio of the results can be improved by assuming that the functional form of the PSF is either a Gaussian peak, or (in the case where a birefringent blur filter is used) the sum of two offset Gaussian peaks. Since these functions are described by only a few parameters, consistent results can be obtained in spite of the presence of noise (however if the actual form of the PSF is not Gaussian, some accuracy is lost).

c) Linearity Measurement

A front surface mirror is used in place of the grating in the monochrometer to obtain white light for linearity evaluation. The camera is placed in front of the integrating sphere, and the output slit is adjusted to control the illumination intensity. If the optical elements in the monochrometer are adjusted correctly, the output will be almost a linear function of slit width. If more dynamic range is required, a spectrally non-selective neutral density filter can be inserted between the lamp and the monochrometer to obtain lower light levels. The exposure period of the camera is fixed in this experiment. The average values of the different color planes are measured as a function of slit width. These values can be tabulated against the reference detector current in order to determine the opto-electronic response functions (OECF) for each color plane. This linearity measurement method is consistent with the alternative focal plane OECF measurement method from ISO 14524 [13].

$$OECF = \langle C_k(w) \rangle / I_{ref}(w) \qquad (1)$$

In this equation, $\langle C_k \rangle$ is the average value of the pixels in the k^{th} color channel, I_{ref} is the reference detector current and w is the slit width.

d) Spectral Sensitivity Measurement

The quantum efficiency is defined as the efficiency of conversion from incident photons to detected electrons. The *relative QE* of a camera is determined by dividing the camera's raw output signal by the current induced in the reference photodiode (with the appropriate correction factor) at a number of different wavelengths.

$$Relative\ QE(\lambda) = \langle C_k(\lambda) \rangle / (I_{ref}(\lambda) / QE_{ref}(\lambda)) \qquad (2)$$

In this equation, QE_{ref} is the quantum efficiency of the reference detector at the operating wavelength. In this measurement the camera's output signal must never exceed its linear range. One can perform these measurements at low signal levels without losing accuracy because many pixels are averaged in order to compute the mean responses of the color channels.

The spectral sensitivity is a measure of CCD response as a function of incident optical power, rather than of incident photon flux. The photon energy is $E = h\ c\ /\ \lambda$, where h is Plank's constant, c is the speed of light, and λ is the wavelength. Since the energy varies as the inverse of the wavelength, the *Relative Spectral Sensitivity* can be determined by multiplying the *QE* by the wavelength.

$$Relative\ Spectral\ Sensitivity\ S_k(\lambda) = \lambda \langle C_k(\lambda) \rangle / (I_{ref}(\lambda) / QE_{ref}(\lambda)) \qquad (3)$$

Spectral sensitivity measurements and their application are described in ISO WD 17321.4 [14].

e) Testing of camera calibration

The OECF measurement in equation (1) measures the individual channel responses compared to the input stimuli. Additional measurements that are important for confirming the accuracy of the spectral sensitivity measurement depend on the crosstalk between channels. This crosstalk could arise in either the optical configuration of the color filter array and the photosensitive area of the CCD, or in the electrical properties of the CCD and/or the A/D converter. Some crosstalk due to CCD clocking errors can show up as differences between two channels that are expected to be the same (such as from the two green filters in a Bayer pattern).

After this check, an additional final confirmation of accurate data involves measuring the spectral power distribution of light reflected off a standard target, and comparing these measurements with the signals the camera creates from the target. The spectral power distributions from the target surfaces are measured with a tele-spectrophotometer. The camera responses are predicted by integrating these distributions with the measured camera spectral sensitivities. Camera images of the targets (this must be using raw camera data) are then decomposed into separate color planes, and the average values of the pixels in the center of each patch in the target are computed. By comparing the results on a two-dimensional chromaticity plot, we are able to evaluate the accuracy of the signals produced by the camera.

By comparing the color errors between the saturated and desaturated colors we can get some indication of the amount of crosstalk between channels in the camera. Saturated colors are obtained when there are large differences between the *RGB* values. A saturated green signal is one where the green channel is stronger than the red and blue channels (at least it is stronger after correction for the illuminant). In this case, crosstalk will cause a percentage of the signal from the green channel to leak into some or all of the other channels, pulling the chromaticity towards white. In contrast, in pastel colors the strength of the *RGB* values is similar, and the crosstalk of the system will have little or no effect. For example, if a 12 bit A/D converter is specified to have 0.5% crosstalk error, then a saturated green having *RGB* values of 10, 4090, 10 could be misrepresented as 30, 4030, 30. This corresponds to a significant color change.

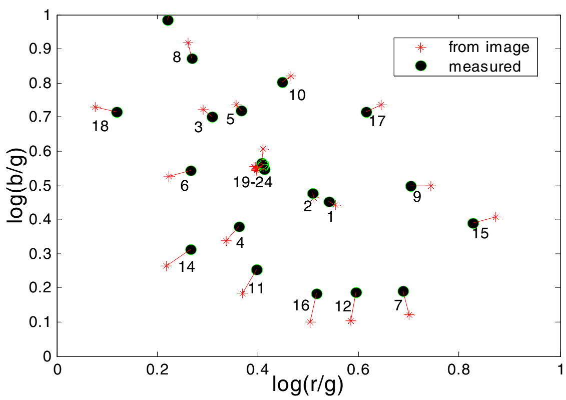

Figure 1. Chromaticity diagram showing desaturation of the more saturated colors caused by crosstalk in camera system. (Asterisks are chromaticities calculated from digital camera image data and dots are chromaticities predicted from measurements of light reflected off surfaces and camera spectral sensitivities.)

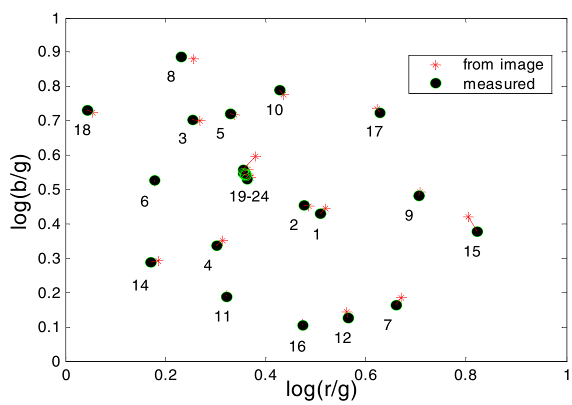

Figure 2. Chromaticity diagram showing more accurate prediction of image colors calculated from measured data. (Asterisks are chromaticities calculated from digital camera image data and dots are chromaticities predicted from measurements of light reflected off surfaces and camera spectral sensitivities.)

Figure 1 shows a comparison of colors measured from a Macbeth ColorChecker® Chart with the colors computed from the camera data. In this case a timing error caused significant crosstalk that manifests itself as desaturation of the chart's most saturated patches. After fixing the timing, the results shown in Figure 2 were obtained. This result shows much more evenly distributed color errors.

Generating Scene Data

Having successfully completed the calibration of a digital still camera, we can now proceed to the algorithms and transformation methods for estimating scene data. Although treated briefly in this paper, camera measurement and calibration is not only extremely time consuming and difficult, but also fundamentally necessary in developing a thorough understanding of the digital photographic process.

a) White-point balance

The first step in our computation of scene-referred data is the determination of the adopted white in the scene. Both our methods of image demosaicing and of determining linear transformations into a color space depend upon knowledge of the scene adopted white.

The method used for white-balance in the new HP and Pentax digital cameras is known as 'Color by Correlation' [3,4]. In this method, we correlate the colors seen by the camera with colors that are physically realizable under a set of plausible light sources. Then the results of this test are correlated with statistical probabilities of the light sources having the same chromaticity as the adopted white balance of the scene.

Having accurate measurements of the spectral sensitivities, $S_k(\lambda)$, of the digital camera system, we can pre-compute correlation matrices that compare captured image data to both possible and probable scene light sources. These matrices depend on a choice of n reference illuminants, $E_n(\lambda)$, and a set of m reference surface reflectances, $R_m(\lambda)$, that together encompass most colors found in nature.

As discussed in previous articles [15], we can use several different methods to produce the correlation matricies once the spectral sensitivities are measured. All of these methods use the basic correlation framework to compute the estimated white point of the scene by taking the maximum likelihood of the following:

$$\mathbf{v}^t\mathbf{M}, \tag{4}$$

where **M** is the correlation matrix and **v** is the image data in a vector form of the image chromaticity histogram. Having now determined our estimate of the scene illuminant, we can either balance the raw sensor data before the demosaic step, or include this into the convolution operator explained in the next section.

b) Mosaiced data reconstruction

We have cast the problem of generating the fully populated planes of color image data from the sensor data as a reconstruction problem rather than one of interpolation. As our solution is mathematically intense, it will only be summarized here. There are two parts, the generation of a convolution kernel, **T**, and the actual convolution,

$$\mathbf{Y} = \mathbf{T}\,\mathbf{X}, \qquad (5)$$

which is performed in real time just after an image is captured. The image data from the sensor is represented by **X**, and **Y** is the reconstructed image. The operator **T** is calculated *a priori.* The measured capture subsystem attributes (spectral sensitivities, PSF, OECF) and observations of the natural world (noise spectral power distributions in scenes, the estimated illuminant spectral power distribution, and probabilistic descriptions of the colorimetry of scene surfaces) are used as inputs for a newly developed reconstruction algorithm.[11] Deterministic mathematical models for the image and the capture process, and statistical models for noise and scene surfaces are generated and incorporated into a tensor operator of high dimensionality. Aided by simplifying mathematical assumptions, a linear mean-square optimization of this tensor is performed, resulting in the set of coefficients for the final reconstruction operator, **T**. This operator typically has an 8x8x3 pixel structure, and all sensor data within the footprint are used in calculating the missing color values for the target pixel in the reconstruction.

The result of this convolution is an estimate of the color values not measured at a given pixel. Because the capture subsystem is calibrated and therefore the sensor data is physically traceable, the result is a best-estimate reconstruction of the color triplets describing the original scene. Once transformed into an appropriate color space, the scene data is delivered to the rest of the camera pipeline for pictorial processing. In the nature of the mathematical constraints used, and the scope of physical processes considered, this technique is a significant improvement over other Bayesian reconstruction techniques [16]. Because of the computational complexity, a hardware accelerator capable of operating at over 500 MOPS was built into the custom ASIC used in these cameras. [17]

c) Linear transformation

The final step to get to scene-referred data is to apply a linear transformation to take the camera data from raw response space to a standard color space. Again, there are many methods available for computing the linear transformations. These methods are currently under investigation [18, 19] and several are under consideration for international standardization [14]. Both manufacturing cost and signal-to-noise limitations have prevented digital camera sensors being developed that have spectral sensitivities that are color matching functions. As a result, the raw response space cannot technically be considered a color space until a transformation has been applied to bring the data to a close approximation to a standard color space.

The method we have developed and used in our devices uses the camera's spectral sensitivities $S_k(\lambda)$, and the predicted scene white-point to compute the linear transformation. Although this particular method does not depend on any expected surface reflectance correlation statistics (the so called *'maximum ignorance'* method) it does depend on the accurate measurement of the camera sensors discussed above in the measurement section.

At this point we now have our best approximation of the scene-referred data for each color dimension at every pixel location. The next step, which goes beyond the scope of this paper, is to produce an optimized, output-referred reproduction of the scene through the application of color rendering algorithms. Some details have been previously described [5], and others will be presented in future publications.

Conclusions

We have based a digital still camera color reproduction solution on the new paradigm of generating physically meaningful scene data in the camera. We have described the methods that we employed for physical characterization of camera capture subsystems in order to routinely generate physically traceable data. Next, we presented our methods for establishing the adopted scene white point, for constructing an estimate of the full set of color triplets describing the scene, and for defining the transformation matrix to convert from the camera's raw response space into a color space suitable for use with the subsequent pictorial processing algorithms.

The result presented here is a physically accurate estimate of the color data in the original scene. This has proven to be a very powerful approach for enabling the use of new pictorial processing algorithms to deliver improved photographic performance.

References

1. Holm, J., A Strategy for Pictorial Digital Image Processing (PDIP), *Proc. IS&T/SID Fourth Color Imaging Conference: Color Science, Systems, and Applications*, Scottsdale, AZ, pp. 194-201. (1996).
2. Saito, O., et.al., "Signal Processing System for High Resolution Digital Camera," *Proc. IS&T PICS Conference*, Savannah, GA, pp.76-79. (1999).
3. Finlayson, G.D., Hubel, P.M., and Hordley, S., "Colour by Correlation", *Proc. IS&T/SID Fifth Color Imaging Conference: Color Science, Systems, and Applications*, Scottsdale, AZ, 6-11, 1997.
4. Hubel, P.M., Holm, J., and Finlayson, G.D. "Illuminant Estimation and Color Correction," *Colour Imaging in Multimedia*, Lindsay MacDonald, editor, Wiley, 1999, pp. 73-95.
5. Holm, Jack, "Photographic Tone and Colour Reproduction Goals", *Proc.CIE Expert Symposium '96: Colour Standards for Image Technology*, Vienna, Austria, pp. 51 56. (1996).

6. Shaw, Rodney, Digital Photography: Towards Ultimate Performance, *Proc. IS&T PICS Conference*, Montreal, Canada, March 2001, (in press).
7. Shaw, Rodney, Quantum Efficiency Considerations in the Comparison of Digital and Analog Photography, *Proc. IS&T PICS Conference*, Portland, OR, pp 165-168. (1998).
8. Shaw, Rodney, Signal-to-Noise Analysis for Digital Photography, *Proc. IS&T 50th Annual Conference*, Cambridge, MA, pp 335-338. (1997).
9. Baer, R.L., CCD Requirements for Digital Photography, *Proc.1999 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, pp.199-202. (1999).
10. Baer, R.L. & Holm, J., A Model for Calculating the Potential ISO Speeds of Digital Still Cameras based upon CCD Characteristics, *Proc. IS&T PICS Conference*, Savannah, GA, pp. 35-38. (1999).
11. Taubman, David, Generalized Wiener Reconstruction of Images from Color Sensor Data Using a Scale Invariant Prior, *Proc. IEEE International Conference on Image Processing*, Vancouver, Canada, September 2000, (in press).
12. ISO 12233:2000, *Photography – Electronic still-picture cameras – Resolution measurements*
13. ISO 14524:1999, *Photography – Electronic still-picture cameras – Methods for measuring opto-electronic conversion functions (OECFs)*
14. ISO/WD 17321.4 (1999), *Graphic Technology and Photography - Colour characterisation of digital still cameras (DSCs) using colour targets and spectral illumination*
15. Finlayson, G.D., Hordley, S., and Hubel, P.M., “Unifying Color Constancy”, *Proc. IS&T/SID Seventh Color Imaging Conference: Color Science, Systems, and Applications*, Scottsdale, AZ, 120-126. (1999).
16. Brainard, D., Bayesian Method for Reconstructing Color Images from Trichromatic Samples, *Proc. IS&T 47th Annual Conference*, Rochester, NY, pp. 375-379. (1994).
17. Agilent Technologies HDCP-3200 processor, www.agilent.com/view/imaging
18. Hung, P.C., Comparison of Camera Quality Indexes, *Proc. IS&T/SID Eighth Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*, Scottsdale, AZ, pp. 167-171. (2000).
19. Holm, J., Tastl, I, & Hordley, S., Evaluation of DSC (Digital Still Camera) Scene Analysis Error Metrics – Part 1, *Proc. IS&T/SID Eighth Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*, Scottsdale, AZ, pp. 279-287. (2000).