Camera Color Correction Using Two-Dimensional Transforms

Jon S. McElvain and Walter Gish; Dolby Laboratories, Inc.; Burbank, CA (USA)

Abstract

For digital camera systems, transforming from the native camera RGB signals into an intermediate working space is often required, with common examples involving transformations into XYZ or the sRGB. For scene-linear camera signals, by far the most common approach utilizes 3x3 matrices. For color pipelines designed for Rec709 displays, matrix-based input transforms are capable of producing reasonable accuracy in this domain. However, the associated colorimetric errors can become significant for saturated colors, for example those beyond Rec709. To address this shortfall, a novel input color transformation method has been developed that involves the use of two-dimensional lookup tables (LUTs). Because the surfaces associated with the 2D LUTs possess many degrees of freedom, highly accurate colorimetric transformations can be achieved. For several cinematic and broadcast cameras tested, this new transformation method consistently shows a modest reduction of mean deltaE errors for lower-saturation colors. The improvement in accuracy becomes much more significant as saturation increases, such that the mean deltaE errors are reduced by more than a factor of three.

Introduction

Recently, several factors have contributed to wider adoption of digital capture systems in the professional environment, for example in cinematic and episodic productions. Among these include improved noise performance, extended dynamic range capability, and importantly the creation of cost-effective digital workflow ecosystems. As productions continue to migrate toward digital, lower cost camera systems have been introduced, giving many small to medium budget productions access to high quality content creation. For example, both the Canon C500 and the RED Epic are relatively affordable in the $20K-$30K range; even prosumer cameras, such as the GoPro Hero3 or the Canon 5D mII/mIII, have been successfully used in numerous productions [1,2]. Although at higher price points, the ARRI Alexa and Sony F55/F65 have both found widespread use, and have produced imagery with quality that rivals that of modern cinematic film.

The image formation process for these systems is illustrated in Figure 1, for single sensor designs and more complex cameras utilizing multiple sensors. In the case of a single sensor configuration, a scene is imaged through the optical system onto the sensor. A color filter array (CFA) is patterned onto the sensor, and in the case of a Bayer design produces individual pixels with either a red, green, or blue response. With this CFA design, the spatial sampling of the green pixels is twice that of the red or blue channels, and to produce separate red, green, and blue images with the full sensor pixel count, various demosaicing algorithms are employed [3]. For three chip configurations (typically found in broadcast camera systems), dichroic mirrors in conjunction with red, green, and blue trimming filters produce the full resolution RGB channels without the need for demosaicing [4]. The spectral transmittance of the color filters when combined with the absorption of silicon form the spectral response characteristics of each channel. Example curves for single chip and multi-sensor cameras are shown in Figure 2.

Analog RGB signals, initially in the form of electrons in the well of the photodiode associated with each pixel, are formed by taking the projection of the focal plane spectral power distribution L(λ) and the RGB spectral sensitivity functions $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$ over all wavelengths:

$$\begin{aligned} R_{in} &= \int_0^\infty L(\lambda)\bar{r}(\lambda)d\lambda + n_R \\ G_{in} &= \int_0^\infty L(\lambda)\bar{g}(\lambda)d\lambda + n_G \\ B_{in} &= \int_0^\infty L(\lambda)\bar{b}(\lambda)d\lambda + n_B \end{aligned} \quad (1)$$

where the term $n_{R,G,B}$ refers to the signal-dependent electronic noise generated in the photosite wells. These electrons are converted to analog current or voltage signals during image readout, and subsequently digitized via an analog to digital conversion stage. The digital values produced at this point are typically designated as "native raw", to reflect the fact that no color processing has been applied to the RGB signals.

From Figure 3a, the color signal processing pipeline that follows involves conversion of these native raw pixel values into pleasing images on the destination display device. Various elements of the pipeline may occur in the camera firmware or offline using a color grading system operated by a colorist. Often the pipeline can be separated into two distinct stages: 1) transformation of the native RGB signals into a space with defined colorimetry such as XYZ; 2) rendering in this color space to eventually produce the final imagery. As an example, workflows using ACES [5] adopt this concept of separation into two stages, shown in Figure 3b. The Input Device Transform (IDT) is responsible for converting the camera signals into the ACES colorimetry. Following this, the data are passed through a block whereby rendering choices are made by the artists involved with the production, and this stage may include the Reference Rendering Transform (RRT) and a Look Modification Transform (LMT). The final element of this workflow is the Output Device Transform (ODT), which is responsible for transforming from the rendered signals (still in the ACES color space) to the destination display device.

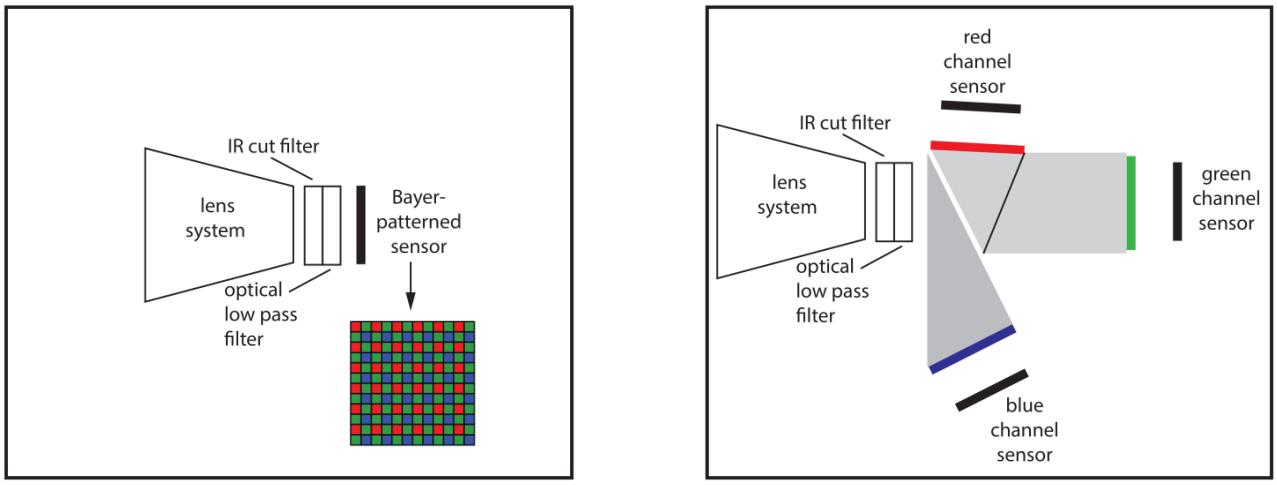

Figure 1*. The image formation elements for a) single and b) multi-chip camera systems.*

In this paper, we focus our attention entirely on the input color transform, as the transformation target output is well-defined by the destination colorimetry. For a workflow using ACES, the problem reduces to emulating the Reference Image Capture Device (RICD) which assumes a D60 illuminant; for other systems, the destination may be defined by an XYZ colorimeter. Regardless of the workflow, *accuracy* at this stage carries a particularly high value, as color errors introduced here will propagate through the entire system unless corrected. Saturation and/or hue artifacts created in the input transform may require nonlinear corrective algorithms in later stages, or force a colorist into region-specific secondary color correction. Furthermore, wide-gamut display systems such as those supporting Rec2020 primaries [6] are expected to exacerbate the visibility of these defects.

Existing Methods for Input Color Transforms

The most common method of input color transforms involves the use of 3x3 matrices, as these are simple to implement both in hardware and software. However, for transformations into XYZ or ACES, a camera system that strictly satisfies the Luther-Ives condition [7] does not exist. As depicted in Equation 1, the source and destination signals are formed via two separate projections from a Hilbert space, and their relationship is ultimately determined by the power spectral distribution of the incoming light at each pixel. Thus the 3x3 matrix is an approximation at best, and is typically determined using regression methods. Common approaches employ spectral reflectance databases in conjunction with knowledge of the camera spectral response characteristics [8]. From these data, a set of camera native R,G,B signals can be computed, and likewise for the destination space (e.g. XYZ or ACES). A variety of regression methods can be used to compute the matrix coefficients, but the most common approaches impose white point preservation as a constraint to the regression problem [9]. Hubel, et al. provide a performance comparison of several different matrix computation methods [10].

As an extension to the 3x3 matrix approach, transformations involving higher order polynomials have also been explored [11]. Relative to the linear 3x3 matrix transform, this has been shown to produce much greater levels of accuracy. However, this method suffers from its instability with respect to intensity scaling: changing the exposure (while preserving the native RGB ratios) can induce chromatic shifts that do not occur using the simple 3x3 matrix. Finlayson, et al [12] recently described a modification to the polynomial method, whereby each of the polynomial terms included n^{th} root products of each of the channels. For example, the red channel transform model could be represented as follows:

$$R_{out} = a_0 R_{in} + a_1 G_{in} + a_2 B_{in} + a_3 \sqrt{R_{in} G_{in}} + \cdots + a_m \sqrt[3]{R_{in} G_{in} B_{in}} + \ldots + a_n \sqrt[3]{R_{in}^2 G_{in}} \quad (2)$$

where the coefficients a_i are computed via regression. An important feature of the root-polynomial transform is the preservation of linear scaling with respect to intensity, as doubling each of the three input channels results in a corresponding doubling of the output signal. As with the integer-power polynomial models, root-polynomial input transforms demonstrate reductions in colorimetric errors relative to their 3x3 matrix counterparts. However, care should be taken when using either of these polynomial methods, as undesirable values may be generated for input signals that lie outside of the training set used for the coefficient regression. Also from the standpoint of system resources, the polynomial methods require considerably more computation at run time, particularly if there is a need to calculate n^{th} roots.

For systems with larger computational and memory resources, 3D lookup tables have also been widely used in this domain. Architectures usually employ preshaper 1D LUTs to tailor the input node spacing in the 3D LUT, and careful LUT generation can produce reasonably good accuracy for camera color transforms [13,14,15,16]. However, LUT sizes are generally constrained to sizes below 65^3 (and perhaps less for older generation hardware) and this can result in a highly sparse design where accuracy between nodes may be compromised. During run time, algorithms such as tetrahedral interpolation may improve the color accuracy particularly along the neutral axis, but inevitably errors will persist in some regions of the cube.

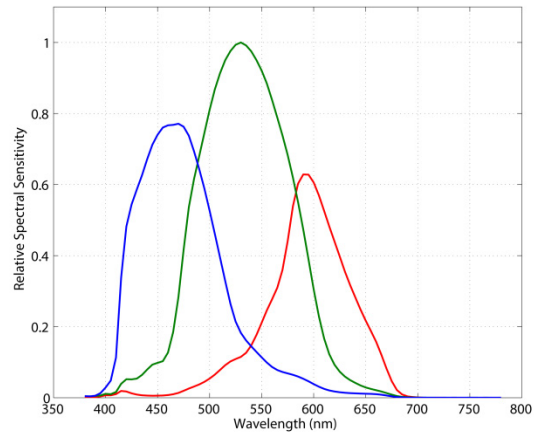

***Figure 2**. Example spectral sensitivity curves for a) single chip DSLR and b) three-chip camera, courtesy of [4].*

Color Transforms Based on 2D LUTs

As discussed, 3x3 matrices offer low-complexity solutions for color transforms from camera native to the working space, but may suffer from a reduction in accuracy, particularly for more saturated colors. Alternatively, 3D LUTs provide the ability to map between input and output Hilbert space projections in a nonlinear fashion, but they come at the cost of higher complexity and have the potential of introducing quantization artifacts due to system memory constraints. The polynomial methods are capable of substantial accuracy improvements over the 3x3 matrices, but at the cost of higher computational complexity during pixel processing.

We propose an alternate approach to improved-accuracy camera color transforms, that requires a memory footprint comparable to the 3D LUT but carries a relatively low complexity during pixel processing. The architecture is illustrated in Figure 4, and consists of a pre-processing block that precedes a 2D lookup, and finally a post-processing block. In the initial stage, chromaticity-like coordinates (p,q) are computed using the native camera signals. These are used to index into the 2D LUT and produce intermediate output values $(\tilde{R}, \tilde{G}, \tilde{B})$. An overall scale factor (Σ) is also calculated from the incoming camera signals, and this is used to provide radiometric scaling for the final output signals. The quantities computed in the pre-processing stage, which are assumed to include only the absolute value of the camera signals, are summarized below in Equation 3,

$$\Sigma = R_{in} + G_{in} + B_{in}$$
$$p = \frac{R_{in}}{\Sigma} \quad (3)$$
$$q = \frac{G_{in}}{\Sigma}$$

As seen from Equation 3, the pre-processing stage is relatively simple and only requires a small number of add and divide operations. Likewise the 2D lookup stage will have comparable complexity relative to common image interpolation blocks that are used for rotation and scaling, and either bilinear or bicubic methods will be sufficient. Finally the post-processing block will require a single multiply for each channel. In order to preserve linear scaling in the system, the nodes of the 2D LUT are populated with the following ratios:

$$\tilde{R}_{LUT}(p_i, q_j) = {R_{out}(p_i, q_j)}/{\Sigma_{in}(p_i, q_j)}$$
$$\tilde{G}_{LUT}(p_i, q_j) = {G_{out}(p_i, q_j)}/{\Sigma_{in}(p_i, q_j)} \quad (4)$$
$$\tilde{B}_{LUT}(p_i, q_j) = {B_{out}(p_i, q_j)}/{\Sigma_{in}(p_i, q_j)}$$

where (i,j) represent the LUT node coordinates. The importance of the 2D LUT is evident; it provides many degrees of freedom and thus the ability to tailor the LUT surfaces to minimize colorimetric errors across the entire (p,q) space spanned by the input signals.

The 2D LUT nodes can be populated using reflectance training data, pure functional representations, or a combination of both. When using training data, the challenge is to create a uniformly sampled LUT surface from scattered and often sparse data in the (p,q) domain. Also of importance is the imposition of a smoothness constraint on the surface, as there can be a tendency to over-fit to the training data. Lastly, the LUT surfaces should be well-behaved in the extrapolation regions outside the (p,q) spectral locus of the camera, as camera noise can produce (p,q) values in these regions.

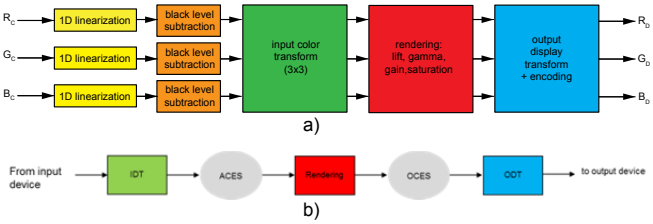

***Figure 3**. a) Signal processing pipeline for digital camera systems; b) block diagram of a workflow using the Academy Color Encoding Specification (ACES).*

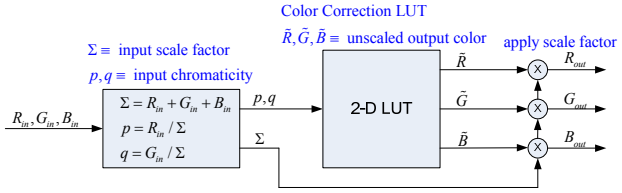

***Figure 4**. a) Block diagram of proposed color transform, where a 2D LUT is used to provide better accuracy.*

It is interesting to recast the conventional 3x3 matrix transformation into the equivalent 2D LUT representation in order to more easily understand the differences between the two methods. The matrix equations relating the input and output signals are as follows,

$$R_{out} = c_{11}R_{in} + c_{12}G_{in} + c_{13}B_{in}$$
$$G_{out} = c_{21}R_{in} + c_{22}G_{in} + c_{23}B_{in} \quad (5)$$
$$B_{out} = c_{31}R_{in} + c_{32}G_{in} + c_{33}B_{in}$$

where c_{ij} are the 3x3 matrix coefficients. Dividing each of these equations by Σ results in the following expression for the LUT nodes for a 3x3 matrix:

$$\tilde{R}_{LUT}(p_i, q_j) = (c_{11} - c_{13})p_i + (c_{12} - c_{13})q_i + c_{13}$$
$$\tilde{G}_{LUT}(p_i, q_j) = (c_{21} - c_{23})p_i + (c_{22} - c_{23})q_i + c_{23} \quad (6)$$
$$\tilde{B}_{LUT}(p_i, q_j) = (c_{31} - c_{33})p_i + (c_{32} - c_{33})q_i + c_{33}$$

Equation 6 indicates the 3x3 matrix can be represented as three planar LUT surfaces, with three degrees of freedom per channel. If white point preservation is imposed in the 3x3 matrix regression, then there are only a total of six degrees of freedom for the system (two per channel).

Figure 5 shows false color plots of the red channel 2D LUT surfaces for a cinematic camera (camera raw to ACES), for both a 3x3 matrix and optimized LUT surface. Near the (p,q) white point, it is evident that the optimized LUT surface is approximately planar in nature. However, moving away from the white point, the differences between the 3x3 and 2D LUT approaches become apparent. The optimized LUT shows a much more complex surface structure since it is designed to minimize colorimetric errors throughout the entire physical (p,q) domain. In contrast, the 3x3 matrix transform maintains a constant planar surface and thus colorimetric errors will in general be more pronounced as color saturation is increased.

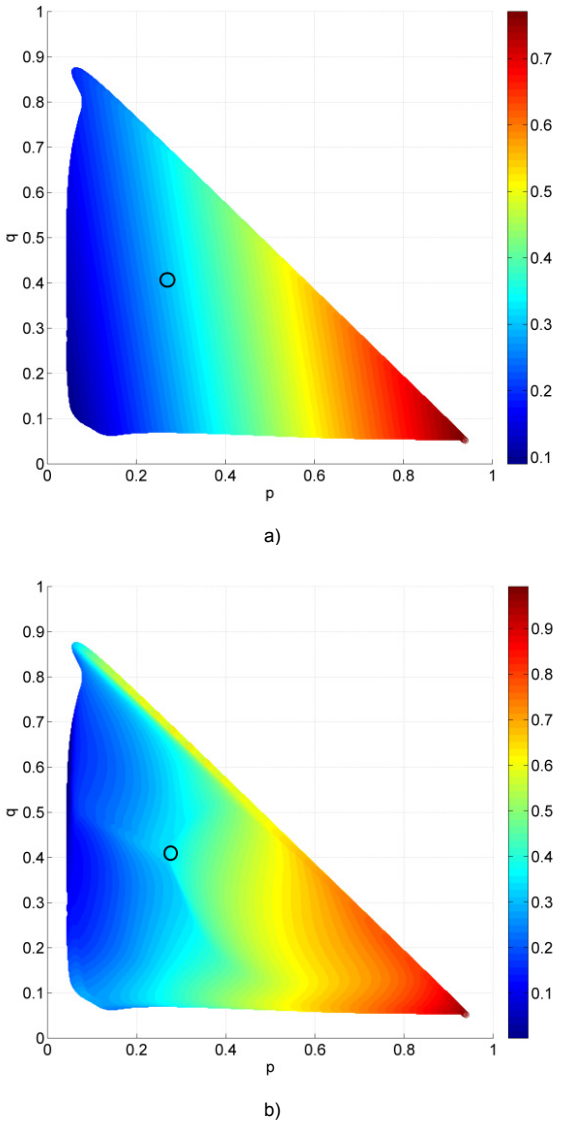

***Figure 5**. Comparison of 3x3 matrix and 2D LUT surfaces for the red channel of a cinematic camera. a) 3x3 matrix; b) optimized 2D LUT. The circle indicates the white point, and the plots are restricted to the physical camera (p,q) domain.*

Experimental Results

Several cameras, both cinematic and broadcast, were analyzed in the context of 2D LUT input transforms, and compared to the 3x3 matrix approach. In our initial analysis, 525x525 2D LUTs were constructed as a proof of concept, as this roughly corresponds to the memory footprint required for a 65^3 3D LUT. As expected, the 2D LUT approach consistently yielded significantly lower average color errors, particularly for sample points corresponding to saturated colors. It should be noted that smaller LUTs were not explored, but it is believed that similar performance can be achieved when combined with 1D preshaper LUTs.

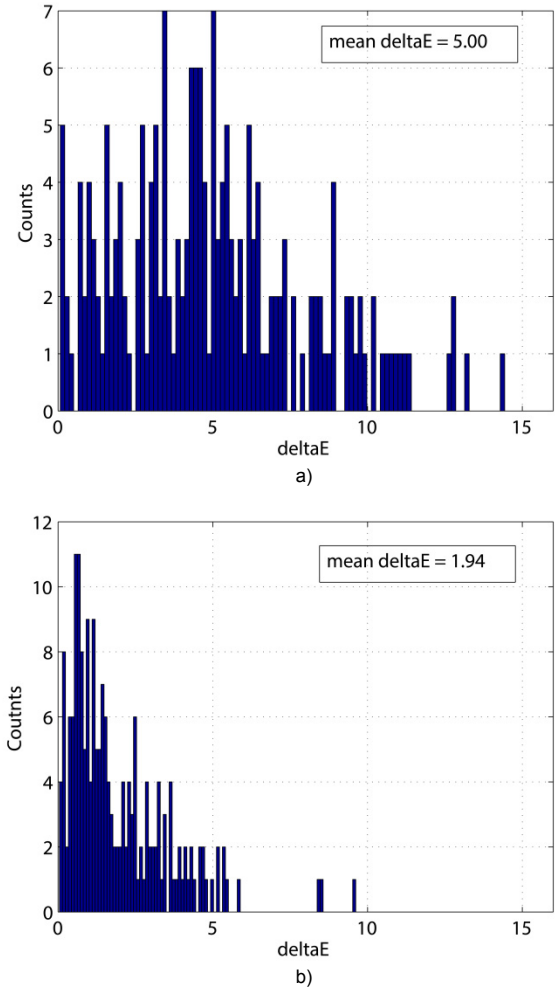

***Figure 6**. Error histograms for cinematic color transformations from camera native to ACES RGB : a) 3x3 matrix; b) 2D LUT.*

As an example, consider the Input Device Transform (camera native to ACES) for a cinematic camera for which we directly measured the spectral response curves. Using these characteristics in conjunction with a 190-element reflectance dataset, native RGB signals were computed assuming a blackbody illuminant at 3200K. These same reflectance values were used to compute reference ACES RGB values according to the ACES Reference Image Capture Device (RICD) [5] specification with a 6000K D-series illumination. Using a manufacturer-supplied 3x3 matrix IDT optimized for 3200K, the native camera RGB signals were then used to calculate predicted ACES RGB values. Following conversion from ACES to XYZ, Figure 6a shows the computed colorimetric errors via CIELAB 1976 (relative to the reference ACES RGB values) realized when using the 3x3 matrix IDT. The results indicate an average ΔE=5, and the important observation here is that many of the predicted RGB values will generate highly visible differences relative to the actual reference colors. Separately, the 190 native camera signals were processed through the 2D LUT IDT to form a new set of predicted ACES values, and after conversion from ACES to XYZ the colorimetric errors are significantly reduced (Figure 6b). The mean error is now below ΔE=2, and it is apparent that the histogram is strongly clustered near the origin. There are a few samples with large errors, but this fundamentally cannot be avoided due to different metameric behavior between the input camera and the RICD. However the number of samples with ΔE>5 is greatly reduced when using the 2D LUT.

For another example, color transforms from 3200K camera native to XYZ D65 were considered for a broadcast camera, with spectral response curves provided by the authors of [4]. As with the previous example, a set of 190 reflectances were used for the test database, and native RGB and reference XYZ values were generated for this dataset. Because a manufacturer-supplied 3x3 transform matrix was not available, one was generated using a separate 772-element training set and Finlayson-Drew white-point preserving regression [9]. Two sets of predicted XYZ values were computed, one using this matrix, and the other using the optimized 2D LUT. Figure 7 shows the results for both the 3x3 matrix and the 2D LUT approaches. A modest improvement in average color errors can be observed for the 2D LUT, as once again the histogram is clustered more toward the origin.

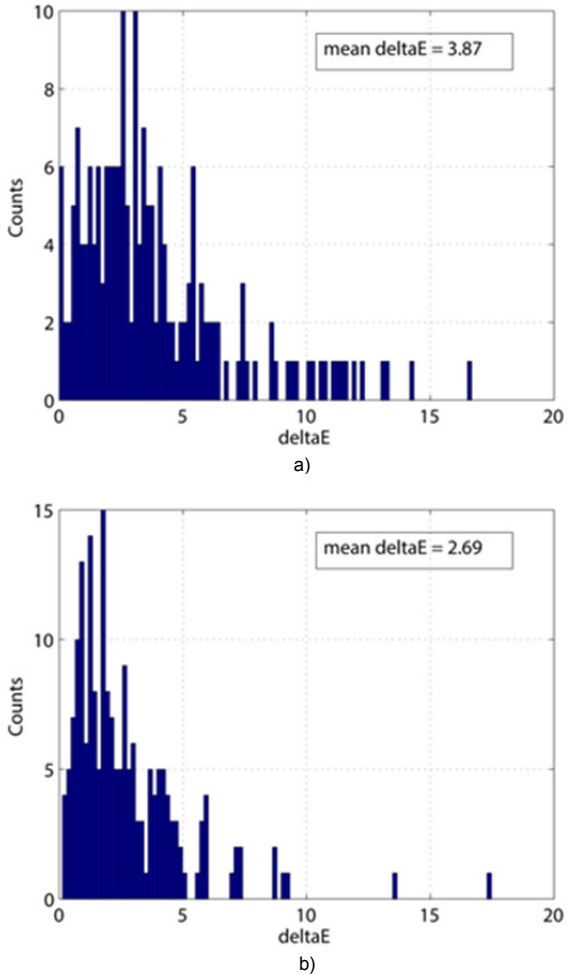

***Figure 7**. Error histogram computed for a broadcast camera using a) 3x3 matrix and b) 2D LUT transforms from camera native to XYZ D65.*

Figure 8 shows false color plots depicting the color errors as a function of reference chromaticity, this time using a 12,000+ element natural reflectance set. For illustration purposes, the Rec709 and Rec2020 boundaries are shown on both plots. It can be readily observed that the 3x3 matrix transform performs well near the white point, but begins to exhibit significant errors for larger saturation values. Conversely, the 2D LUT transform is able to maintain low colorimetric errors throughout the domain of the data set, and in particular performs well near and beyond the Rec709 boundaries. Figure 9 shows error histograms for colors beyond Rec709, for both the 3x3 and 2D LUT methods. The mean error for the 3x3 matrix is in excess of ΔE=8, indicating the majority of these samples will be transformed with highly visible errors. In contrast, the 2D LUT transform results in significantly lower errors (ΔE=2.21), indicating this method would be more appropriate for processing imagery containing more vibrant colors, in particular when intended for wide-gamut display devices.

Conclusions

A novel input color transformation method that involves the use of two-dimensional lookup tables (LUTs) has been described. Because the surfaces associated with the 2D LUTs can have many degrees of freedom, highly accurate colorimetric transformations can be achieved. For several cinematic and broadcast cameras tested, this new transformation method consistently shows a modest reduction of mean ΔE errors for colors within the Rec709 primaries. The improvement in accuracy becomes much more significant for saturated colors.

There are a number of ways in which this technology can be applied. First and foremost, it offers more accurate processing of raw images from existing digital cameras. Furthermore, by decoupling the need for special sensitivities to satisfy the Luther-Ives condition in order for a 3x3 matrix to be effective, it could allow greater freedom in the design of camera sensors and systems overall. Lastly, the general form of this 2D transformation may have application elsewhere in the image processing pipeline as it preserves linear scaling while offering a completely general color transformation with a more modest footprint than a 3D LUT.

Until now, the inaccuracies in camera input processing have been viewed largely as an annoyance to be corrected in later stages of the camera color pipeline. We anticipate that this problem will become more significant as display systems migrate to wider color gamuts. 3x3 matrices may be sufficient for imagery than does not contain many saturated colors. However, we assert that higher-complexity camera transforms, such as the 2D LUT method described here, will be crucial for workflows intended for wide-gamut display systems such as those compatible with ITU recommendation 2020.

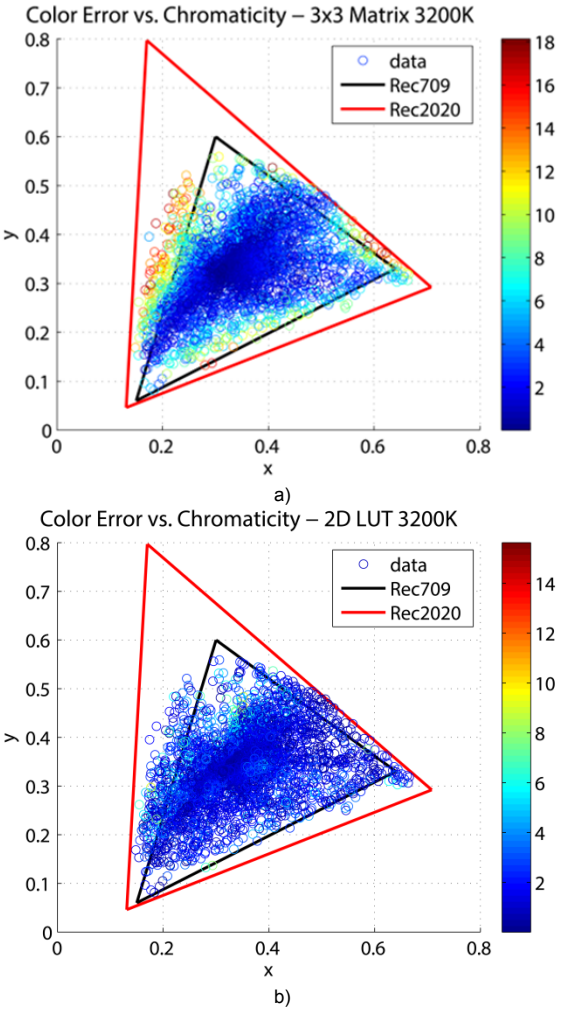

***Figure 8**. Color errors vs. reference chromaticity for a broadcast camera: a) 3x3 matrix, showing good performance near the white point but significant failures for more saturated colors; b) 2D LUT, exhibiting low errors throughout the domain of the data set.*

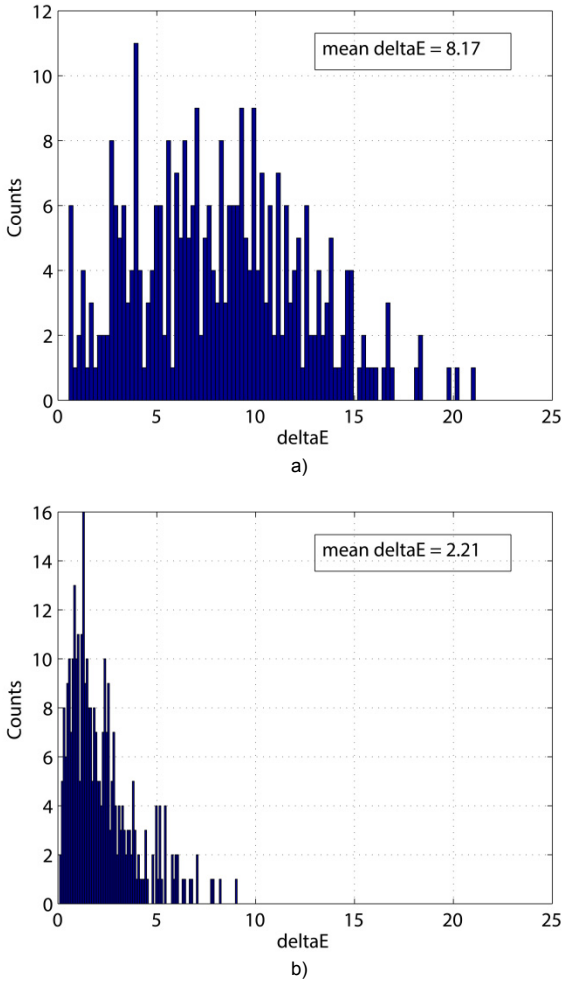

***Figure 9**. Color error histograms for colors outside Rec709: a) 3x3 matrix transform; b) 2D LUT transform.*

References

[1] *Act of Valor*. Motion Picture. Directed by Mike McCoy and Scott Waugh. Bandito Brothers (2012).

[2] *The Art of Flight*. Motion Picture. Directed by Curt Morgan. Red Bull Media House (2011).

[3] X. Li, B. Gunturk, and L. Zhang, Image Demosaicing: A Systematic Survey, Proc. SPIE-IS&T Electronic Imaging: Visual Communications and Image Processing, 6822, pg. 1J1. (2008).

[4] P. Böhler, J. Emmett, and A. Roberts, "Toward a Standard Television Camera Color Model," SMPTE Mot. Imag. J 122, 30 (2013).

[5] SMPTE ST 2065-1, Academy Color Encoding Specification (ACES). (2012).

[6] ITU-R BT.2020, BT.2020 : Parameter values for ultra-high definition television systems for production and international programme exchange, International Telecommunication Union. (2012).

[7] R. Luther, "Aus dem Gebiet der Farbreizmetrik," Zeitschrift für technische Physik **8**, 540 (1927).

[8] M. Brady and G. E. Legge, "Camera Calibration for Natural Image Studies and Vision Research," J. Opt. Soc. Am. A 26(1), 30 (2009).

[9] G. Finlayson, and M. Drew, White-Point Preserving Color Correction, Proc. IS&T Fifth Color Imaging Conference, pg. 258. (1997).

[10] P. Hubel, J. Holm, G. Finlayson, and M. Drew, Matrix Calculations for Digital Photography, Proc. IS&T Fifth Color Imaging Conference, pg. 105. (1997).

[11] G. Hong, M. R. Luo, and P. A. Rhodes, "A Study of Digital Camera Characterisation Based on Polynomial Modeling," Color Research and Application 6(1), 76. (2001).

[12] G. Finlayson, M. Mackiewicz, and A. Hurlbert, Root Polynomial Colour Correction, Proc. IS&T 19th Color Imaging Conference, pg. 115. (2011).

[13] S. Srivastava, T. Ha, E. Delp, J. Allebach, Generating Optimal Look-up Tables to Achieve Complex Color Space Transformations, Proc. 16th IEEE International Conference on Image Processing (ICIP), pg. 1641. (2009).

[14] P. Hung, "Colorimetric Calibration in Electronic Imaging Devices Using a Look-up Table Model and Interpolations," Journal of Electronic Imaging 2(1), 53 (1993).

[15] J. Morovic, A. Alberran, J. Arnabat, Y. Richard, and M. Maria, Accuracy-Preserving Smoothing Of Color Transformation LUTs, Proc. IS&T 16th Color Imaging Conference, pg. 243. (2008).

[16] Z. Wang, A. Aristova, and J. Hardeberg, Evaluating the Effect of Noise on 3D LUT-based Color Transformations, Proc. 2010 Conference on Colour in Graphics, Imaging, and Vision (CGIV), pg. 88. (2010).

Author Biography

Jon McElvain is currently with the Advanced Technology Group of Dolby Laboratories, specializing in camera and display systems, image quality quantification, and image processing algorithm development. He received a B.S. in physics from the University of California, San Diego, an M.A. in physics from the University of California, Berkeley, and a Ph.D. in physics from the University of California, Santa Barbara. He is a member of SMPTE and IS&T.

As Principal Member for Dolby's Advanced Technology Group, Walter Gish leads research and development efforts for next generation imaging technology. He has over 30 years of experience in signal and image processing, high performance computing and hardware development. Since the early 1980s he has focused on computer graphics, video, and display technology. Mr. Gish received his BS and MS in physics and applied mathematics from the California Institute of Technology in 1968 and 1970.