Camera Design Using Locus of Unit Monochromats

James A. Worthey, Gaithersburg, Maryland; Michael H. Brill, Datacolor, Lawrenceville, New Jersey

Abstract

The Maxwell-Ives criterion (MI) says that for color fidelity a camera's spectral sensitivities must be linear combinations of those for the eye[1]. W. A. Thornton's research found certain wavelengths, the "prime colors" (PC), with special importance for color vision. At CIC 6, M. H. Brill et al. spoke in favor of "cameras that have peak sensitivities at the PC wavelengths."[2] MI and PC are not independent ideas. MI implies symmetry between the camera and the eye: the camera has its own prime colors, which should be similar to the eye's. At CIC 12, J. A. Worthey presented an orthonormal opponent set of color matching functions as a path to J. B. Cohen's Locus of Unit Monochromats (LUM), an invariant representation of color-matching facts[3]. We now present a concise method to evaluate a sensor set by comparing its LUM to the eye's. Equal LUMs would mean that MI is met, and equal PC wavelengths would tend to mean that MI is loosely met. We notice that two sets of camera sensors can have the same LUM, but differ in the net effect that sensor noise will have. A numerical noise example illustrates the point.

Introduction

The Maxwell-Ives criterion says that for color fidelity a camera's spectral sensitivities must be linear combinations of those for the eye[1]. Two underlying ideas are: (1) That the color sensitivities act together as a set—one sensor by itself is not right or wrong. (2) That there is symmetry between the camera and the eye. Turning to Jozef Cohen's theory of color mixture [4-6], two basic ideas are: (3) That the color sensors act as a set, the same as in Maxwell-Ives. (4) That the facts of color mixture can be expressed by an invariant graph in 3 dimensions. Extending Cohen's method to a camera allows its color-mixing traits to be summarized and compared to the eye's.

Camera sensors can be evaluated for overall goodness [7], but since one may wish to work with an existing sensor set [8], to deal with variability [9], or simply to make compromises, there is a need for a conceptual framework in which details can be discussed. The method below takes ideas from Cohen's [4-6] and Thornton's [10] research, and from the orthonormal-opponent scheme [3] presented at CIC 12.

Projector Matrix

Suppose that **A** is an $N \times 3$ matrix, whose columns are a set of color-matching functions (CMFs), $\mathbf{A} = [a_1\, a_2\, a_3]$. Cohen found a projection operator to be "matrix R," given by

$$\mathbf{R} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T, \tag{1}$$

where superscript T denotes matrix transpose. The original application was, if L is the SPD of a light, to find L^*, the projection of that light into the vector space of the CMFs:

$$L^* = \mathbf{R}L\ . \tag{2}$$

L^* is called the fundamental metamer [4-6] of L. The projection operation is based on a least-squares fit, so it is convenient (and accurate) to think of various curve-fitting steps as projections. As Cohen discovered [4-6], **R** is invariant: transforming the CMFs in A to a different representation leaves **R** unaltered. The columns (or rows) of **R** are the L^*'s of unit-power monochromatic lights, which can be plotted as vectors in a 3-space. The curve those vectors trace is the Locus of Unit Monochromats, an invariant embodiment of the facts of color mixing.

A camera's sensor functions determine its LUM. Then the Maxwell-Ives criterion is that the camera's locus of unit monochromats should be the same as the eye's. Comparing a camera's locus to the eye's offers some insight in the realistic case that MI is in fact not met.

Abbreviations summarized

MI = Maxwell-Ives criterion; LUM or just 'locus' = locus of unit monochromats for the eye, or for a camera; PC = prime colors. The words sensitivity and sensor will generally refer to spectral properties.

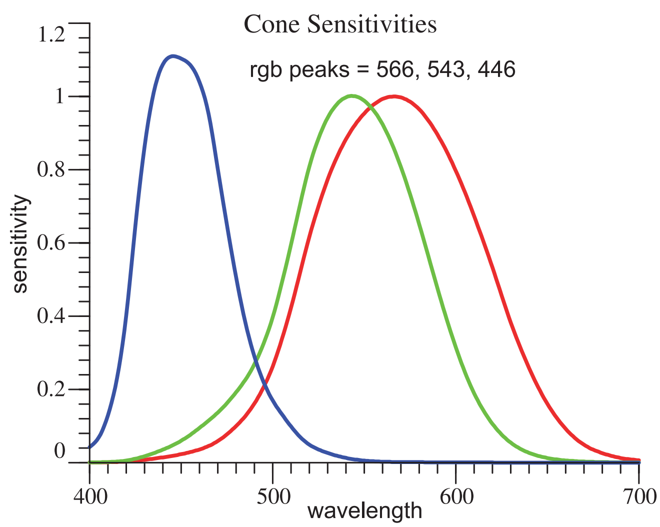

Figure 1. Human cone sensitivities, consistent with the CIE's 2° observer. [11] The red, green, and blue functions peak at 566, 543, and 446 nm.

Red-Green Overlap

In its simplest statement, MI is a pass-fail test that most cameras will fail. Fig. 1 shows a version of human cone sensitivities, which are consistent with the CIE's 2° observer. The overlapping functions are a defining feature of human vision, and in particular the red and green sensitivities show marked overlap. A direct way to satisfy MI would be for a camera's sensors to mimic the eye, including the spectral overlap of red and green sensors. Unfortunately, that would lead to highly correlated signals from the red and green sensors and when those signals are subtracted to recover hue information, the ratio of signal to noise would be poor.

Table 1

system	sensors	peak sep.	dir. cos
Eye	r-g	23 nm	0.918
	g-b	97	0.121
Nikon D1	r-g	59	0.175
	g-b	62	0.362

The camera designer's task then is to choose the sensors with MI in view, but also to minimize noise. The camera may map colors into a color space different from the eye's, but one hopes that a further mapping (a linear transformation perhaps) can then map most objects near to their proper place in human color space. Fig. 2 shows the sensitivities of a Nikon D1 camera, as presented at an earlier Color Imaging Conference [12]. Compared to the 2° observer, the camera does not show the same breadth and overlap of the red and green sensitivities. The wavelength locations of the peaks are indicated in the figures. Combining those numbers with a little further calculation leads to Table 1, where peak sep. denotes peak separation, and dir. cos. denotes the direction cosine between two functions. These measures support the observation that the camera's red and green sensitivities overlap less than the 2° observer's. The camera's green and blue functions overlap more than the human's.

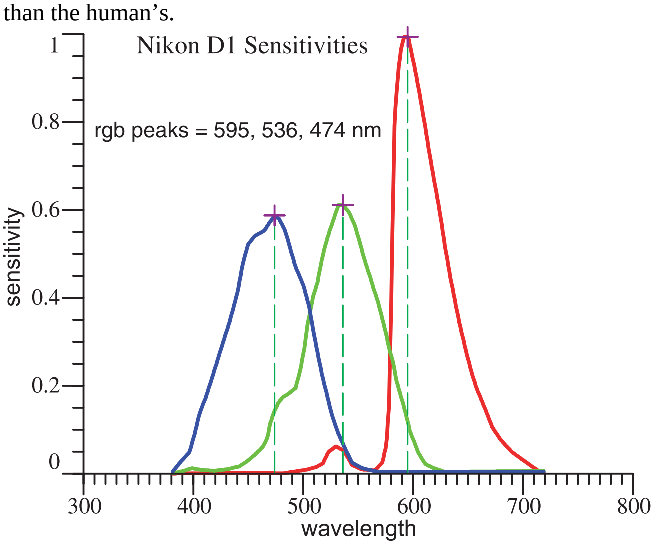

Figure 2. Sensitivities of a Nikon D1 camera as reported at an earlier Color Imaging Conference [12]. The red, green, and blue functions peak at 595, 536, and 474 nm. The purple plus signs and dashed vertical lines show the peak points as found by a simple algorithm.

Fundamental Metamers

If two or more lights with different spectra match to a standard observer, then traditionally one would say that they have the same tristimulus vector $[X\ Y\ Z]^T$. Another proxy for the matching lights is their fundamental metamer, the function L^* defined above [4-6]. L^* uses the facts of color mixing, but transcends the arbitrariness of XYZ or any particular system. Dividing the visible spectrum into narrow bands, then finding the fundamental metamer for unit power in each band, leads to a series of vectors in a color space, tracing the LUM [4-6]. The columns (or rows) of the projector matrix **R** are a set of vectors tracing the LUM, which shows that its shape is invariant.

Orthonormal Basis

Color matching data, such as the 2° observer, can be linearly combined to form orthonormal opponent color matching functions, with the interpretation of achromatic (proportional to $\bar{y}$), red-green opponent, and a kind of blue-yellow function, Fig. 3. [3] The first two functions involve only red and green receptors, so they deal with the key issue of red-green overlap. The third function has input from all three cones. Combining the three functions of this orthonormal basis in a parametric plot gives Cohen's Locus of Unit Monochromats, now graphed with respect to meaningful axes. The tristimulus values of any L are the coefficients for the expansion of L^* in terms of the orthonormal basis. [3] Colors add vectorially, so its vector denotes a light's direction in color space and its strength of action in mixtures. In the XYZ system, color mixing is modeled by vector addition, but vector diagrams would be hard to interpret and are not drawn.

The orthonormal functions can become the columns of a matrix **Ω**:

$$\boldsymbol{\Omega} = [|\omega_1\rangle\ |\omega_2\rangle\ |\omega_3\rangle] \tag{3}$$

A ket such as $|\omega_j\rangle$ is a column vector, while a bra such as $\langle\omega_i|$ is a row vector, allowing orthonormality to be written:

$$\langle\omega_i|\omega_j\rangle = \delta_{ij}\,, \tag{4}$$

where δ_{ij} is the Kronecker delta, = 1 if $i = j$, = 0 if $i \neq j$.

A tristimulus vector based on the color matching functions **Ω** can be called V. [3] Its components represent orthogonal directions in color space, and have intuitive meanings. If $|L\rangle$ is a light's spectrum, then its tristimulus vector is

$$V = \boldsymbol{\Omega}^T|L\rangle\ . \tag{5}$$

Letting $|L\rangle$ be a narrow-band light of unit power stepped through the spectrum, V traces out the LUM. Virtual-reality 3D graphs of the LUM are available on http://www.jimworthey.com . Applying orthonormality, Eq. (4), in Eq. (1), simplifies the formula for **R**:

$$\mathbf{R} = \boldsymbol{\Omega}\boldsymbol{\Omega}^T, \tag{6}$$

only when the CMFs are the orthonormal set.

Prime Colors and Longest Vectors

Thornton coined the phrase "Prime Colors" for those wavelengths that act most strongly in mixtures. [2,10] Within the LUM, the prime-color wavelengths (e.g., 446, 538, 603 nm for the

2° Standard Observer) are approximately the wavelengths of the longest tristimulus vectors (e.g., 445, 536, 604 nm). [3,13]

Naive Hypothesis

Thornton found that metameric spectra tend to cross at the prime color wavelengths. [10,2] It would then make sense that those wavelengths also be important to a camera's detection of color. A simple hypothesis is that a successful camera will have red and green sensors somewhat narrower than the eye's, and somewhat more separated in wavelength, with the net effect that it mimics the eye's prime colors. Reality may not be so simple, but a plan suggests itself: compare the camera's LUM to that of the eye, and look especially at the camera's longest-vector wavelengths or prime colors.

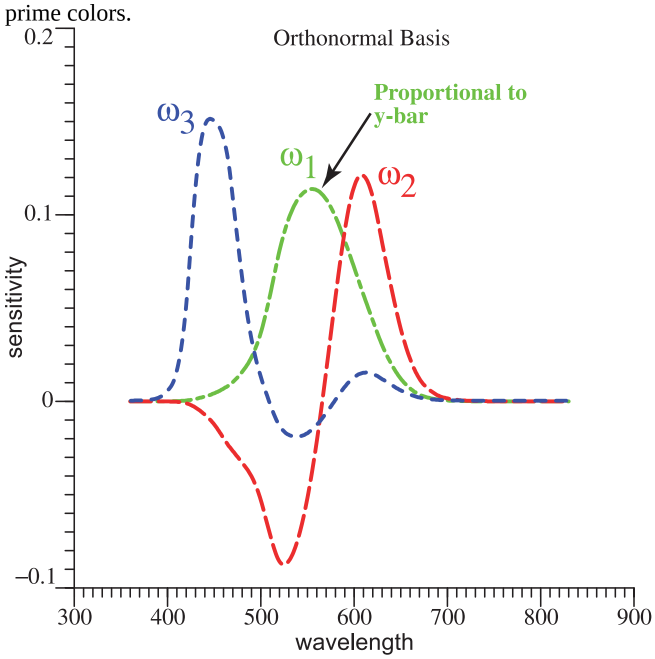

Figure 3. The orthonormal color matching functions for humans, based on the 2° observer. [3]

Implementation

Since the camera's sensors define a different vector space from the eye's, there is not an inherently correct way to graph them together. The following method has some logic. We assume that orthonormal functions [3] for the eye are in hand, Fig. 3, along with *rgb* functions for the camera. Then orthonormal functions are generated for the camera, and they determine its LUM. The goal is not curve-fitting as such, but we begin by making a fit of one visual function by 2 of the 3 camera functions. A projector matrix is used for the least-squares fit:

1. Call the camera's sensor functions r, g, b. Define $\mathbf{A} = [|r\rangle\ |g\rangle]$, so that the red and green functions become the columns of A. Then compute the projector matrix [4-6] $\mathbf{R}_{rg} = \mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}$.
2. Find $|\omega_1(\text{camera})\rangle = \mathbf{R}_{rg}|\omega_1\rangle$, where $|\omega_1\rangle$ is the known function for the eye. $|\omega_1(\text{camera})\rangle$ is normalized in the next step.
3. Assemble 3 vectors into a temporary matrix, $[|\omega_1(\text{camera})\rangle$, $|r\rangle$, $|b\rangle]$ and do Gram-Schmidt orthonormalization on them in that sequence. Then $\omega_2(\text{camera})$ will be a red-green opponent function involving only the red and green sensors, and $\omega_3(\text{camera})$ will be a blue or blue versus yellow function that involves all three sensors.
4. Combining the camera's orthonormal functions into a single 'parametric plot' gives its LUM, positioned for comparison to that of the eye.

These steps maintain transparency of cause and effect; in particular, $|\omega_1\rangle$ and $|\omega_2\rangle$ are always orthonormal linear combinations of $|r\rangle$ and $|g\rangle$ only. Steps 1-3 are intended to align the camera's LUM with the eye's. We resist the temptation to further optimize the alignment.

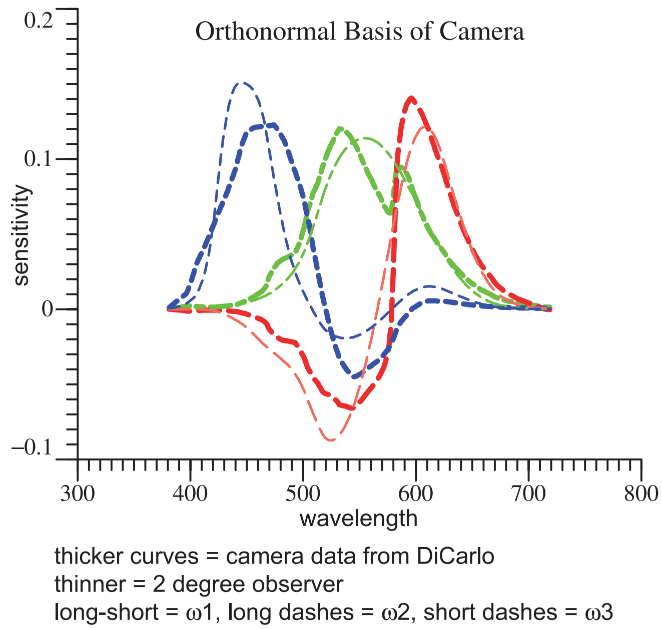

Figure 4. The orthonormal basis for the camera, with human orthonormal basis shown as thinner lines.

Examples

By the 4 steps above, the orthonormal basis in Fig. 4 was computed for the camera sensors of Fig. 2. The camera's LUM is compared to the eye's in two views, Figs. 5 and 6. The solid curves show the camera's intrinsic color-matching properties, independent of any human observer. The dashed curves are the eye's locus [3]. Figs. 5 and 6 are messy, but show the reality of the situation. There is freedom in how the LUMs are situated with respect to the axes, but otherwise they are invariant shapes. The axes have intuitive meanings, enhanced because an orthonormal basis means no double-counting.

Ignoring the short arrows for now, the curves in Figs. 5 and 6 express the key result. In these figures, we see the vector to which the camera will map a light of unit power and wavelength λ. We see the mapping of one wavelength in relation to another, and the camera's mapping compared to the eye's. The eye data are the same Cohen space and LUM for which other uses have been outlined [3]. The terse Maxwell-Ives idea blossoms into a detailed description of the camera's sensor set.

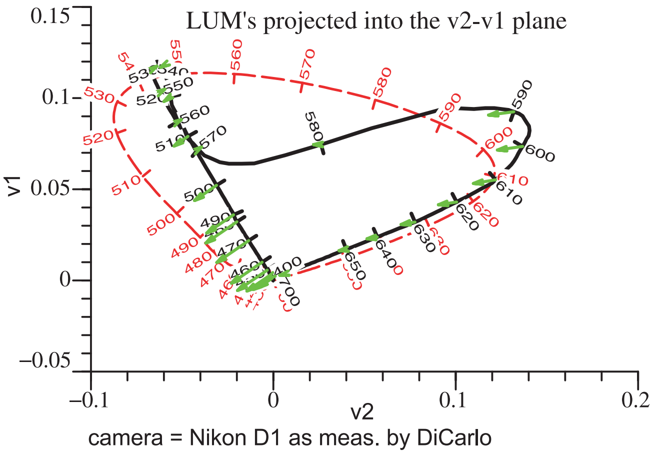

Figure 5. This figure and the next one are projections of the same 3-dimensional graph. The dashed curve is the LUM for the 2° human observer. The solid curve is the LUM for a camera. v1 is the achromatic dimension.

Additional "Best Fit" Step

When the steps above succeed at aligning the camera's LUM with the eye's, they suggest a working transform from camera signals to visual space. We now seek the further effect of a "best fit" step. Eq. (6) can give us the camera's projector matrix:

$$\mathbf{R}_{cam} = \mathbf{\Omega}_{cam}\mathbf{\Omega}^{T}{}_{cam} . \tag{7}$$

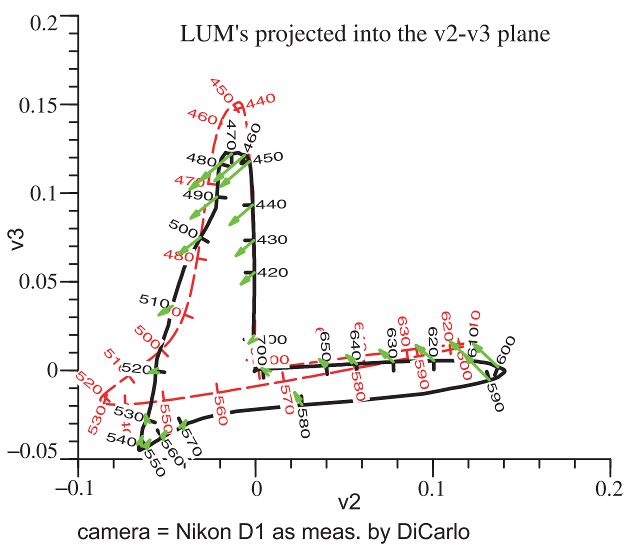

Figure 6. The same 3D curve as in Fig. 5 is now projected into v_3 vs v_2, which can be called the chromatic plane. v_2 is the red vs green dimension, while v_3 is blue vs. yellow. The arrows in both figures show the effect of making a best fit to the human functions with those of the camera.

The fit which we seek is the projection of the eye basis into the space of the camera, using $\mathbf{R}_{cam}$:

$$\mathbf{\Phi} = \mathbf{R}_{cam}\mathbf{\Omega} , \tag{8}$$

where $\mathbf{\Omega}$ is the eye's basis, as in Eq. (3). The columns of $\mathbf{\Phi}$ are the fit functions, linear combinations of the camera's orthonormal basis, $\mathbf{\Omega}_{cam}$. In Figs. 5 and 6, the small arrows show the correction from the camera's intrinsic $\mathbf{\Omega}_{cam}$ to $\mathbf{\Phi}$. Eq. (8) is not a method for mapping a 3-vector from the camera to human color space. To derive the needed 3×3 matrix, combine Eqs. (7) and (8):

$$\mathbf{\Phi} = \mathbf{\Omega}_{cam}[\mathbf{\Omega}^{T}{}_{cam}\mathbf{\Omega}] . \tag{9}$$

The product in square brackets results in a 3×3 matrix. Call it

$$\mathbf{X} = \mathbf{\Omega}^{T}{}_{cam}\mathbf{\Omega} , \tag{10}$$

no connection to the XYZ system. Then $\mathbf{X}$ is a transform from camera basis to "fit" basis:

$$\mathbf{\Phi} = \mathbf{\Omega}\mathbf{X} . \tag{11}$$

If $|L\rangle$ is any radiance, take the matrix transpose on both sides of Eq. (11) and find the tristimulus vector on both sides, according to Eq. (3). Then

$$\mathbf{\Phi}^{T}|L\rangle = \mathbf{X}^{T}\mathbf{\Omega}_{cam}{}^{T}|L\rangle . \tag{12}$$

Referring again to Eq. (5),

$$V_{fit} = \mathbf{X}^{T}V_{cam} , \tag{13}$$

where V_{cam} and V_{fit} are the 3-vectors before and after correction. The numerical matrix for the data just considered is:

$$\mathbf{X}^{T} = \begin{bmatrix} 0.97437 & 0 & -0.093857 \\ -0.070569 & 0.93879 & -0.13688 \\ 0.036341 & 0.07741 & 0.88096 \end{bmatrix} . \tag{14}$$

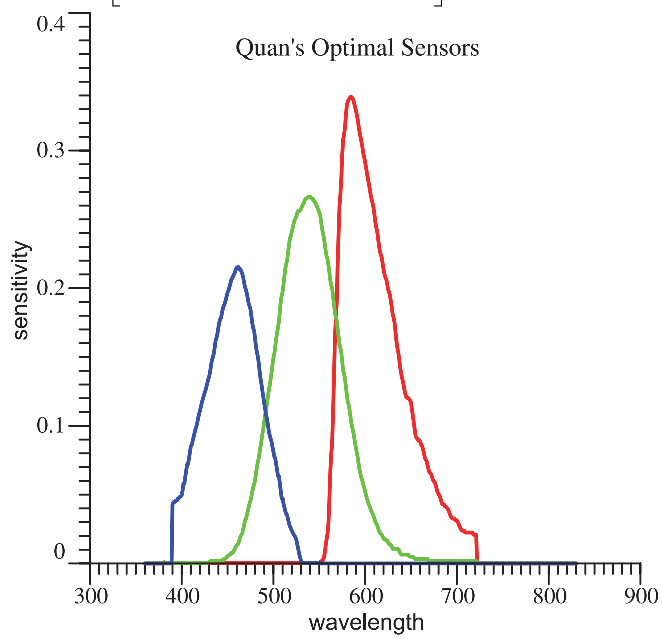

Figure 7. Quan's optimal sensors.

Quan's Optimized Sensitivities

Quan addressed the problem of optimal camera sensitivities, describing the problem as above: the camera should have red and green sensitivities narrower than the eye's in order to improve signal-to-noise [14]. After an analysis that considers lights and objects that the camera might encounter, he arrived at an optimal set of sensitivities, Fig. 7. These sensors can be analyzed as was the set above, Figs. 8 and 9. To the extent that the loci are different, the curve-fit adjustment does little. The camera's *r* and *g* sensors are narrower than for human, which apparently causes distortion as seen in Fig. 5 and to a lesser degree in Fig. 8.

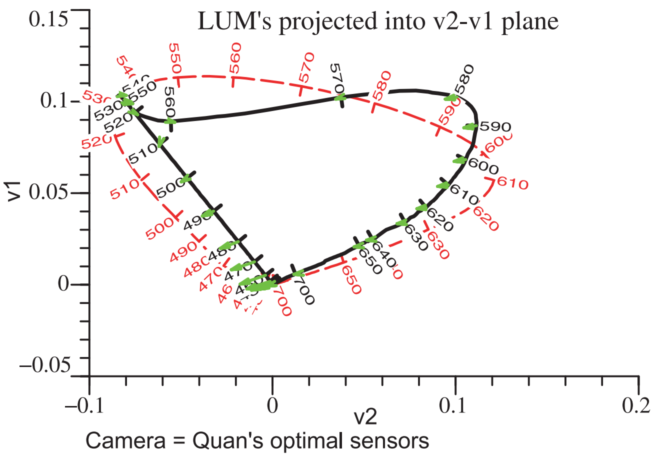

Figure 8. projection into the v_2-v_1 plane of the locus based on Quan's optimal sensors. Again, the dashed curve is the locus for the 2° standard observer.

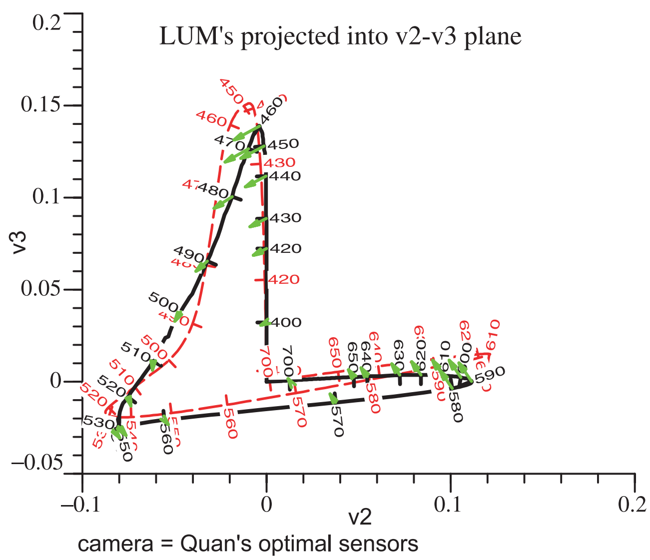

Figure 9. Projection into the v_2-v_3 plane of the locus based on Quan's optimal sensors.

Prime Color Wavelengths, etc

Again, the prime color wavelengths for the 2° observer are 603, 538, and 446 nm. The prime colors for the Fig. 2 camera are 595, 536 and 474 nm, while its longest vector wavelengths are 592, 536, 474 nm. For Quan's sensors, prime wavelengths are 585, 538, 461 and longest vectors are at 584, 539, 461. Ref. [13] reviews methods for finding the prime wavelengths. The red prime colors of both cameras speak to a mapping of reds different from human, seen in Figs. 5, 6, 8 and 9. To that extent, the naive hypothesis (above) is not supported.

Observation about noise

The results above relate color mixing to the vector space of a camera's sensor functions, expressed by the LUM. The sensors also affect a camera's signal-to-noise properties in a way that the LUM does not predict. In toying with the idea of anomalous color vision, Fig. 10 was generated. The left and right curves are green and red cones. The middle graphs show hypothetical "anomalous" cones, computed as mixtures of red and green.

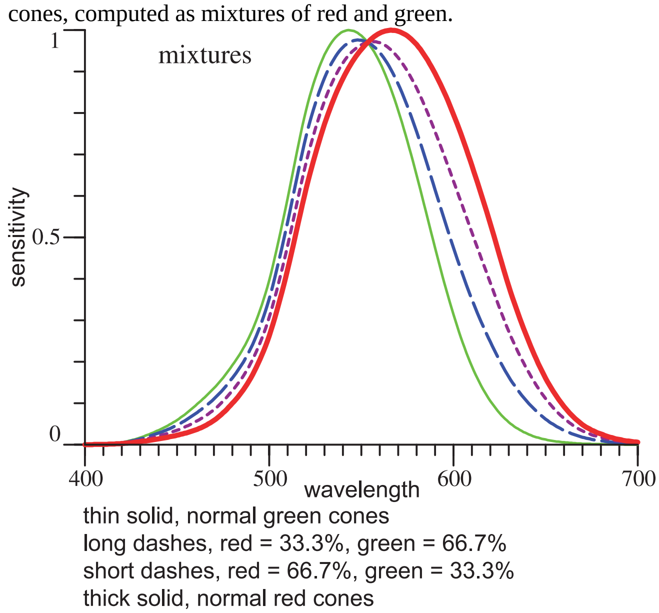

Figure 10. Red and green cones and 2 intermediate mixtues.

Now compare two hypothetical cameras, considering only their *r* and *g* sensors, represented as the columns of [|*r*⟩ |*g*⟩]. One camera has sensors with the exact sensitivities of human red and green cones, Fig. 1 or Fig. 10. The "anomalous" second camera has the same *g* sensors, but its *r* sensors have the sensitivity drawn with long dashes in Fig. 10. Those "red" sensors are only a little different from the green ones, but they must play the red role. The anomaly is assumed to be in the red sensors, not simulated by adding and subtracting. Both cameras satisfy MI, meaning that both have the orthonormal basis of Fig. 3, and both have the LUM shown for human in Figs. 5 and 8, where only red and green cones play a role. The first function is achromatic, the second is red-green, and there is no third. Doing Gram-Schmidt on $[|\bar{y}\rangle$, |normal $r\rangle]$ will give **Ω**. Noise arises in the sensors, including the quantum noises of detected photons and of dark current. Such camera sensitivities [|*r*⟩ |*g*⟩] can be transformed to orthogonal functions in a matrix **Ω** via a 2×2 matrix Y:

$$\mathbf{\Omega} = [|r\rangle\ |g\rangle]\mathbf{Y} \quad , \tag{15}$$

from which the required **Y** can be obtained by pre-multiplying by $\mathbf{\Omega}^{T}$:

$$\mathbf{Y} = (\mathbf{\Omega}^{T}\ [|r\rangle\ |g\rangle])^{-1} \quad . \tag{16}$$

For camera 1, with normal human sensors, when the cone amplitudes are adjusted as in Fig. 10,

$$\mathbf{Y}_1 = \begin{bmatrix} 0.0725 & 0.267 \\ 0.0447 & -0.310 \end{bmatrix} . \qquad (17)$$

For camera 2, the anomalous one,

$$\mathbf{Y}_2 = \begin{bmatrix} 0.212 & 0.782 \\ -0.100 & -0.843 \end{bmatrix} . \qquad (18)$$

For any light $|L(\lambda)\rangle$, we wish to relate signal and noise in a camera's signal V to the red and green signals R and G, and the noises N_R and N_G. Eqs. (15), (17) and (18) are used with the definition of V, Eq. (5). Let v_1, v_2 be the achromatic and red-green components of a signal V. Then the components for Camera 1 are:

$$v_1 = 0.0725\, R + 0.0447\, G . \qquad (19)$$

$$v_2 = 0.267\, R - 0.310\, G . \qquad (20)$$

The noise signals add in quadrature, giving the noises in Camera 1's two channels:

$$n_1(\text{Camera 1}) = [(0.0725\, N_R)^2 + (0.0447\, N_G)^2]^{1/2} \quad , \qquad (21)$$

$$n_2(\text{Camera 1}) = [(0.267\, N_R)^2 + (-0.310\, N_G)^2]^{1/2} . \qquad (22)$$

Now make a further simplifying assumption that $N_R = N_G = N$, for both cameras. Then, n_1(Camera 1) = $0.0851N$, n_2(Camera 1) = $0.409N$. By similar logic, n_1(Camera 2) = $0.235N$, n_2(Camera 2) = $1.1495N$. Two unsurprising conclusions are that the v_2 signal is noisier than the v_1 signal in both cameras, and the noise is more than doubled in each of v_1 and v_2 for Camera 2, because of the red sensor being so similar to the green. Signals v_1 and v_2 don't need to be evaluated; whatever they are, they are the same for both cameras, because $\mathbf{\Omega}$ is the same for both cameras.

One conclusion was stated above: two sets of receptor sensitivities can both satisfy the Maxwell-Ives criterion, but differ in their noise properties. Also, notice that the orthonormal basis played the role of a standard signal format. The signals v_1 and v_2 measure independent stimulus dimensions in a standardized way.

General conclusion

Our main message is that the Maxwell-Ives criterion can be expressed as a graphical comparison. A locus of unit monochromats is generated for the camera so its color-mixing properties can be compared to the eye's. The effect of a further adjustment can be added to the diagram. Even the remapping of example object colors by a camera's sensors could well be shown in Cohen's color space, though it was not done here. The method given above under "Implementation" is easy to apply. Using orthonormal bases for the eye and camera sensitivities eases the algebra. Our claim is not that any one equation above is original and undreamed of, but that the graphical method can be used in an open-ended way to aid sensor and camera design.

References

[1] Ives, Frederick E., "The optics of trichromatic photography," *Photographic Journal* **40**, 99-121 (1900).

[2] Brill, Michael H., Graham D. Finlayson, Paul M. Hubel, William A. Thornton, "Prime Colors and Color Imaging," 6th Color Imaging Conference: Color Science, Systems and Applications, Nov. 17-20, 1998, Scottsdale, Arizona, USA. Publ. IS&T, Springfield, Virginia.

[3] Worthey, James A., "Color matching with amplitude not left out," Proceedings of the 12th Color Imaging Conference: Color Science and Engineering, Scottsdale, AZ, USA, November 9-12, 2004. Published by IS&T, Springfield, VA 22151, http://www.imaging.org .

[4] Cohen, Jozef B. and William E. Kappauf, "Metameric color stimuli, fundamental metamers, and Wyszecki's metameric blacks," *Am. J. Psych.* **95**(4):537-564 (1982).

[5] Cohen, Jozef B. and William E. Kappauf, "Color mixture and fundamental metamers: Theory, algebra, geometry, application," *Am. J. Psych.* **98**(2):171-259, Summer 1985.

[6] Cohen, Jozef, *Visual Color and Color Mixture: The Fundamental Color Space*, University of Illinois Press, Champaign, Illinois, 2000.

[7] Vora, Poorvi L. and H. Joel Trussell, "Measure of goodness of a set of color-scanning filters," J. Opt. Soc. Am A 10(7):1499-1508, July 1993.

[8] Wandell, Brian A. and Joyce E. Farrell, Water into Wine: Converting Scanner RGB to Tristimulus XYZ, *Proceedings of SPIE* Vol.**1909**, pp. 92-100.

[9] Vora, Poorvi L. and H. Joel Trussell, "Mathematical methods for the analysis of color scanning filters," IEEE Transactions on Image Processing **6**(2):321-327, February 1997.

[10] Thornton, William A., "A simple picture of matching lights," *J. Illum. Eng. Soc.* **8**(2):78-85 (1979). Later articles revisit the theme of wavelengths where metamers are likely to cross.

[11] Worthey, James A., "Color rendering: a calculation that estimates colorimetric shifts," Color Res. Appl. **29**(1):43-56, February 2004.

[12] DiCarlo, Jeffrey M., Glen Eric Montgomery and Steven W. Trovinger, "Emissive chart for imager calibration," *Proceedings of the 12th Color Imaging Conference: Color Science and Engineering*, Scottsdale, AZ, USA, November 9-12, 2004. Published by IS&T, Springfield, VA 22151, http://www.imaging.org.

[13] Brill, Michael H., James A. Worthey, "Color Matching Functions When One Primary Wavelength is Changed," *Color Res. Appl., in press*.

[14] Quan, Shuxue, Evaluation and optimal design of spectral sensitivities for digital color imaging, Ph. D. dissertation, Rochester Institute of Technology, 2002.

Author Biography

James A. Worthey has a BS in Electrical Engineering, and an MS in Physics. His PhD in Physiological Optics is from Indiana University of Bloomington, Indiana. He is particularly interested in lighting and the interaction of lights with objects and the eye. He has published work on light source size, color rendering, object color metamerism and color constancy.