Prime colors and color imaging

M.H. Brill
Sarnoff Corporation
Princeton, NJ
USA

G.D. Finlayson
Univ. of Derby
Derby
UK

P.M. Hubel
Hewlett Packard
Palo Alto, CA
USA

W.A. Thornton
Prime Color Inc.
Cranford, NJ
USA

Abstract

In modeling color vision, certain visible wavelengths have special significance. A growing body of scientific work shows that the wavelengths around 450nm, 540nm and 605nm, the so called prime-color (PC) wavelengths, are fundamental to color vision. Perhaps unsurprisingly, these same wavelengths are often discussed in the color imaging literature. Monitors that can display a large gamut of colors and are visually efficient have phosphor-primary peaks at the PC wavelengths. Color cameras that have peak sensitivities at the PC wavelengths have favorable color-balancing properties. Why are the PC wavelengths so important? This paper provides a start toward a mathematical theory to answer this question.

1. Introduction

Suppose we acquire an arbitrary scene with a digital camera and display the captured image on a monitor. Ideally, ignoring issues of preference and rendering intent, the displayed image should be a visual match to the original scene. The goodness of the match depends on three things: the spectral characteristics of the camera, the spectral characteristics of the monitor, and the processing applied to the image. In this paper we consider the nature of *optimal* device characteristics and *optimal* processing.

In discussing optimal processing, we will repeatedly encounter the *prime-color* (PC) wavelengths, and so define them here. The three PC wavelengths are those at which unit-power monochromatic lights induce the largest tristimulus gamut (volume of the parallelepiped spanned by the tristimulus vectors of these lights). An older definition (in Part I of[1]), which was shown by Brill[2] to be equivalent to the above, is based on a color-matching experiment using these wavelengths for monochromatic primaries. Color-matching functions derived from any three monochromatic primary lights are such that, for each of the primary wavelengths, one color-matching functions is 1 and the other two are 0. Choosing the monochromatic primaries at the PC wavelengths additionally ensures that for each of the derived color-matching functions, at the primary wavelength for which its value is 1, it is also maximum (and maximum in absolute value). The older definition was based on an argument of visual efficiency: no more than one watt of a PC primary is needed to match one watt of any other wavelength. We prefer the new definition because the existence of prime-color wavelengths is clearer by the new definition.

Now it is important to understand what we mean by "optimum." In defining optimum three characteristics were considered: gamut size, efficiency, and balance.

<u>Gamut Size</u>

The optimal monitor should have a large color gamut; otherwise colorful scenes will be poorly reproduced. Two important issues relating to gamut are the range of saturations that can be displayed and the volume of the displayable color space (defined in terms of the colors we see). To maximize saturation we consider only the ideal of monochromatic monitor primaries, each triplet of which induces a volume in tristimulus space. Over all possible triplets of wavelengths we find that primaries anchored at approximately 450nm, 540nm and 605nm induce the largest gamut size. This agrees with the result one of us obtained earlier when computed in cone response space[3]. The wavelengths 450nm, 540nm and 605nm, the *PC wavelengths* have been shown to play an important role in many aspects of vision (including some of the aspects we discuss below). In this paper we give an mathematical explanation of *why* the PC wavelengths are so important.

<u>Efficiency</u>

Of course, large gamut size cannot be our only concern; energy efficiency must also be considered. If, for example, a monitor primary mixture of $10k$ Watts is required to match a physical (scene) stimulus of k Watts, then the monitor is visually inefficient. Visual inefficiency, by definition, implies a large power consumption (undesirable). More seriously, because the power output is bounded, it limits the display signal. The efficiency of this signal as compared to the noise (due to display artifacts and viewing conditions) limits the dynamic range of scenes that can be reproduced. This poses the question, 'which monochromatic set of primaries is the most visually efficient'. Again, the PC wavelength set was found to be optimal in this regard. That is, optimizing gamut size also optimizes visual efficiency and the scene dynamic range that can be displayed.

We now consider the spectral sensitivities of the camera that will drive our display. In the absence of noise, it is sufficient that the camera sensitivities are a linear transform from the color response, defined by the color matching functions, of our own visual system[4]; in this case the RGBs measured by the camera are linearly related to the required mixture coefficients. Unfortunately, the signal measured by the camera **is** confounded by the noise[5] and the linear matrixing operations increase this noise. Thus, it is advantageous to build a camera that sees exactly the required mixture coefficients without any matrixing. Classical colorimetry tells us that in this case the camera should have sensitivities that are matching functions for the monitor primaries.

<u>Balance</u>

A further complication is the role that the illumination spectral power distribution plays in our visual perception. It is well known that human observers have some degree of color constancy; that is, we see colors as more or less stable over a wide variety of illuminants. It is imperative then that this color constancy should be mimicked in the color reproduction process. The simplest, and most commonly used, method to discount the illumination is to divide the camera RGBs by the RGB for the illuminant (the RGB for putative white surface). Relative to this operation, it is clear that white will always be mapped to $(1, 1, 1)$ and so is always illuminant independent. It is less clear that dividing by white discounts illuminant color for other surface colors. Indeed, it is well known that white-balancing will work for unrestricted surface colors if and only if the camera has monochromatic sensitivities[6]. Failing this, the sensitivities should be as monochromatic as possible[7], but still be a linear combination of the color matching functions to avoid metamerism. We show that over all linear combinations of the color matching functions the set whose primaries are at the PC wavelengths behave most like a monochromatic sensor set. Relative to this particular set, dividing by white accurately discounts illuminant color bias for nearly all surface colors.

The reader may feel a little uncomfortable with the discussion about monochromatic primaries and sensors, but we believe our arguments tie in with practical color imaging. As an example, the dominant wavelengths for most monitor primary sets are at, or near, the PC wavelengths. The trade-off between good color balancing properties (in tri-band camera sensors) and low metamerism (found in color matching functions) results in a transformation of the matching functions that are as narrow as possible; these functions peak at the PC wavelengths.

The work that precedes this article focussed on providing experimental evidence for the importance of the PC wavelengths; a review of this work is given in section 2. The next three sections (sections 3-5) are a detailed analysis of the three main points: gamut size, visual efficiency, and color balance. In section 6, we place our work in the context of the large literature on PC wavelengths that is especially relevant to color imaging. The relationship between our idealized monochromatic primaries and those used in practice is made clear.

2. Visual system sensitivities and PC wavelengths

The importance of the PC wavelengths has been long reported in the color vision literature. W. D. Wright, in speaking of the three characteristic intersections of the spectral power distributions of lights (and speaking before the term prime color wavelengths was coined) that match to a normal human observer[8], states "...each crossing point tends to be located near to the three maxima of the sensitivity curves." Following Wright's lead, one of us made a 20-year study[9, 10] of the intersections of matching lights, concluding that the modal wavelengths of intersection lie near the PC wavelengths, and that these mark the spectral colors to which the normal human visual system responds most strongly (i.e., the peaks of the visual system sensitivities). It should be noted however, that these modal wavelengths depend on the particular color matching functions used. As an example, the modal wavelengths for the CIE 2 degree standard observer functions were found to be 447nm, 541nm and 604nm, and for the 10 degree standard observer, 446nm, 538nm and 600nm; modern data for six human observers show 450nm, 533nm, and 611nm. In this paper, we refer to the CIE 2 degree observer PC wavelengths.

In 1975[11](Part I) Thornton discovered that transformation of primaries of either the 1931 or 1964 CIE Standard Observer to real primaries that coincide in wavelength with the peaks of the resulting color-matching functions results in the same three wavelengths; such coincidence signifies that the resulting three spectral lights require minimum power content in visual matches in which they occur, i.e., they invoke maximum visual-system response per watt. White lamplight composed of the prime colors was shown to afford high visual efficiency and good color-rendering[1], as well as a gamut of coloration exceeding that of daylight of the same color[12]. The chromaticity of an element in any visual scene is established with minimum power input to the eye when light from the element is composed of a mixture of spectral colors near the PC wavelengths[13]. Systems of color television, color photography, and colorants (inks, paints and dyes) were proposed[14], the latter for reasons of improved color-constancy[15, 16, 17] as well as large gamut of color. Three articles discuss the relation between PC wavelengths and peak system sensitivities[18, 19, 20]. Part I of [11] shows that the prime colors are resident in both the 1931 and 1964 CIE Standard Observer data. The remaining Parts II - VI of [11] make clear, however, that the CIE color-matching functions are not to be relied upon as weighting functions in color imaging, unless (1) metamerism is very weak indeed, or (2) all of the viewed lights involved are composed predominantly of PC wavelength components; in those cases, tristimulus errors by one or other of the CIE Standard Observers will be relatively small, but not zero. Finally, peak sensitivities of the normal human visual system are shown[21] to be slightly but importantly different from those of either CIE Standard Observer, and are to be found near the following wavelengths: 450 +/- 1nm, 533 +/- 1nm, and 611 +/- 3nm.

In the context of the current article one might wonder whether any of the various proposed ”cone functions” lead properly to the sought three sensitivity functions of the normal human visual system, to be used in color imaging. In 1974 Stiles and Wyszecki made a Herculean effort[22] to show the opposite: that three acceptable absorption curves of visual pigments could be inferred from the CIE color-matching data. Stiles had shown in 1953 that the longwave spectral sensitivity of Pitt, as well as that of Stiles himself, ”could not correspond to absorption by a single visual pigment of the rhodopsin type...” The 1974 work agrees, and concludes that the CIE color-matching data demand ”...a peak wavelength for the ’red’ sensitive pigment at 617nm which other evidence shows to be much too far in the red” and ”...a completely unacceptable transmission curve representing light losses in the eye prior to visual absorption.”[22] One of us suggested[18] in 1978 that this indicated ’red’ peak position near 617nm (unacceptable as a rhodopsin absorption), as well as the acceptable peaks near 432nm and 535nm, represent ”peak system responses.” It follows that the valid peak visual-system responses, needed in color imaging, cannot be expected to be derivable from ”cone functions” of the type that have been proposed.

3. Gamut size

Let $C(\lambda)$ denote a spectral power distribution (SPD) of light that enters the visual system. Assuming the visual system is trichromatic with sensitivities proportional to the CIE XYZ standard observer matching functions. The visual system responds linearly as follows:

$$\begin{aligned} x &= \int_\omega X(\lambda)C(\lambda)d\lambda \\ y &= \int_\omega Y(\lambda)C(\lambda)d\lambda \\ z &= \int_\omega Z(\lambda)C(\lambda)d\lambda \end{aligned} \tag{1}$$

where $X(\lambda), Y(\lambda)$ and $Z(\lambda)$ are the color matching functions for the standard observer and ω represents the visible spectrum (roughly 400 to 700nm).

By sampling each spectral function every 5nm it is possible to rewrite equation (1) in the notation of vector algebra. Let $\underline{C}$ denote the 61-vector (61 samples across the visible spectrum) corresponding to $C(\lambda)$. Similarly $\underline{X}$, $\underline{Y}$ and $\underline{Z}$ denote vector approximation of the three standard observer functions. For ease of notation we group these three vectors into the three columns of the 61×3 matrix $\mathcal{R}$. The vector $\underline{\rho} = [x \;\; y \;\; x]^t$ is equal to:

$$\underline{\rho} \;=\; \mathcal{R}^t \underline{C} \tag{2}$$

A typical monitor has three primaries and mixtures of the primaries are set to match scene colors. Let the column of the 61×3 matrix $\mathcal{P}$ contain the three spectral power distributions for a three primary monitor. The relationship between color mixture, defined by the 3-vector $\underline{\alpha}$, and color response is equal to:

$$\underline{\rho} \;=\; \mathcal{R}^t \mathcal{P} \underline{\alpha} \tag{3}$$

To address gamut size, we must look at (3) in more detail. We begin by pointing out that because the power of light is non negative, the entries in $\underline{C}$, $\mathcal{P}$ and $\underline{\alpha}$ must be positive. The standard observer sensitivity functions $\mathcal{R}$ are also all positive. We will assume that all color signal spectra, including the monitor phosphors, have bounded power. That is,

$$\int_\omega C(\lambda)d\lambda \leq k \tag{4}$$

without loss of generality we set $k = 1$. Rewriting (4), in discrete terms:

$$\sum_{i=1}^{61} C_i \leq 1 \tag{5}$$

Again without loss of generality, we assume that the columns of $\mathcal{P}$ sum to one (the power of the primary mixtures is bounded). For now, we will assume that the coefficient terms in $\underline{\alpha}$ are positive and sum to one. Thus, all mixtures have at most unit power and so our monitor has bounded power output.

A measure of the range of colors we can see is the volume gamut subtended in tristimulus space. For an additive system such as a CRT, this volume is proportional to that of the tetrahedron subtended by the color primaries (at maximum output power). Figure 1 shows the tetrahedron formed when the three vectors corresponding to the three primaries intersect the spectral locus. Let the function $v(\mathcal{M})$ return the volume spanned by the convex combinations of the columns of a matrix $\mathcal{M}$. Since $\underline{\alpha}$ is the positive convex vector of mixture coefficients of $\mathcal{R}^t\mathcal{P}$, the volume of the monitor gamut is equal to:

$$\text{monitor-gamut-volume} = v(\mathcal{R}^t\mathcal{P}) \tag{6}$$

To find the largest monitor gamut we need to maximize the expression in equation 6.

We must also ensure that our optimal monitor can display a reasonable range of saturated colors. The most saturated colors that we can see are monochromatic, so we restrict our attention to monochromatic primaries (later, we will consider whether this was really a good thing to do).

If M is a $3 \times m$ matrix then the $M_{i,j,k}$ is the 3×3 matrix comprising the ith, jth and kth columns of $\mathcal{M}$. The 61×3 matrix $\mathcal{P}^{i,j,k}$ denotes a monochromatic primary matrix such that all entries of $\mathcal{P}$ are 0, save the ith, jth and kth rows; $[\mathcal{P}^{i,j,k}]^t_{i,j,k} = \mathcal{I}$ (the 3×3 identity matrix). Clearly,

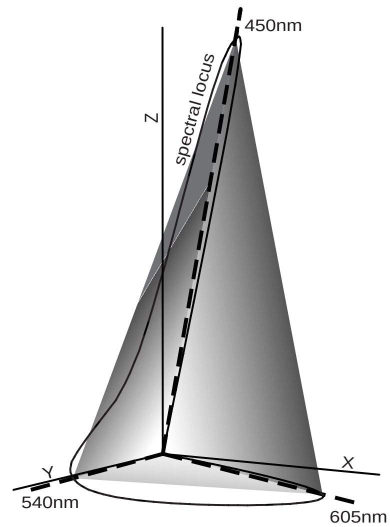

Figure 1: In any tristimulus space the three vectors that intersect the spectral locus at the prime-color wavelengths give maximal gamut volume.

$$\mathcal{R}^t\mathcal{P}^{i,j,k} = \mathcal{R}^t_{i,j,k} \tag{7}$$

The maximum gamut monitor is found by maximizing:

$$v(\mathcal{R}^t_{i,j,k}) = (1/6) * |\mathit{determinant}(\mathcal{R}^t_{i,j,k})| \quad \mathit{over} \quad i,j,k = 1,2,\cdots,61 \tag{8}$$

We found that $i = 11$, $j = 29$ and $k = 42$ (corresponding to wavelengths 450nm, 540nm and 605nm) maximized (8). It follows that the optimal monitor should have primaries anchored at these wavelengths.

So far we have carried out our analysis in tristimulus space to develop our argument, but the argument carries over easily to any tristimulus space, (such as cone response space[3]), including some opponent-color spaces. This generalization can be made because of the following mathematical fact:

It is a classical result in linear algebra[23] that the volumes of two regions that are a linear transform apart are related by the volume of the linear transform:

$$v(\mathcal{TM}) = v(\mathcal{T}) * v(\mathcal{M}) \tag{9}$$

It follows from this fact that the optimization of Equation 6 applies to any basis transformation of tristimulus space, because the transformation incurs only a constant multiplier on the tristimulus volume subtended by $\mathcal{R}^t\mathcal{P}$.

4. Visual efficiency

We say that a monitor is visually efficient if a stimulus of k Watts can be matched by primary mixture stimuli of no more than αk Watts, where α is close to one. Moreover, each individual mixture coefficient is constrained to be less than or equal to 1 (since we are assuming output power is bounded).

From equation (3) the mixture $\underline{\alpha}$ required to match a response vector $\underline{\rho}$ is equal to:

$$[\mathcal{R}^t\mathcal{P}]^{-1}\underline{\rho} \;=\; \underline{\alpha} \qquad (10)$$

The color matching functions for primaries $\mathcal{P}$ are the mixture coefficients required to match each monochromatic stimulus across the visible spectrum. The matching functions $\mathcal{F}^t$ are a linear transform of the standard observer functions:

$$\mathcal{F}^t \;=\; [\mathcal{R}^t\mathcal{P}]^{-1}\mathcal{R}^t \qquad (11)$$

The matching curves for the PC wavelength monitor are shown in Figure 2. By visual inspection, it is clear that the largest absolute value of the matching curves is 1 and so the maximum mixture coefficient, needed to match any monochromatic wavelength of light, is also 1.

In the context of this paper we assume that scene stimuli have bounded power. What then is the maximum power, per primary channel, required to match a non-monochromatic stimuli? The ith mixture coefficient is the average value of the i matching function weighted by the color signal spectrum $\underline{C}$:

$$\alpha_i \;=\; \sum_{j=1}^{61} C_j \mathcal{F}_{ji} \qquad (12)$$

To make α_i large, $\underline{C}$ should have maximum power. Without loss of generality let the magnitude of $\underline{C}$ equal 1. In this case we can interpret $\underline{C}$ as being a probability distribution and α_i is the expected value (or weighted average) of the ith color matching function. By definition the weighted average of a distribution of numbers must fall between the maximum and minimum of the distribution. It follows then that the mixture coefficients are

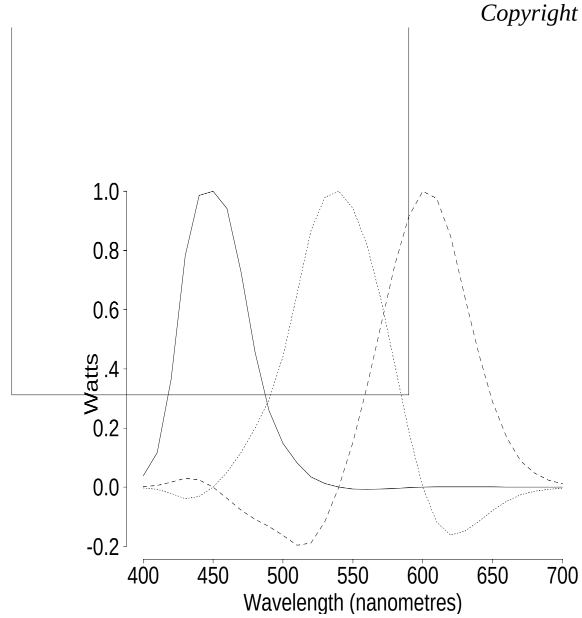

Figure 2: Color matching functions derived from primaries placed at wavelengths 450nm, 540nm and 605nm (assuming visual system sensitivities are linearly related to CIE 1931 2^o matching curves)

bounded by the absolute maximum of the matching functions. That is, one watt or less of each primary suffices to match all color signal spectra of one watt or less.

Based on this suggestion of efficiency of the PC wavelengths, it is plausible to conjecture that the most watt-efficient metamer of any tristimulus vector in the PC wavelength gamut is a linear combination of the PC wavelengths. Strictly speaking, this conjecture is false. Counterexamples can be constructed as follows (see figure 3). For each wavelength λ, find the wattages for the following color-match: a positive combination of two PC wavelengths (prime-color side of match), and 1 watt of wavelength λ plus a positive wattage of the third PC wavelength (non-prime-color side of the match). Each chromaticity so composed is on a leg of the prime-color triangle in chromaticity space. The wattages can be read directly from the color-matching functions in Figure 2. The primary whose wattage is negative is the ”third prime color” and is part of the non-prime side of the match. The other two primaries are on the prime side of the match. Now define ”relative nonprime efficiency” as the power on the prime side of the match, divided by power on the nonprime side. Values of λ for which this efficiency is greater than 1 are counterexamples to the conjecture. Figure 4 shows a plot of this efficiency as a function of λ. It can be seen that the value 1.175 at 570nm is the maximum value of the nonprime efficiency (a clear counterexample), but that for most wavelengths this efficiency

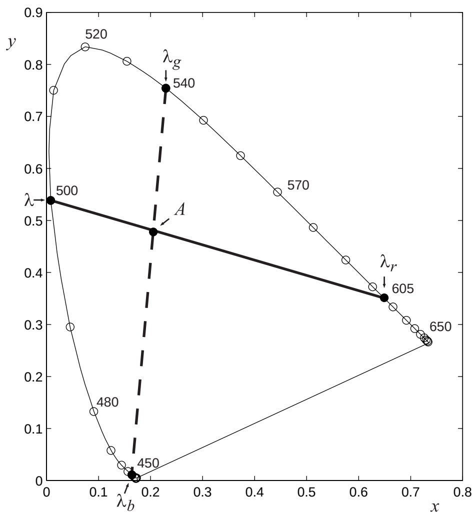

Figure 3: *Two metamers with chromaticity A, one composed of energy at the green and blue prime-color wavelengths (connected by a dashed line) and the other composed of energy at the red prime-color wavelength and a non-prime wavelength (connected by a solid line). The dashed line denotes the prime side of match, and the solid line denoted the non-prime side.*

is less than 1. The tendency here for the prime-color side of the match to require fewer watts than the nonprime side confirms the authors' experience that exceptions to this rule tend to be neither very strong nor very numerous. The PC wavelengths tend to be close to maximal in watt-efficiency.

Hence the PC wavelength monitor is visually efficient, not only per primary but in toto. This is an important observation. In maximizing monitor gamut size we restricted ourselves to monochromatic primaries and primary mixtures of 1 watt or less. It is straightforward to show that if the maximum wattage required to match a color signal stimulus is always less than one watt then the gamut volume is maximized by a monitor with monochromatic primaries. That is, we do not need to appeal to saturation in order to justify our choice of monochromatic primaries.

Are these rather nice properties a matter of chance or is there some reason why the monitor that maximizes gamut size should also be visually efficient? In fact, chance is not at work here, but rather maximum gamut size implies visual efficiency.

To see that this is so, let us begin with an arbitrary color sensitivity functions $\mathcal{Q}'$ (perhaps based on an alternate (non CIE) standard observer). Let the wavelengths

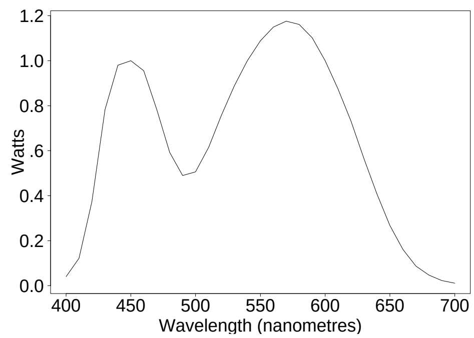

Figure 4: *Relative non-prime efficiency: total power of prime mixture divided by power of matching stimulus lying on the gamut boundary (formed by mixing a monochromatic light with one of the PC wavelengths).*

a, b, c induce the largest gamut. Let $\mathcal{Q}$ denote the color matching functions associated with monochromatic primaries anchored at wavelengths a, b, c. It follows that $\mathcal{Q}^t_{a,b,c} = \mathcal{I}$. Let us now suppose that $\mathcal{Q}$ is not visually efficient: there exists an entry in the dth row of $\mathcal{Q}$ (d is a wavelength other than a, b or c) that is bigger than 1 (breaking our bounded power constraint).

It follows that the matrix $\mathcal{Q}^t_{a,b,d}$ has the form:

$$\begin{bmatrix} 1 & 0 & x \\ 0 & 1 & x \\ 0 & 0 & 1+\epsilon \end{bmatrix} \quad (\epsilon > 1) \qquad (13)$$

where x denotes a dummy variable and we have placed a value greater than 1 in the third matching function. It is easy to show that the determinants of $\mathcal{Q}^t_{a,b,c}$ and $\mathcal{Q}^t_{a,b,d}$ are equal to 1 and $(1 + \epsilon)$ respectively. It follows then that $v(\mathcal{Q}^t_{a,b,d}) > v(\mathcal{Q}^t_{a,b,c})$. This cannot be the case since a, b, c are those primary wavelengths that maximize gamut size. We have a contradiction and so maximum gamut size *does* imply visual efficiency.

As noted in the introduction, the above result was originally derived by looking at color matching and visual efficiency without considering gamut size[11]. Once the connection is seen between the volume-gamut and visual-efficiency definitions of PC wavelengths, the interchanging of definitions can be useful. Maximizing over all three unit-power monochromatic lights the volume subtended in tristimulus space, is an effective means for *finding* the PC wavelengths. This result is significant in connecting several appearances of the PC wavelengths in color science and technology. Moreover, in the context of this paper, this result is now seen to link with the question of gamut size.

5. Color balance and camera sensitivities

The color matching functions $\mathcal{F}$ for the PC wavelength primaries are functions of wavelength and interact with color signal spectra in forming a mixture coefficient vector in the same way that the standard observer functions interact with the color signal spectra in forming a tristimulus ($\underline{\alpha} = \mathcal{F}^t\underline{C}$ and $\underline{\rho} = \mathcal{R}^t\underline{C}$). It follows then that the matching functions can be regarded as a set of visual sensitivities.

Thinking of matching functions in this way helps instruct how to build a camera to drive the PC wavelength monitor. By definition, the color matching functions *see* the required mixture coefficients. That is, $\mathcal{R}^t\mathcal{P}\mathcal{F}^t\underline{C} = \mathcal{R}^t\underline{C}$. If instead a camera had standard observer sensitivities then the 3×3 linear transform $[\mathcal{R}^t\mathcal{P}]^{-1}$ needs to be applied to the camera response to recover the mixture coefficients (see equation 13). Avoiding the need for a matrixing step is important since image noise increases under linear transformation.

In designing a color camera it is important to consider the role played by illumination. Under different illuminants a color camera records signals that have color shifts as compared to the perceived color of the scene being rendered. Thus the image must be color balanced prior to display. This balance operation is one of the fundamental properties of color appearance (referred to as chromatic adaptation). The simplest, and most widely used method, for discounting the illuminant is to map the camera RGB measurements by scale factors inversely proportional to the response for a perfect white diffuser. This operation renders white equal to $(1, 1, 1)$ under all illuminants. But does dividing by white also remove illuminant bias for other surfaces?

Let us examine this problem in more detail. First, we express the color signal $C(\lambda)$ as a product of illumination, $E(\lambda)$ and reflectance $R(\lambda)$: $C(\lambda) = E(\lambda)R(\lambda)$. Now let us assume that $X(\lambda) = \delta(\lambda - \lambda_X)$; that is the long wave sensor in (1) is a delta function anchored at wavelength λ_X (a delta function is non zero only at the anchor wavelength). Similarly the medium and short-wave standard observer responses are the delta functions $\delta(\lambda - \lambda_Y)$ and $\delta(\lambda - \lambda_Z)$. Rewriting equation (1):

$$\begin{aligned} x &= \int_\omega \delta(\lambda - \lambda_X)E(\lambda)R(\lambda)d\lambda = E(\lambda_X)R(\lambda_X) \\ y &= \int_\omega \delta(\lambda - \lambda_Y)E(\lambda)R(\lambda)d\lambda = E(\lambda_Y)R(\lambda_Y) \\ z &= \int_\omega \delta(\lambda - \lambda_Z)E(\lambda)R(\lambda)d\lambda = E(\lambda_Z)R(\lambda_Z) \end{aligned} \quad (14)$$

Notice with respect to delta functions the integrals vanish. Sensor response is equal to the color signal at the anchor wavelength.

The response of a perfect white diffusing surface: ($R(\lambda) = 1$) is $(E(\lambda_X),\ E(\lambda_Y), E(\lambda_Z))$ and so it follows that dividing by white cancels illumination for all reflectances:

$$\begin{aligned} \frac{E(\lambda_X)R(\lambda_X)}{E(\lambda_X)} &= R(\lambda_X) \\ \frac{E(\lambda_Y)R(\lambda_Y)}{E(\lambda_Y)} &= R(\lambda_Y) \\ \frac{E(\lambda_Z)R(\lambda_Z)}{E(\lambda_Z)} &= R(\lambda_Z) \end{aligned} \quad (15)$$

It has been shown that, for unrestricted reflectance, white-balancing exactly discounts illumination if and only if delta functions are used[6]. Indeed, delta functions are the only type of sensor for which a linear model of illumination change is justified[24]. Camera sensitivities made up of delta functions would, of course, have unacceptable levels of metamerism. So, for the purposes of discounting illumination we wish to transform the matching functions to behave more like delta functions.

Let us develop an error measure of the closeness of the matching functions, or linear transforms thereof, to delta functions. If $\mathcal{A}$ and $\mathcal{B}$ are matrices of device sensitivities, then the best linear transform $\mathcal{T}$ that minimizes $|\mathcal{B}\mathcal{T} - \mathcal{A}|$, in a least-square sense, is defined by the Moore Penrose inverse: $\mathcal{T} = [\mathcal{B}^t\mathcal{B}]^{-1}\mathcal{B}^t\mathcal{A}$. Let $\mathcal{P}^{i,j,k}$ denote a 61×3 matrix of device sensitivities where the first, second and third columns contain delta functions anchored at wavelengths i, j and k. It is easy to show that $[\mathcal{P}^{i,j,k}]^t\mathcal{P}^{i,j,k} = \mathcal{I}$. $\mathcal{P}^{i,j,k}$ mapped to the standard observer sensitivities $\mathcal{R}$ in a least-squares sense is equal to:

$$\mathcal{P}^{i,j,k}[\mathcal{P}^{i,j,k}]^t\mathcal{R} \approx \mathcal{R} \quad (16)$$

If $|\mathcal{P}^{i,j,k}[\mathcal{P}^{i,j,k}]^t\mathcal{R}|$ is close to $|\mathcal{R}|$ then it follows that the set of delta functions $\mathcal{P}^{i,j,k}$ samples light similarly to the standard observer (and vice versa). Note that we are not comparing (differencing) the sensor sets but rather determining the closeness of the sensor sets by examining the closeness of their respective magnitudes. At first glance such an approach appears bizarre. However, it is justified because the two sensor sets are related by a least-squares fit. As the fit gets better the magnitudes must converge.

To calculate $|.|$ we begin by calculating the covariance ellipsoid associated with $\mathcal{P}^{i,j,k}[\mathcal{P}^{i,j,k}]^t\mathcal{R}$

$$cov(\mathcal{M}) = \frac{\mathcal{M}^t\mathcal{M}}{61} \tag{17}$$

and

$$cov(\mathcal{P}^{i,j,k}[\mathcal{P}^{i,j,k}]^t\mathcal{R}) = \frac{\mathcal{R}^t\mathcal{P}^{i,j,k}[\mathcal{P}^{i,j,k}]^t\mathcal{R}}{61} \tag{18}$$

If $\mathcal{R}$ samples light like a particular delta function set then the magnitude of its covariance matrix should approximate the magnitude of the covariance matrix of the standard observer functions themselves. That is,

$$|\mathcal{R}^t\mathcal{P}^{i,j,k}[\mathcal{P}^{i,j,k}]^t\mathcal{R}| \approx |\mathcal{R}^t\mathcal{R}| \tag{19}$$

Using the volume function $v()$ (defined at Equation 6) as our measure of magnitude $|.|$, we seek to maximize:

$$\frac{v(\mathcal{R}^t\mathcal{P}^{i,j,k}[\mathcal{P}^{i,j,k}]^t\mathcal{R})}{v(\mathcal{R}^t\mathcal{R})} \tag{20}$$

Inserting identity (7) and using (9) this simplifies to

$$\frac{v(\mathcal{R}^t_{i,j,k})^2}{v(\mathcal{R}^t\mathcal{R})} \tag{21}$$

For all triplets of narrow-band functions, $v(\mathcal{R}^t\mathcal{R})$ is constant. It follows then that (21) is maximized when (8) is maximized. That is the volumetric argument that delivers the monitor primaries that maximizes gamut size also delivers the narrow band sensors that behave proportionally most like the human visual system. An attractive feature of this argument is that the selection of narrow band sensors is not contingent on the particular basis (linear transform) of the matching functions used.

However, as a final step we must actually tie down the particular basis that behaves like delta function. The argument, set forth in (14) and (15) is basis specific. From (16), $\mathcal{P}^{i,j,k}[\mathcal{P}^{i,j,k}]^t\mathcal{R} \approx \mathcal{R}$. It follows then that

$$[\mathcal{P}^{i,j,k}]^t \approx [\mathcal{R}^t\mathcal{P}^{i,j,k}]^{-1}\mathcal{R}^t \tag{22}$$

That is the optimal camera functions, from a color balancing perspective, are color matching functions for PC wavelengths primaries; these are shown in Figure 2.

Of course the functions in figure 2 are still far from being narrow band and so we might wonder whether they really behave like narrow band sensors. Simulations[25] have shown that this is in fact the case.

6. This work in context

The PC wavelengths have been reappearing for some time in a wide range of literature relating to colorimetry.

The chromaticities of the phosphors and of the white point prescribed by standards bodies (for television technology, computer imaging, digital photography, and other fields) show salience of the PC wavelengths. This salience is to be seen by examining the dominant wavelength of each phosphor–i.e., the wavelength obtained by extrapolating the line from the white point through the primary until it is incident on the spectrum locus. This construction is easily done using the data and figure in Poynton's book[26]. The following phosphor sets are included there: NTSC primaries (developed in 1953 now obsolete); the current ITU-R BT.709 standard; and the SMPTE 240M and EBU standards.

It is remarkable that, in every case, the dominant wavelengths of the phosphor primaries are very close to the PC wavelengths, except for a discrepancy between the NTSC and the other standards (especially in the green). In addition to gamut size and visual efficiency discussed in previous sections the above salience has theoretical significance related to chromatic adaptation: scaling the signals from the primaries is equivalent to a Von Kries transformation of the tristimulus values using color-matching functions that would have been associated with the primaries by a color-matching experiment. Hence the white-point corrections are calculated in a tristimulus basis for which this kind of adaptation ensures the greatest color constancy. In fact, Hubel and Finlayson[27] have given psychophysical evidence that chromatic adaptation using the Sharp transformations gives better correlation to chromatic adaptation in the visual system than either the cone sensitivities or CIE tristimulus functions (as used in color appearance models). They also showed that matching functions derived using the Sharp transformations were almost identical to matching functions derived from a standard set of primaries[28].

Other literature also reveals salience of the PC wavelengths. MacAdam[29] noted that the moment per watt of a monochromatic light (defined as the power of the complement needed to neutralize one watt of the light) is greatest when the wavelength of the monochromatic light is at 448 or 605nm; Thornton[1] continued the complementary lights into the purple (two monochromatic lights) and found a maximum moment-per-watt at the green PC wavelength (540nm). Wright[8] noted that natural metameric reflectance spectra tend to cross each other near three particular wavelengths, which Thornton[9, 10] later recognized as the PC wavelengths. By using laser lines near the PC wavelengths Hubel[30] recorded a color reflection hologram which replayed a color three dimensional scene by diffraction of incident white-light into PC wavelength components. A theoretical study extending Thornton's work and applying it to holography also gave wavelengths near the PC wavelengths as optimum for both gamut area and color error[31]. Upon examining the first four principal components of daylight, Brill[32] found that the linear combination of these components that is orthogonal to all the color-matching functions has its zero-crossings near the PC wavelengths. Neugebauer[33] found a similar salience of the PC wavelengths when he examined the least-square residual of a narrow-band camera sensitivity function compared with the best linear combination of the CIE 1931 color-matching functions. When plotted as a function of the dominant wavelength of the camera-sensitivity function, the residual shows three distinct minima near the PC wavelengths.

The PC wavelengths have also found other technological applications besides color imaging. For example, Thornton's design of fluorescent lamps with concentrations of light at the PC wavelengths has led to substantial commercial success because of the superior color-rendering properties as well as their visual efficiency. For example, metamerism is minimized by confining light to these wavelengths, because naturally metameric reflectances tend to cross at these wavelengths. Finally, a design of glasses with transmission mainly near these wavelengths (Thornton, [34]) allows restriction of the total radiation into the eye without compromising vision.

7. Conclusion

All this evidence points to the importance of the prime-color wavelengths for color imaging and our understanding of the visual system. In this paper we have used the maximum-volume proof to bring a formal definition to prime-color wavelengths and shown that these also imply visual efficiency and an optimal space for color balancing operations. Using such a framework, we can begin to understand how these and many other seemingly disparate concepts are interrelated by the prime-color wavelengths. There is clear benefit to choosing the primaries of color imaging systems at the prime-color chromaticities.

Acknowledgements

The support of Hewlett Packard and Sarnoff corporations is gratefully acknowledged. We would also like to thank Jack Holm, Steve Hordley, and Steve Marschner for helpful conversations and assistance.

References

[1] W.A. Thornton. Luminosity and color-rendering capability of white light. *J. Opt. Soc. Amer.*, pages 191–194, 1971.

[2] M.H. Brill. A theorem on prime-color wavelengths. *Color Res. Appl.*, 21:239–240, 1996.

[3] P.M. Hubel. *Color Reflection Holography*. PhD thesis, University of Oxford, Department of Engineering Science, 1990.

[4] F.E. Ives. The optics of trichromatic photography. *Photographic Journal*, 40:99–121, 1900. reprinted in D. L. MacAdam, Ed., Sources of Color Science, MIT Press, 1970, pp. 127-133.

[5] A.K. Juenger. Color sensitivity selection for electronic still cameras based on noise considerations in photographic speed maximization. In *Image Processing, Image Quality, Image Capture Systems Conference*, pages 79–83. IS&T, 1998.

[6] J.A. Worthey and M.H. Brill. Heuristic analysis of von Kries color constancy. *Journal of The Optical Society of America A*, 3(10):1708–1712, 1986.

[7] G.D. Finlayson, M.S. Drew, and B.V. Funt. Spectral sharpening: Sensor transformations for improved color constancy. *J. Opt. Soc. Am. A*, 11(5):1553–1563, May 1994.

[8] W.D. Wright. *The Measurement of Colour*. Reinhold Co., 1967. 4th Edition.

[9] W.A. Thornton. Matching lights, metamers and human visual response. *Journal of Color and Appearance*, 2(1):23–29, 1973.

[10] W.A. Thornton. Intersections of spectral power distributions of lights that match. *Color Res. Appl.*, 18:412–421, 1993.

[11] W.A. Thornton. Toward a more accurate and extensible colorimetry. *Color Res. Appl.* Part I, 17, 2, 79-122 (1992); Part II, 17, 3, 162-186 (1992); Part III, 17, 4, 240-262 (1992); Part IV, 22, 3, 189-198 (1997); Part V, 23, 2, 92-103 (1998); Part VI, 23, 4, 226-233 (1998).

[12] W.A. Thornton. Color-discrimination index. *J. Opt. Soc. Amer.*, pages 191–194, 1972.

[13] W.A. Thornton. Three-color visual response. *J. Opt. Soc. Amer.*, pages 457–459, 1972.

[14] W.A. Thornton. Relating commercial uses of color to human color vision. *Westinghouse ENGINEER*, 1972.

[15] W.A. Thornton. The design of safety colors. *J. Illum. Engineering Soc.*, pages 92–99, 1977.

[16] W.A. Thornton. The psychophysics of image coloration. *Photographic Science and Engineering*, pages 102–107, 1978.

[17] W.A. Thornton. Improving color constancy of object colors. *Color Res. Appl.*, 11:278–286, 1986.

[18] W.A. Thornton. Reply to Ohta-Wyszecki on location of nodes of metameric stimuli. *Color Res. Appl.*, 3:202–204, 1978.

[19] W.A. Thornton. Note on visual responses: system vs. retinal. *Color Res. Appl.*, 11:176–177, 1986.

[20] W.A. Thornton. Evidence for the three spectral responses of the normal human visual system. *Color Res. Appl.*, 11:160–163, 1986.

[21] W.A. Thornton. Spectral sensitivities of the normal human visual system, color-matching functions and their principles, and how and why the two sets should coincide. *Color Res. Appl.* In press.

[22] W.S. Stiles and G. Wyszecki. Colour-matching data and the spectral absorption curves of visual pigments. *Vision Research*, 14:195–207, 1974.

[23] J.B. Fraleigh and A.R. Beauregard. *Linear Algebra*. Addison Wesley, 3rd edition, 1995.

[24] D. Forsyth. A novel algorithm for color constancy. *Int. J. Comput. Vision*, 5:5–36, 1990.

[25] G.D. Finlayson and B.V. Funt. Coefficient channels: Derivation and relationship to other theoretical studies. *Color Res. Appl.*, 21(2):87–96, 1996.

[26] C.A. Poynton. *A Technical Introduction to Digital Video*. Wiley, 1996.

[27] P.M. Hubel and G.D. Finlayson. Sharp transforms for color appearance. In *Device-Independent Color*, volume 3300. SPIE, 1998.

[28] P.M. Hubel, J. Holm, and G.D. Finlayson. Illuminant estimation and color correction. In *CIM'98 Colour Imaging in Multimedia*, pages 97–105. University of Derby, 1998.

[29] D.L. Macadam. Photometric relationships between complementary lights. *J. Opt. Soc. Am.*, 28:103–111, 1938.

[30] P.M. Hubel and A.A. Ward. Color reflection holography. In *Pratical Holography III*, volume 1051, pages 18–24. SPIE, 1989.

[31] P.M. Hubel and L. Solymar. Color-reflection holography: theory and experiment. *Applied Optics*, 30(29):4190–4202, 91.

[32] M.H. Brill. Statistical confirmation of Thornton's zero-crossing conjecture for metameric blacks. *Color Res. Appl.*, 12:51–53, 1987.

[33] H.E.J. Neugebauer. Quality factor for filters whose spectral transmittances are different from color mixture curves, and its application to color photography. *J. Opt. Soc. Am.*, 46:824–828, 1956.

[34] W.A. Thornton. Filter with threee-band transmission for good seeing. 1989. U.S. Patent 4,826,286.