Converting RGB Images to LMS Cone Activations

Judah B. De Paula
Department of Computer Sciences
The University of Texas, Austin, TX 78712-0233
judah@cs.utexas.edu

February 22, 2005

Abstract
A process for converting color RGB images to LMS photoreceptor activations.

1 Introduction

This article describes a method to convert an RGB image, where each triple represents phosphor luminosity in a CRT monitor (a.k.a. Cathode Ray Tube, or CRT,) to an LMS triple where each value simulates the retina cone activity viewing that RGB triple. The first part of the article discusses retina cones and computer monitors. The following sections discuss the conversion process.

2 Photoreceptors in the Retina

Light enters the cornea and is focused upon the retina where the photoreceptors activate based upon the number of photons absorbed. There are two classes of photoreceptors called rods and cones: The cones are used for color and daytime sight and rods are used for low-light vision. There are three types of cones separated into short, medium, and long based upon their relative light sensitivity functions (Stockman and Sharpe, 2000; Stockman et al., 1999). Long (L) wavelength sensitive cones are often called red-selective cones, medium (M) cones are called green-selective, and the short (S) are called blue-selective. The peak selectivity of these cells do not map to what we would call pure red, green or blue but the names are useful in describing their general behavior. See the bottom plot in Figure 1 for the normalized sensitivity curves of L, M, and S cones as measured in humans.

Most computational vision simulations ignore color entirely and assume an achromatic architecture. Also, many lower animals do not have color vision, so doing a monochrome
simulation is the most biologically realistic representation. In trichromats such as macaque monkeys, achromatic neurons do exist, but there are also many color sensitive cells that are an integral part of their vision system.

Modeling a brain exposed to natural scenes requires that the simulation must have realistic retina activation patterns. A color in the natural world is not a single wavelength, rather it is a combination of multiple wavelengths each with a different intensity. To reproduce a color in the laboratory, experimenters usually must resort to using a computer monitor to simulate the retinal activation experienced in a natural scene.

A color image is displayed on a CRT monitor by the light emitted from phosphors that are excited by three electron guns. By varying the mix of the three color channels red, green, and blue (RGB) a monitor can reproduce the majority of colors observable by the human eye. The mixture of wavelengths from the CRT can activate the cones in the eye in almost the same ratios as the natural light being duplicated. For most experiments, the shortcomings of CRT monitors are negligible. The displays can be calibrated to show near lifelike images within the limits found in the camera optics, CRT phosphors, and ambient lighting of the testing environment.

The goal of the engineer is to make a computer monitor that can match as many colors as possible, as cheaply as possible, which might not necessarily duplicate the wavelength spectrum of what the eye sees in a natural environment. For example, a perceived single wavelength of orange could be duplicated with a mixture of red and yellow wavelengths. Computer displays rely upon this phenomenon to give the computer user an impression of natural colors when looking at a picture.

Figure 1 shows the difference between the light that computer monitor phosphors emit, and the cone sensitivity of the human eye. Notice how the M and L cones overlap a great deal, yet there is little overlap in the G and R guns.

The problem with using RGB channels when training or testing a neural network is that a pure wavelength stimulus, for example pure red \((R[i] = 255, G[i] = 0, B[i] = 0,)\) does not stimulate only one cone type. An RGB picture of a natural scene still has correlational information of the natural environment but it is in a different form than what the biological eye would see in nature. It is not clear what effect the RGB color transformation has on self-organizing neural networks.

Ideally, the raw light from nature should be used as input to a neural network, but it may not be available. It is not possible to use a traditional three-dimensional (RGB) digital camera to capture the naturally occurring high-dimensional wavelength mixtures. However, a special transformation can be used to simulate photoreceptors viewing photographs of natural scenes. The RGB to LMS transformation simulates a retina that is looking at a CRT monitor. Since neuroscientists also use CRT monitors to measure brain activity, it is reasonable to expect that the simulated and biological maps will have similar color selectivity properties. The next section describes the transformation used to convert RGB images to LMS cone activations.
The top plot shows RGB electron gun emission curves for a specific CRT monitor ($E_{R}^{\text{max}}$, $E_{G}^{\text{max}}$, $E_{B}^{\text{max}}$). The color of the curve specifies which gun was measured. The bottom plot shows normalized long (red), medium (green), and short (blue) human cone sensitivity functions (Stockman and Sharpe 2000; Stockman et al., 1999). An image designed to be displayed on an RGB monitor cannot be used as a photoreceptor activation function because it is designed for a different spectral curve in each channel.

3 The Transformation

Converting from an RGB image where each triple represents phosphor luminance, to an LMS image where each triple represents cone stimulation, requires three data sets:

- Wavelength sensitivity functions of retina cone photoreceptors.
- Phosphor photon emission functions for a specific computer monitor.
- RGB images of natural scenes that will be converted.

The emission spectra of a specific CRT monitor must be measured because each monitor has different energy emission functions. Variations between monitors are small enough that any display with realistic looking colors is sufficient to use in the conversion process.

An LMS cone activation triple must be calculated for each RGB color pixel. This is done by summing together the emission function for each monitor gun at the specified RGB pixel intensity. The dot-product of the cone sensitivity function and the summed emission intensity function, gives the final cone activity.
Notation

- All uppercase letters ($L, M, S, \ldots$) represent single dimensional numerical row-major column vectors. Subscripted vectors are distinct: $P_1$ is different from $P_2$.
- $(A)^T$ means the transpose of $A$.
- Lowercase letters ($\alpha, i, j, \ldots$) are scalars or variables.
- $\text{mod}$ is the remainder of integer division ($\text{div}$).
- All normalizations are $\infty$-norms. The $\infty$-norm of vector $A$ is the largest element of $\text{abs}(A)$.

3.1 Step 1: Extract the R,G,B components of each image pixel

A square RGB image named $Image$ contains normalized ($n \times n$) triples where,

$$\forall i, i \in [0, n * n), (R[i], G[i], B[i]) = Image[i \text{ div } n][i \text{ mod } n]$$

and

$$\forall i, i \in [0, n * n), 0 \leq R[i], G[i], B[i] \leq 1$$

$R[i]$, $G[i]$, and $B[i]$ are the Red, Green, and Blue components of an individual pixel within the image.

3.2 Step 2: Calculate the monitor spectrum emission functions

CRT monitors emit photons in a linear function relative to the input current, however, cone luminosity sensitivity is non-linear. Monitors include an exponential function, usually called a gamma function, to adjust the display. The exponentiation of the pixel values create a linear increase in perceptual luminosity across the bytes representing color.

The exact gamma function may vary depending upon the monitor, but a standard equation is: $\gamma(x) = x^{2.2}$. Ideally, the exact emission spectrum should be measured for each pixel value possible, but spectral emission measurements usually simply record when the phosphor is set at its brightest value. It is still possible to estimate the complete spectrum emission function for a monitor using a single set of luminance measurements. The top plot of Figure 1 shows the emission vectors for when $R[i]$, $G[i]$, and $B[i]$ are maximized.

By using the gamma function $\gamma(x) = x^{2.2}$; $x \in [0, 1]$ and the maximum spectral power distribution vectors $E_{R}^{\text{max}}$, $E_{G}^{\text{max}}$, and $E_{B}^{\text{max}}$ from Figure 1, all possible emission values can be calculated:
∀j, j ∈ \{R, G, B\}, E_j(x) = \gamma(x)E_j^{max} \quad (3)

3.3 Step 3: Calculate total spectral power emission for each image pixel

To calculate the luminance power for all wavelengths coming from a CRT pixel, the following equation is used:

∀i, i ∈ [0, n * n), P_i = E_R(R[i]) + E_G(G[i]) + E_B(B[i]) \quad (4)

E_R(\cdot), E_G(\cdot), and E_B(\cdot) each return vectors containing the strength of the phosphor emissions for discrete wavelengths at the specified intensity.

3.4 Step 4: Cone activity in the retina

The final step is to measure the activity of a cone observing a single CRT pixel. This is done by taking the dot product of the spectral sensitivity vector of the cone (L, M, or S, as shown in Figure 1) by the emission power spectrum \(P_i\) of pixel \(i\) in the RGB image.

\[\alpha_L[i] = (P_i)^T \cdot L, \quad \alpha_M[i] = (P_i)^T \cdot M, \quad \alpha_S[i] = (P_i)^T \cdot S\] \quad (5)

\(\alpha_L[i], \alpha_M[i], \text{ and } \alpha_S[i]\) are the scalar L, M, and S cone activation values for image pixel \(i\).

4 Normalization

The many levels of processing requires careful normalization of the data. Each subsection will discuss a different input and what type of normalization will be required.

4.1 CRT Spectral Power Distributions

When the CRT is showing pure white \((R[i] = 1, G[i] = 1, B[i] = 1,\) the peak energy between each gun will not be the same. Figure 1 shows that the red gun has a higher peak than the others. The relationship between emission functions must be preserved, but it is reasonable to normalize all three functions simultaneously to reduce rounding errors or make the data easier to plot. Dividing all three vectors by the wavelength with largest energy will create a dataset that has a maximum wavelength energy of 1.

\[\text{peak} = \| \|E_R^{max}\|,\|E_G^{max}\|,\|E_B^{max}\|\| \] \quad (6)
\[ E_{R}^{\text{norm}} = \frac{E_{R}^{\text{max}}}{\text{peak}}, \quad E_{G}^{\text{norm}} = \frac{E_{G}^{\text{max}}}{\text{peak}}, \quad E_{B}^{\text{norm}} = \frac{E_{B}^{\text{max}}}{\text{peak}} \] (7)

The area under each curve will be different, and the peak energy of each function will be different, as they should be.

4.2 RGB Images

Most RGB images are already adjusted to take advantage of computer monitor displays. Some regions of the image will be dark, and some images will contain values at or near maximum brightness. It should not be necessary to adjust the RGB images, but if it is required, the three channels can be scaled using equations similar to Equation 6 and Equation 7.

Note: A corpus of training images may have biases within the separate color channels. For example, a picture of grass and trees may have a mean Green pixel average of 0.80 out of 1.0, while the Blue channel mean may be a mere 0.2 out of 1.0.

4.3 Photoreceptor sensitivity functions

Photoreceptor sensitivity functions are very difficult to measure. It is meaningless to have the peak sensitivity wavelength of one cone be higher than another, unless the relationship between the long, medium, and short cone sensitivity functions are known.

There are three solutions to normalizing the photoreceptor sensitivity functions.

4.3.1 Method 1: Find the relationship

Find a dataset that has an exact peak-sensitivity correlation between the three cone types. This also requires an understanding of the behavior of the photoreceptors under different lighting conditions not included in this model.

4.3.2 Method 2: Preserve sensitivity correlations

This method preserves whatever relationship the original data contains. Here, the cone sensitivity functions are normalized similar to Equation 7, but the normalization term \( \text{peak} \) is calculated using Equation 9. Looking at a pure white display (Equation 8,) one of the three cones will have the highest activity, possibly beyond the required limit of 1. The highest cone activity is used as the normalization term \( \text{peak} \).

\[ P_{\text{max}} = E_{R}^{\text{max}} + E_{G}^{\text{max}} + E_{B}^{\text{max}} \] (8)

\[ \text{peak} = \| \left[ (L)^T \cdot P_{\text{max}}, (M)^T \cdot P_{\text{max}}, (S)^T \cdot P_{\text{max}} \right] \| \] (9)
After normalization, a pure white display will cause the most sensitive cone to fully activate to a value of 1, while the other two channels will peak out at some value less than 1. This assumes that there is no channel independent gain control, unlike the final method.

4.3.3 Method 3: Independent channel normalization

The third method assumes that each cone type has independent gain control so that when looking at a pure white light, each cone becomes fully driven. Each cone sensitivity function is independently normalized so that pure white light causes an activation value of 1. Equation [10] shows the formula for this form of normalization.

\[
L_{\text{norm}} = \frac{L}{(L)^T \cdot P_{\text{max}}}, \quad M_{\text{norm}} = \frac{M}{(M)^T \cdot P_{\text{max}}}, \quad S_{\text{norm}} = \frac{S}{(S)^T \cdot P_{\text{max}}}
\]

This form of normalization provides the best features for training a neural network. Each cone type can take all possible values, and a pure white light \((R[i] = 1, G[i] = 1, B[i] = 1)\) generates maximum cone activation \((L[i] = 1, M[i] = 1, S[i] = 1)\).

5 Conclusion

Upon completion of the RGB to LMS conversion, each triple of information contains simulated retinal cone activations as if the retina contained three overlapping regions of L, M, and S cones.

6 Appendix: Python Conversion Code

Python code is available from the author that performs the above transformation.

References
