

# COLOR IS NOT A METRIC SPACE

Implications for Pattern Recognition, Machine Learning, and Computer Vision

Thomas Kinsman<sup>1</sup>, Mark Fairchild<sup>2</sup>, Jeff Pelz<sup>3</sup>

Multidisciplinary Vision Research Labs, Rochester Institute of Technology, Rochester, NY 14623

{<sup>1</sup>Thomas.Kinsman,<sup>3</sup>Jeff.Pelz}@gmail.com

Associate Dean, College of Science, Rochester Institute of Technology, Rochester, NY 14623, <sup>2</sup>mdf@cis.rit.edu

**Abstract**—Using a metric feature space for pattern recognition, data mining, and machine learning greatly simplifies the mathematics because distances are preserved under rotation and translation in feature space. A metric space also provides a “ruler”, or absolute measure of how different two feature vectors are. In the computer vision community color can easily be mis-treated as a metric distance. This paper serves as an introduction to why using a non-metric space is a challenge, and provides details of why color is not a valid Euclidean distance metric.

**Keywords**—metric dimension; pattern recognition; feature space; distance learning; color space; CIELAB

## I. EXAMPLE NON-METRIC MEASUREMENT SPACE

Color is often used as an appearance cue [19] for image or object clustering.[20] As a result it is easy to think of color as a metric space, which is not the case. Here we demonstrate the problem of using a non-metric space, and the extension to color measures.

### A. Non-commensurate axes

Suppose that there are two surveillance suspects: Smith and Thomas. To tell them apart, we measure their features. Smith is 6 feet tall, with 1 FPSI. Thomas is 5.5 feet tall, with 3 FPSI. When plotted in feature space, (Figure 1.) we can ask, “What is the difference between Smith and Thomas?” It would be very easy to compute a “Euclidean” distance:

$$D = \sqrt{(6 - 5.5)^2 + (3 - 1)^2} = 2.06$$

But, what are the *units* of this distance? The units here are intentionally vague. Whatever units FPSI are, they are *not* feet. Consequently, D cannot be in feet. D is not a metric distance.

Furthermore, there is no understanding of the relative importance between the height measurement in feet, versus the skin complexion, in freckles per square inch [FPSI]. The only way for D to be a valid metric distance would be if both axes had the same units.

Consider the consequence of changing units. If the units of height were changed to centimeters, the value of D would change to 15.4. Height would be more important than FPSI. If the height was in kilometers, the difference in heights becomes nearly zero, and the D would only reflect the difference in FPSI, or 2.0. The bias between any two dissimilar axes is inherent in the measurement units.

### B. Application to color

Consider when one the units of the axes are changed to the amounts of green and red. The distance, D, still cannot have units associated with it. We can call it a color difference, but

the distance is *neither* only a difference in red, *nor* only a difference in green.

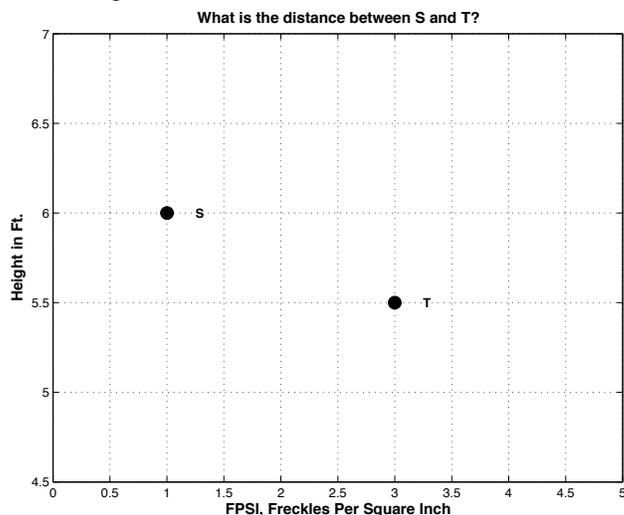


Figure 1 – Suspects Smith (S) and Thomas (T) plotted on a 2D graph.

The distance between two colors is a visual, perceptual, difference. Consider the red = green diagonal line on a pair of axes that are red versus green. One might say that the diagonal direction was “reddish-green-ness”, however there is no human perception of “reddish green”. [1]

### C. Finding a Metric Distance

In terms of feature extraction and feature measurement, determining the relative importance between features is termed “finding a metric distance”. [2] This relative scaling between the features is used to predict how important a difference between data points is in that space. In terms of color, the importance is application dependent. In dermatology, a slight change in skin redness could indicate infection. When trying to tell if a strawberry is ripe, a much larger change in redness is required. The metric for the space, the relative importance of features, depends on the application.

## II. NON-RECTILINEAR SPACES

### A. Skewed Axes in Analytic Geometry

In the 1980’s, when people plotted all data by hand, it was sometimes convenient to use skewed axes, axes that were neither rectilinear nor orthogonal. Graphical techniques were figured out using graph paper that encompassed relative relationships. Triangular graph paper was used in chemistry to

plot relative concentrations. Skewed axes helped automatically visualize relationships.

At present, the idea that axes are orthogonal is our default mindset. We assume Cartesian Coordinates.[3]

Figure 2 shows data plotted on *skewed* axes. It shows three points. The point S is located at (0.0, 0.8), T is at (0.2, 0.6), and R is located at (0.0, 0.4). In skewed space, on the skewed axes, the distance from S to T is the same as the distance from R to T. However, in reality (on paper), the distance  $\overline{RT}$  is more than twice the distance  $\overline{ST}$ . ( $\overline{ST} \approx 0.20$ ,  $\overline{RT} \approx 0.44$ ).

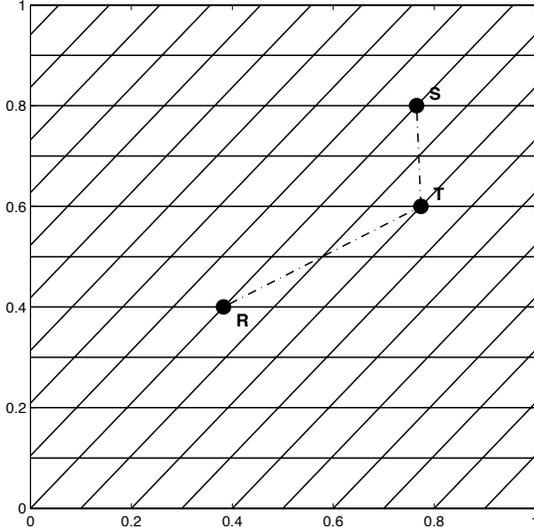


Figure 2 - An example of non-Euclidean, or skewed, axes.

### B. Skewed axes in color

In 1931, J. Guild and W. D. Wright created standardized sensitivity functions that modeled how the human visual system [HVS] responded to different frequencies of light. [Berns et al.] Their functions modeled how the HVS matches luminance levels, and chrominance levels for wavelengths of light between 380 nm to 780 nm. These color *matching* functions,  $(\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda))$ , form the basis vectors for the way color is perceived by the HVS.

The tabulated values for these basis vectors, the color matching functions, are publically available over the internet or in publications.[4,5] The integral of these basis functions over all wavelengths, are called the (X,Y,Z) tristimulus values.

Equation 1 is given for computing tristimulus values, XYZ. The amount of light going into the eye (after reflecting off an object) at each wavelength is  $s(\lambda)$ , and  $k$  is a normalizing constant. (Simplified from Berns.[6]) If the illumination changes,  $s(\lambda)$ , will change, and X, Y and Z will all change.

$$\begin{aligned} X &= k \int_{\lambda} s(\lambda) \cdot \bar{x}(\lambda) d\lambda \\ Y &= k \int_{\lambda} s(\lambda) \cdot \bar{y}(\lambda) d\lambda \\ Z &= k \int_{\lambda} s(\lambda) \cdot \bar{z}(\lambda) d\lambda \end{aligned} \quad (1)$$

A practical rule of Analytical Geometry is to work in the positive quadrant whenever possible. The  $(\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda))$

basis vectors, the color matching functions, do just that. However, they are also skewed.

### C. Computing the skew of the XYZ Tristimulus Axes

As per (1), the (X,Y,Z) tristimulus values are the weighted integral along the basis directions  $(\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda))$ .

The steps involved in computing the angles between two basis vectors are: take their dot product, divide by the norms of the component vectors, and take the inverse-cosine. Using [4] or [5], we can solve for the angles between axes. (Table 1).

$$\angle(\bar{x}(\lambda)\bar{y}(\lambda)) = \cos^{-1} \left( \frac{\bar{x}(\lambda) \cdot \bar{y}(\lambda)}{|\bar{x}(\lambda)| |\bar{y}(\lambda)|} \right) \quad (2)$$

Table 1 - Angles between the color matching functions used for the 1931 Standard Color Observer

$$\begin{aligned} \angle(\bar{x}(\lambda)\bar{y}(\lambda)) &= 76.95^\circ, \\ \angle(\bar{x}(\lambda)\bar{z}(\lambda)) &= 83.53^\circ \\ \angle(\bar{y}(\lambda)\bar{z}(\lambda)) &= 86.41^\circ \end{aligned}$$

None of these angles are  $90^\circ$ . The axes that Guild and Wright *designed* were never intended to compute absolute differences.[7] Small distances will be approximately linear, but the error increases as the distance grows. Any formula based on XYZ tristimulus values incorporates the skew built into the  $(\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda))$  axes.

### D. Numerical Consequence of using Skewed Axes

Skewed axes are well suited for visualizing some types of data. Guild and Wright used them to force the resulting tristimulus points to have the desired properties: always positive, and to bend the spectrum locus into a convex shape.

While skewed axes are beneficial for visualization, they are generally detrimental to analysis. When two lines intersect at an angle, the numerical precision is inversely proportional to the sine of the angle between the two lines. In an extreme case, when two lines are nearly parallel, (intersect at a very small angle), the intersection point is very sensitive to noise.

The outcome of the fact that the color matching functions are skewed is that numerical accuracy of intersections between  $\bar{x}(\lambda)$  &  $\bar{y}(\lambda)$ , will be worse than the intersections of  $\bar{x}(\lambda)$  &  $\bar{z}(\lambda)$ , which are worse than for intersection of  $\bar{y}(\lambda)$  &  $\bar{z}(\lambda)$ . Had the angle  $76.95^\circ$  been much farther than  $90$  degrees, the errors might have been bad enough to cause better color *difference* functions to be created from the ground up. As it is, the cumulative numeric errors caused by the non-orthogonality of these basis functions is  $\approx 5$  percent.

### III. COMPUTING 1/U – THE ERROR OF THE INVERSE

In Color Science, the chromaticities of points are sometimes used. Chromaticity values are computed:

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z} \quad (3)$$

In science, the function  $1/u$  is important. It is used to convert electrical resistance to conductivity, in the thin lens equation for optics, and here in color science to compute

chromaticity values. As the  $u$  goes down towards zero, the inverse goes towards infinity, *and* it goes towards infinity at a faster and faster rate.

Numerically, division by zero is difficult to ignore because computers either complain or give dramatically inaccurate results. Yet, it is easy to ignore that when the  $u$  gets small, the slope gets large. The slope of  $f(u) = 1/u$ , is  $f'(u) = -1/(u^2)$ . If  $u$  is decreased to a tenth of its value, the function increases to 10 times its value, and the slope increases to 100 times its value.

The error in a linear approximation (a Taylor's series) is proportional to the slope [8]. This means that when the value of  $u$  in  $f(u) = 1/u$  drops one order of magnitude, the error increases by two orders of magnitude.

Linear models are commonly used in colorimetry. When using the von Kries chromatic adaptation model [9], there is a division by the amount of ambient illumination. When transforming between color primaries there is another division [10]. If the result of the division is close to 1, the magnitude of the slope of  $f'(u)$  is close to 1, and errors are not amplified.

In darker regions, noise and measurement errors are amplified. Any machine learning algorithm must either weight these errors appropriately, or learn to ignore them. Otherwise, a machine learning algorithm would tend to over-fit this amplified noise.

This inverse relationship is used to model white-point accommodation. The accommodation is performed by putting the inverse relationship into a matrix and using the matrix to perform a linear transform. However, what goes into the creation of the matrix is not linear with the white point.

Many linear approximations we use are true only in the neighborhood of an optimization point. In a feature space that is not linear, the approximations we use can change as a function of the feature space. This is why data must be collected carefully supervised methodology and measurements made in the vicinity of the data, instead of having functions that are purported to hold for the entire data set.

#### IV. USING AN $L^1$ NORM INSTEAD OF AN $L^2$ NORM

In this section we describe why chromaticity points are computed by dividing each tristimulus value by the sum of the squares of the three tristimulus values.

Figure 4 shows two circles centered on the origin. The first one, in the dotted line, shows the typical circle, all the points are equidistant from the origin. This distance *from the origin* is defined using the common  $L^2$  definition of the distance, as per the familiar formula (4).

$$D_{L^2} \equiv L^2 = \sqrt{X^2 + Y^2 + Z^2} \quad (4)$$

All of the points on the figure that looks like a diamond are also equally distant from the origin, but using the  $L^1$  norm. The distant *from the origin* using the  $L^1$  norm is given by (5).

$$D_{L^1} \equiv L^1 = |X| + |Y| + |Z| \quad (5)$$

In a Euclidean space, with a vector  $(X,Y,Z)$ , the  $X$  component would be normalized to a unit vector using (6).

$$x = \frac{X}{\sqrt{X^2 + Y^2 + Z^2}} \quad (6)$$

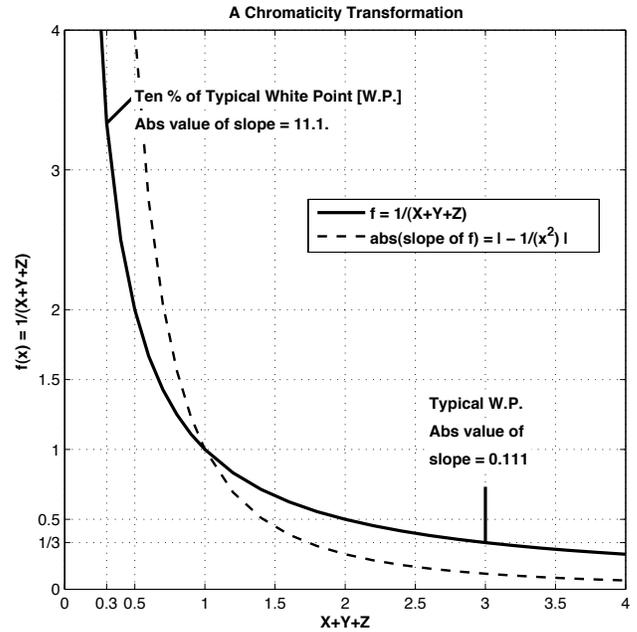


Figure 3 - Transformation from Tristimulus Values to Chromaticities

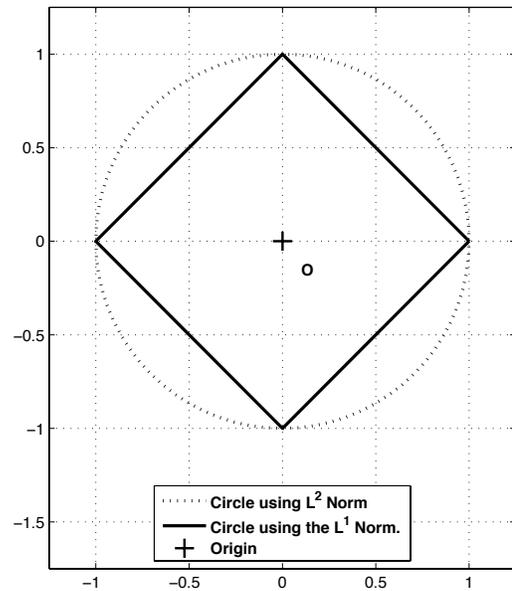


Figure 4 - Two circles of radius 1, centered on the origin, using the  $L^2$  norm and the  $L^1$  norm.

This formula for vector normalization only works for metric spaces. In non-metric spaces, when normalizing a vector, it is common practice to define by the length of the vector using the  $L^1$  norm, and normalize by that:

$$x = \frac{X}{|X| + |Y| + |Z|} \quad (7)$$

Since Guild and Wright specified that their color matching function,  $(x(\lambda), y(\lambda), z(\lambda))$  should always be positive, any integrals of them will also always be positive. Thus, all values of  $X$ ,  $Y$ , and  $Z$  will always be positive using real light sources, and we can remove the absolute value signs:

$$x = \frac{X}{X+Y+Z} \quad (8)$$

This is the transformation that maps (X,Y,Z) tristimulus values to (x,y,z) chromaticity coordinates. It makes no presumption about the relative importance of X, Y, or Z.

In the pattern recognition community when the space is not a metric feature space and the relative importance of the distances along different features is unknown, but when a similarity measure must still be computed, it is common to use the  $L^1$  norm instead of the usual  $L^2$  norm as is done here.

## V. WHY DOES COLOR WORK REGARDLESS?

In Wright's 1981 paper, "50 years of the 1931 standard observer", [11] Wright stated that the reason all the math works is because light sources and reflecting surfaces tend to be unimodal – the distributions have one dominant cluster. Problems occur when light sources are bi-modal or tri-modal. (As the world moves to using more compact fluorescent and LED lights, this will no longer hold.) Wright emphasizes that when they developed the XYZ space, it was to be a color *matching* space, it was never intended to be a color *difference* space.

The CIELAB color space was developed in 1976 [12] in an attempt to be an opponent process color space and so to have a more perceptually uniform color space. While perceptual uniformity was the intent, it is not uniform globally. A line drawn along constant perceptual hue in the blue region of CIELAB does not extend from the center of CIELAB space outwards. Furthermore, CIELAB is computed from XYZ intermediate computations, incorporating all of the errors in computing XYZ values into CIELAB.

## VI. CONCLUSIONS

The Computer Vision practitioner should be acutely aware that color is not a valid metric distance. For global measurements, color is not a Euclidean distance, and not a valid measure for computing metric differences in.

For pattern recognition, two objects that differ by a certain amount of: red, green, and blue are best described as that. The leap from color measurements to a distance metric in color space requires an understanding of the relative relevance of the different color channels, in the region of color space being considered, and for the application being developed.

Over the years the CIE community has come up with several formulas to try to model uniform color differences. These models all try to associate a formula for human visual perceptual differences. Unfortunately they also tend to disagree with each other [13]. The CIEAE1994 and CIEDE2000 standards are available as difference tools, and are based on careful factor analysis.[14,15] The 1994 version is an older standard but has the benefit of computational simplicity.

The human visual system tries to amplify and notice differences in the environment, including differences in color. It adapts to the amount of light, the color of the light source and the color of the surrounding objects. [16]

While CIELAB may be a better difference space for object discrimination or clustering than RGB [19], it still cannot be

treated as a metric space because it is derived from the color matching functions, which are neither orthogonal nor rectilinear.

More recently, it has been suggested that perhaps color should not be treated as a multi-dimensional space at all, but instead be considered as separate "appearance scales". From this view, lightness, hue, and chroma are all separate one-dimensional scales, with no geometric relationship. [17,18]

Color in Computer Vision should be understood as a measurement in a non-metric space. The relative importance of the colors features is application dependent.

## REFERENCES

- [1] Hering E, 1964. *Outlines of a Theory of the Light Sense*. Harvard University Press, Cambridge, Mass.
- [2] Liu Yang; Rong Jin; Mummert, L.; Sukthankar, R.; Goode, A.; Bin Zheng; Hoi, S.C.H.; Satyanarayanan, M.; *A Boosting Framework for Visuality-Preserving Distance Metric Learning and Its Application to Medical Image Retrieval, Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.32, no.1, pp. 30-44, Jan. 2010.
- [3] E. Purcell, *Analytic Geometry*, Appleton-Century-Crofts, Inc., New York, 1958, pp. 1.
- [4] <http://www.cis.rit.edu/mcsl/research/1931.php> Downloaded Sept. 20, 2012.
- [5] R. Berns, *Principles of Color Technology, 3<sup>rd</sup> Ed*, 2000, John Wiley & Sons, New York, pp. 51.
- [6] *Ibid.*, pp. 56.
- [7] W. D. Wright, *50 years of the 1931 CIE standard observer for colorimetry*, AIC Color 81, Paper A3 (1981).
- [8] E. Swokowski, *Calculus with Analytic Geometry, 2<sup>nd</sup> ed.*, Prindle, Weber & Schmidt, Boston, Mass, 1979, pp. 533, (11.11)
- [9] M. Fairchild, *Color Appearance Models, 2<sup>nd</sup> ed.* John Wiley & Sons, Inc, New York, 2005, pp 168-177, eq (9.29)
- [10] R. Berns, *Principles of Color Technology, 3<sup>rd</sup> Ed*, 2000, John Wiley & Sons, New York, pp. 206.
- [11] W. D. Wright, *50 years of the 1931 CIE standard observer for colorimetry*, AIC Color 81, Paper A3 (1981).
- [12] R. Berns, *Principles of Color Technology, 3<sup>rd</sup> Ed*, 2000, John Wiley & Sons, New York, pp. 66-70.
- [13] *Ibid.*, pp. 108-112.
- [14] *Ibid.*, pp. 120-121.
- [15] M. R. Luo, G. Cui, and B. Rigg, *The development of the CIE 2000 colour difference formula*. Color Res. Appl. 26, 340-350 (2001).
- [16] M. Fairchild, *Color Appearance Models, 2<sup>nd</sup> Ed*. John Wiley & Sons, Inc, New York, 2005, pp. 52.
- [17] M.D. Fairchild and R.L. Heckaman, "Deriving appearance scales," IS&T/SID 20th Color and Imaging Conference, Los Angeles, in press (2012).
- [18] M.D. Fairchild, "Is there really such a thing as color space? Foundations of unidimensional appearance spaces," ISCC/IS&T/SID Special Topics Meeting: Revisiting Color Spaces, San Jose 2(2011).
- [19] X. Wang, et. al. "Shape and Appearance Context Modeling", ICCV, Proc. IEEE International Conference on Computer Vision (2007).
- [20] Y. Rubner, C. Tomasi, L. Guibas, "Empirical evaluation of dissimilarity measures for color and texture", Proc. IEEE Sixth International Conference on Computer vision, pp 59-66 (1998).