A Comparison of Color Difference Data and Formulas

Dr. Edward M. Granger*
Key Words: Color, Comparison, IQRGB, Tolerance, Vision

Abstract

A study has been made of color difference data from four laboratories. This was done to test a number of standard color difference formulas along with a new color channel model for performances against the known data. This study shows that the current CIEDE2000 has large systematic errors and is not an improvement on the CIEDE94 metric. It also illustrates that a simple color channel model performs as well as the more complex CIE color difference formulas.

A simple color channel model has been developed based on the ATD color opponent color space first proposed by Lee Guth (1973). The new model is non Euclidean and is based on the actions of the opponent channels of human vision. The ATD model has been simplified by Granger (2001) to yield a uniform chromaticity space that has been specifically tuned to Graphic Arts applications.

The ATD model is expanded in this report to include a color difference formula, Delta Perception (DP), for the perceived subjective color difference. The accuracy of the new color difference formula is compared with DE2000, DIN99, DE94, CMC (1, 1) and DE76. The result of the study shows that a simple three-term formula gives good performance compared to the extant color difference formulas. A secondary discovery was that new color difference formula produces nearly identical error distributions. The other color difference equations produce error distributions with differing means. This indicates that there is a bias built into the each model that does not offer equal treatment to all the test data.

Introduction

The ATD color space has lacked a metric that can accurately predict the perceived difference between color samples that lie close to one another in color space. This report presents a simple three-term model, DP, which predicts perceived color differences. The ability of the model to predict color differences was tested on visual data from four different studies. DP was compared to the performance of DE2000, DIN99, DE94, CMC (1, 1) and DE76. DP was found to be as accurate as the current color difference equations.

The DP metric is based on the principle that the final perception of color and color difference is a linear function of each of the vision's channels. They are luminosity, the Red-Green opponent and the Yellow-Blue opponent. These are hypothesized to act in a non-Euclidian manner to produce the color stimulus.

---

* Ontario Beach Systems, LLC

The most important factor in the model is the impact of the color opponents on the perception of brightness

The chromatic test patches viewed on a neutral background of the same luminosity are usually seen as having a different brightness from the neutral background. This effect is known as the Helmholtz-Kohlrausch Effect. This effect (HKE) is produced by the chromatic channels of the visual system modifying the luminance channel to produce a brightness that is greater or less than that predicted by the CIE Y tristimulus value. The brightness can be greater or less than that predicted depending on the hue of the stimulus. The ATD vision model developed by Guth (1991) has been modified by Granger (2001) to produce a uniform and a linear color space. The brightness, denoted Q is defined by correcting the luminance, A, of Guth's model for the intrusion of the chromatic channels using equation (1).

$$Q = A + T/2 - D \tag{1}$$

The vector, T, is the tristimulus response of the Red-Green opponent channel. The vector, D, is the response of the Yellow-Blue opponent channel. The introduction of Q yields the new QTD color space that has a chromaticity space with uniform color scaling. The new space has been designed to work over the range of lighting conditions common to Graphic Arts.

DP Model Development

A color naming experiment was conducted to determine why, for some hues, the color name changes as a function of darkness relative to the white point while others keep their name. A good example of a color that changes its name is orange. The name for a dark orange is chocolate. The perception of the color chocolate does not relate to orange even though they have exactly the same chromaticity coordinates. In comparison, Blue maintains its hue name until it appears black. In English, the name for Blue with low luminosity is Dark Blue but in Russian, it is still called Blue.

Further study revealed that when the Yellow-Blue opponent is dominated by the Yellow opponent the colors change their name at low luminosity. Conversely, when Blue dominates the Yellow-Blue channel, colors with low luminosity keep their color identity until the color is perceived as Black.

Kuehni (2000) showed that the ratio of lightness to brightness does change as a function of whether Yellow or Blue is dominant in the activity of the Yellow-Blue opponent channel. The data of Wyszecki and Stiles (1982) and Sanchez and Fairchild (2001) was used to create a new Brightness Model for the QTD color space. Wyszecki and Stiles studied the brightness to luminance ratio for reflective tiles. Sanchez and Fairchild studied the brightness to luminance ratio for colors produced on a monitor.

A new model for Q was hypothesized which would compensate for HKE. The form of the model is given by equation (2).

$$
\begin{aligned}
&\text{If } D > 0 \text{ then} \\
&\quad Q = A + C1 * T \\
&\text{Else} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2) \\
&\quad Q = A + C1 * T + C2 * D
\end{aligned}
$$

Where C1 and C2 are constants determined by regression against the two data sets. The constant C1 is common to both forms of Q to maintain continuity at the Yellow-Blue opponent boundary.
Figure (1) shows the data used in the regression. All the data shown has been corrected to a D65 illuminant using the Bradford Transformation. Only the most chromatic data from the Wyszecki and Stiles samples were used in the regression. The pastel colors were felt to have little information to aid in the determination of C1 and C2. All the points of the monitor investigation were included as they were all at the chromatic limits of the monitor.

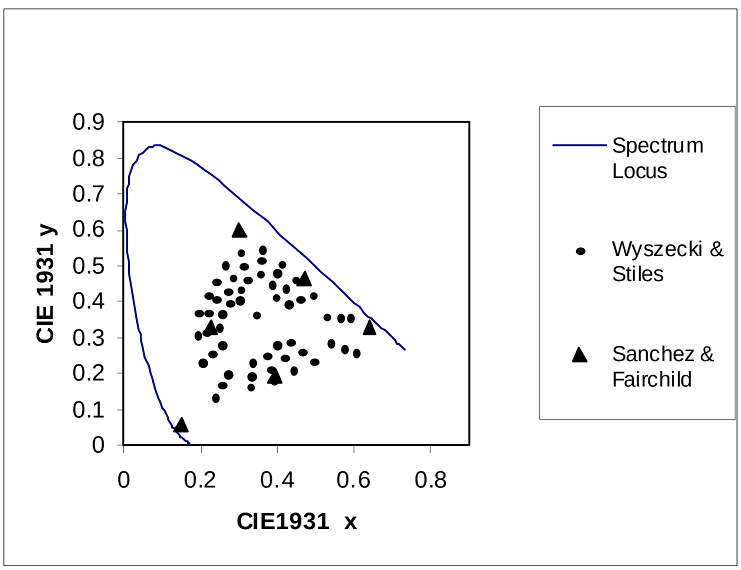

Figure (1) Chromaticity Location of B / L Experimental Samples

The observed and predicted B/L ratios for the Sanchez &Fairchild and the Wyszecki & Stiles data are shown on Figure (2). The regression fit to the Wyszecki & Stiles B/L data is shown on the figure. The Sanchez &Fairchild data was not included in the regression so that we could have an independent check of the model. Kuehni stated that his regression did not fit the observations in the highly chromatic blues. The very chromatic blue of the monitor produced a B/L ratio of nearly 4.0. Figure (2) illustrates that the regression line produced using the reflective tile data also fits all the monitor data.

The Brightness factor formula for the QTD color space is determined to be,

$$
\begin{aligned}
&\text{If } D > 0 \text{ then} \\
&\quad Q = A + T / 2 \\
&\text{Else} \\
&\quad Q = A + T / 2 - 3 * D / 4,
\end{aligned}
\qquad (3)
$$

where C1 = 1/2 and C2 = - 3/4.

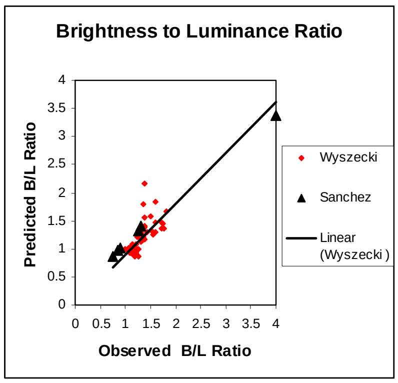

Figure (2) Regression fit to B / L Data

The resulting definition of Q and the close fit to both sets of data supports the hypothesis that the perception of brightness is different for colors when the Yellow-Blue opponent changes from being mainly yellow stimulated to being mainly blue stimulated. This also supports the observations of the color naming study in which color names change for low luminosity when the yellow predominates in the Yellow-Blue opponent.

The ATD color space proposed by Granger (2001) is scaled to produce 10-bit color data that corresponds to the range of luminances found in high range photographs. This definition allows a rendering engine to achieve printed output ranges that are close to the original. The conversion of CIE XYZ tristimulus values to ATD tristimulus values is given by Equation (3). The matrix given by Equation (3) has been scaled to give a one to one transformation of CIE Y to the luminance term A.

$$[ATD] = [XYZ] * \begin{bmatrix} 0.0 & .6265 & 0.1107 \\ 1.0 & -.5765 & .1497 \\ 0.0 & -0.0172 & -.2342 \end{bmatrix} \qquad (3)$$

The combination of Equation (2) and Equation (3) produces the QTD color space.
The chromaticity coordinates for QTD, t and d, are defined in Equation (4). Figure (3) displays the color mixing functions of QTD. The chromaticity coordinates are computed by,

$$t = T / Q \quad \text{and } d = D / Q. \qquad (4)$$

The t and d chromaticity terms combined with Q are used to create a new color difference function for QTD.

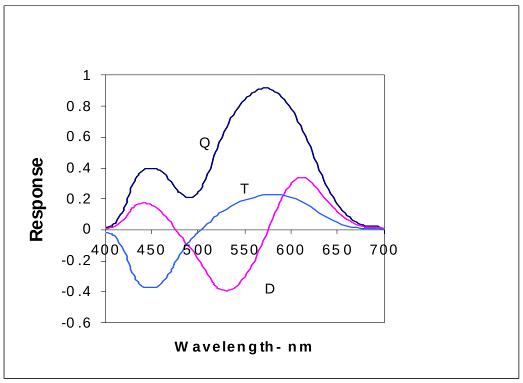

Figure (3) QTD Color Mixing Functions

The QTD color space that will be used in the DP color difference equation is described next. The space is non-Euclidian and is based on the individual contributions of the QTD vectors. The Euclidian approach to color differences is to employ a square root of the sums of squared to determine the difference. A root mean square matrix optimization is used for a target value that is reduced to a minimum by solving a quadratic matrix equation.
The approach taken in this paper is to look at how each of the ATD channels contributes to the perceived color. The result is to a new way of defining what we mean by hue, saturation, and hue angle. Although hue angles will not be used to determine differences of hue in the color difference metric, the hue angle will be defined in terms of hue and saturation as described below. It will be shown that hue angle does correlate well with hue angles of the Munsell color space. This illustrates that the linear channel model is consistent with known color data.

The Qtd perception space uses Q to describe brightness. The chromaticity coordinates are used to define hue, saturation and hue angle in Equation (5) as,

$$
\begin{aligned}
&S = \text{greater of } t \text{ or } d \text{ based on absolute distances}\\
&R = d / S \text{ or } t / S\\
&h = \text{lesser of } |t| \text{ or } |d| \qquad (5)\\
&H = R*h
\end{aligned}
$$

where S is saturation, H is hue and R is the hue angle. The value of R depends on which quadrant that t and d fall.

The Qtd hue angle, R, is compared to the 100 hue data of the Munsell system for all Munsell renotation data, approximately 2700 total points. The Qtd system has 256 hue angles and the Munsell system has 100. This study did not eliminate any of the points that are far removed from colors useful in the graphic arts. The Bradford Transform was used to move the illuminate C white point of the Munsell color space to the D65 white point used in the ATD color space. Since illuminant C and D65 are close, this transformation would not be a source of error in this exercise. The Munsell hue notation is converted to correspond to the ATD hue angle as given in Table 1.

Table (1) Munsell Hue Notation Conversion

Munsell notation	New scale value
RP7.5	0.0
YR7.5	20.0
Y7.5	30.0
GY7.5	40.0
G7.5	50.0
BG7.5	60.0
B7.5	70.0
PB7.5	80.0
P7.5	90.0

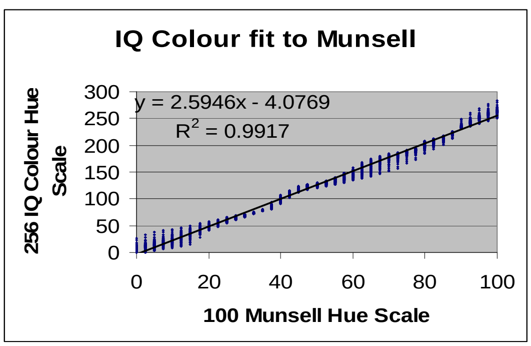

Figure (4) Hue Angle, R vs. Munsell Hue Angle

Figure (4) shows that the definition of hue in the Qtd perception metric is well correlated to the Munsell renotation data over the entire color volume sampled by Munsell. This result demonstrates that definitions of saturation, hue and hue angle given in Equation (5) have an excellent probability of forming a usable color difference metric.

The color difference data sets from the BFD_D65, Leeds, Witt and RIT-DuPont were supplied to me in a comprehensive Microsoft Excel sheet by M. Melgosa. These are the latest color difference data that have been corrected for previous errors. Therefore, these sets of color differences, which were used to develop DE2000, represent the best test of a candidate color difference metric. The new DP color difference equation is compared to the DE2000, DIN99, DE94, CMC (1, 1) and DE76 color difference equations.

The assumption made was that the color channels of A, T and D are independent and that the resulting perception vectors Q, t and d are also independent. The assumption is that a subject in a color difference experiment has these three independent percepts on which to judge the difference between two patches. A Taxi or City Block metric was chosen as a candidate for the DP color differences. The form of the metric is given in Equation (6),

$$DP = a * |\Delta Q|^{\gamma} + b * |\Delta S|^{\gamma} + c * |\Delta H|^{\gamma} \qquad (6)$$

where S and H are defined in Equation (5) and the values for a, b, c and γ were determined using a nonlinear regression.

The regression uses the average of absolute differences as the goal in the minimization. This method was adopted based on the assumption that the errors made by the judges were linearly distributed. An additional hypothesis is that a squared error metric places an uneven weight on judgment errors. Therefore, the least squares analysis does not fit the assumption of linear independent channels adding their activities to produce the sensation of color and color differences.

Torgerson (1967) states that the best measure of goodness of fit using paired comparison or category scaling is the mean absolute difference between the data and the model. The nonlinear regression, on the minimum absolute error, results in the color difference formula given by Equation (7).

$$DP = 80 * |\Delta Q| + 40 * |\Delta S| + 100 * |\Delta H| \qquad (7)$$

The regression is performed on all the data sets. The only exception was that the color difference range was limited to 3.0 just noticeable differences. A review of the data sets indicated that data outside this range are not reliable enough to be included. The results are listed in Table (2).

The error matrix of Table (2) shows that the variability across all metrics for the RIT-DuPont data set was approximately one third less than the other studies.

Table (2) Comparison of Mean Absolute Error

	Delta P	CIEDE2000	DIN99	CIE94	CMC(1:1)	CIEDE76
RIT	0.32	0.14	0.35	0.15	0.35	0.42
Witt	0.51	0.28	0.24	0.38	0.58	0.96
Leeds	0.47	0.17	0.51	0.27	0.50	0.65
BFD_65	0.51	0.24	0.20	0.38	0.44	0.94

The RIT-DuPont data was used to check the performance of DP as a function of luminosity. The data was sorted to run from low to high luminance to find if there was a bias in the error. A bias was anticipated because of the use of a simple linear perception model. The results of the study are shown in Figure (5).

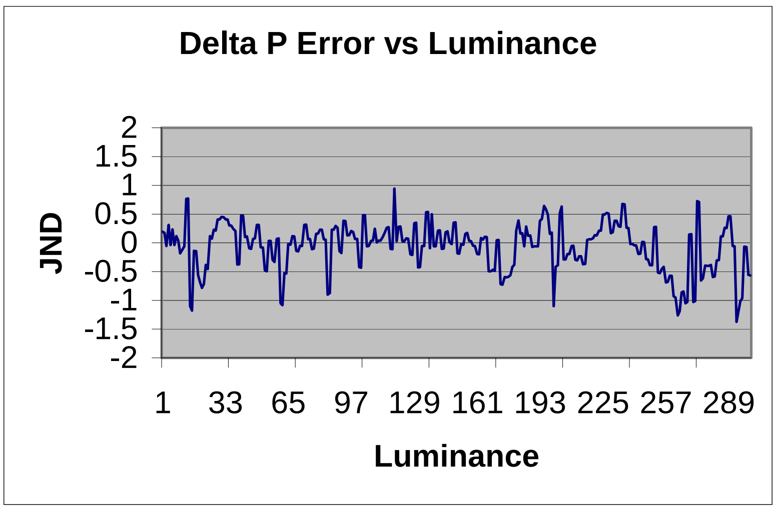

Figure (5)

Therefore, the nonlinear regression was applied to only the RIT-DuPont data set. The other color difference data sets were used to test the efficacy of the model.

The assumption in subjective scaling is that the judges are making random Gaussian distributed errors in their estimation of the number of just noticeable differences between samples. Normally distributed errors should have a symmetric ogive shape and be centered on zero. If the hypothesis is true, then the normal error distributions for all data sets will have a zero mean and the same variance. A cumulative plot of the errors produced by each of the color difference equations will indicate if the errors appear to be normally distributed or if there is a systematic bias introduced in the errors. Figures (6 – 11) are plots of the cumulative errors.

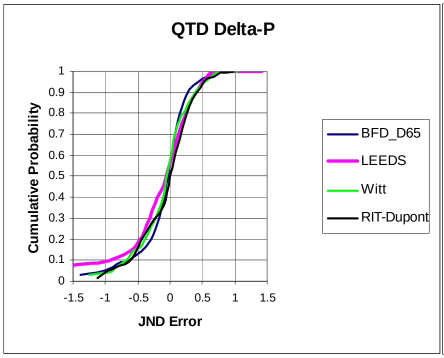

Figure (6)

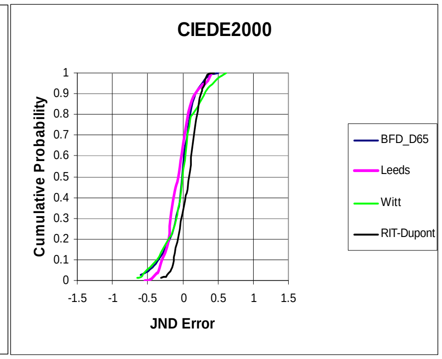

Figure (7)

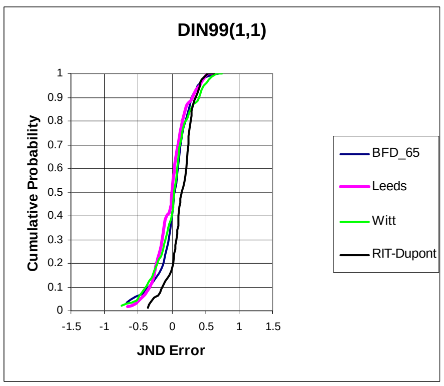

Figure (8)

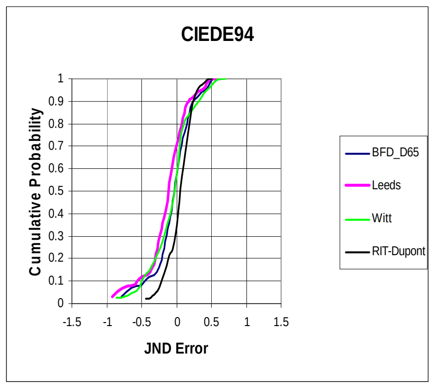

Figure (9)

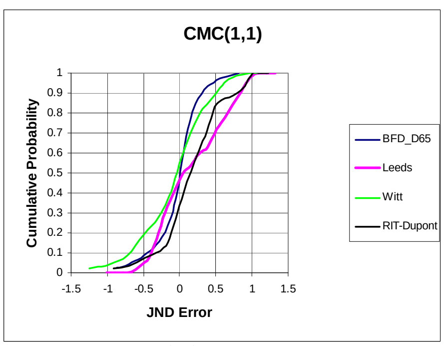

Figure (10)

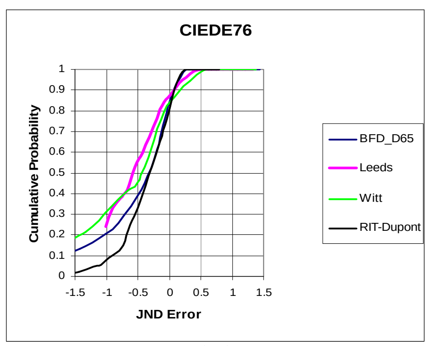

Figure (11)

A comparison of the cumulative plots shows that with the exception of the DP all the others have a mean bias error. The assumption is made that all the judges have the same error dstribution. The DP color difference model ,while having more variance, is treating the data from all studies in a uniform manner. It appears that the simple linear model has the least mean error between treatments and produces nearly identical cumulative error distributions.

Conclusions

The results support the hypothesis that a simple linear vision model can produce a viable color difference metric. The hypothesis used to develop the DP color difference model can not be rejected. DP best fits the underlying premise that all the data used would produce a gaussian error distribution with zero mean. The errors of the DP formula are well within the variblity that exist between the four sets of data used to create the DE2000 color difference formula.

The errors made by DP are independent of the luminance of the samples. The complexity added to the DE94 or DE2000 to fit the data did not yeild the statistical fit that the DP model provided.

The DP model is based on the assumed opponent physiology of the human vision system that is acting at an early stage of the viual process. The additivity of the luminance and the Red-Green and Yellow-Blue opponent channel differences is supported.

The HKE derived brightness is the only hue correction made in the color difference equations. The addition of the color opponents to the luminance channel produces the observed brightness. Our only additional assumption was that the Yellow-Blue opponent did not contribute to brightness when blue was not dominant. The opponent model fits both surface and monitor observations of HKE. This is further support for the three channel linear model.

A new concept for saturation and hue where introduced. The assumption is which ever color opponent is dominant sets the name of the primary color ie, red, green , yellow and blue. This color channel also sets the apparent saturation and the change of saturation of a color.

The secondary opponent sets the differential hues such as orange, yellow-green, cyan and magenta. Therefore, the change in the output of secondary opponent commands hue and the difference of hue.

Acknowledgments

This work would not have been possible without color difference Excel worksheet supplied by M. Melgosa. Franz Sigg and John Seymore have been extremly helpful in helping with the mathematics.The discussions and comments from R. Berns and M. Fairchild are greatly acknowledged.

References

Granger, E. M.
2001 “A Vision Based RGB Color Space”, TAGA, pp 318-
Guth, S.L. and Lodge, H. R.
1973 “Heterochromatic additivity, foveal spectral sensitivity and a new color model” JOSA, Vol. 63, Issue 4, pp 450 -
Guth, S. L.
1991 “Model for color vision and light adaptation” ,
JOSA, Vol 81, pp 976 -
Kuehni, R. G.
2000 “An Opponent-color Model for the Sanders-Wyszecki Helmholtz-Kohlraush Effect Dataset”, Color Research & Application, Vol. 25, No. 4, pp 292-
Sanchez, M. and Fairchild M.
2001 “Perceptual Amplification of Color: Observer Data and Models”, CIC9, Ninth Color Conference, Scottsdale, AZ
Torgerson,W.S.
1967 “Theory and Methods of Scaling,7th Edition”Wiley, New York, New York, pp 241
Wyszecki, G. and Stiles, W. S.
1982 “Color Science, 2nd Edition” Wiley, New York, New York,
pp 410