

Does *CIELUV* Measure Image Color Quality ?

Andrew N Chalmers
Department of Electrical and Electronic Engineering
Manukau Institute of Technology
Auckland, New Zealand

Abstract

A series of 30 split-screen video scenes has been investigated, with one reference image and one test image in each scene. An earlier study had obtained a set of subjective evaluations of the color quality of the test images. In the new work reported here, image digitization and computer-aided analysis of the color content of each image have facilitated the computation of image color statistics aimed at providing a numerical technique for the assessment of color quality. Correlations with the earlier subjective data are investigated, and there are good indications that some relatively simple colorimetric statistics in *CIELUV* space provide a meaningful measure of color quality.

Introduction

The purpose of this work is to assess whether there are specific objective criteria that may be applicable in the prediction of subjective color preferences in video scenes. It is hoped thereby to facilitate the formulation of an automated process for the enhancement of image color quality.

Work with paired images presented on self-luminous displays has led to the suggestion that the ***average*** color difference in *CIELUV* space, between a degraded test image and an undegraded reference image, serves as a useful measure of the color quality of the test image [1,2]. Color quality was based on subjective assessments of the images by a group of 25 observers, using two CCIR grading scales for the assessment of image quality [3]. Two objects were used in the creation of these images: a MacBeth ColorChecker® test chart [4], and a photographic portrait (printed copy).

In addition, investigations of the ***individual*** *CIELUV* color differences between the test and reference images, within selected regions of the portrait image, have shown significant correlation with the subjective ratings [2]. In the case of specific colors in the test-chart image, there was some less convincing evidence - which is thought to be explained by the ability of the observers more easily to relate to the colors in the natural portrait image.

Recent work by De Ridder *et al.* in the Netherlands [5,6] has concluded that the *CIELUV* chroma difference and the *CIELUV* chroma scatter (as measured by the standard deviation of the chroma of all image pixels) can together serve as a measure of the perceived naturalness and image quality. Their work has made use of a set of four different natural images.

A decision was taken by Division 1 (on Vision and Color) of the CIE (International Commission on Illumination) in May 1997 to recommend that the CIE adopt a color appearance model (to be known as *CIECAM-97*) [7]. It is understood that the ICC (International Color Consortium) is interested in the adoption of this model for use in their cross-platform color management system.

This paper examines the De Ridder work, and the CIE recommendation, and compares and contrasts them with our own data and conclusions.

In conclusion, it would appear that there is good evidence to support of the use of the *CIELUV* color model in measuring the color quality of images displayed on CRT monitors.

Color Appearance Models

The CIE has been recognized for more than 65 years as the world's leading authority on color science and as a major source of recommendations and standards. Attempts to find quantitative techniques for the measurement of color difference and the definition of color appearance started to take on their modern form in 1976 when the CIE adopted the *CIELAB* (CIE-1976(L*a*b*)) and *CIELUV* (CIE-1976(L*u*v*)) color spaces and color-difference formulae.

Within the CIE, matters pertaining to color are currently dealt with by Division 1 on Vision and Color. For some years now, this Division has been studying Color Appearance Models, with a view to the recommendation of one preferred model for industry use. Much of the original impetus for this effort was provided by the work of Hunt [8] and Nayatani *et al*. [9]. They independently proposed color models which provide measures of brightness and colorfulness (which are influenced by the viewing conditions) as well as the hue, saturation, lightness and chroma which are also available from the *CIELUV* and *CIELAB* models.

At its meeting held in Kyoto in May 1997, Division 1 decided to recommend that the CIE adopt an interim color appearance model, to be known as *CIECAM-97*. This model is relatively complex in its formulation, but it is said to be capable of providing an almost universal measure of color appearance since it takes into account the viewing conditions as well as the state of adaptation of the observer.

It is understood that this model has been forwarded to the International Color Consortium (ICC) who are said to be considering it for adoption as part of their proposed cross-platform color management system (CMS). The large variety of image source and destination devices currently in use can place particularly heavy demands on the CMS, bearing in mind the different viewing modes for self-luminous screens and hard-copy printout, and the different color analysis characteristics of the diverse range of image sources now in use. Given this range of diversity, it is understandable that the ICC should be considering the use of the *CIECAM-97* color model, despite its relatively high complexity.

Naturalness and Image Quality

De Ridder and co-workers at the IPO in the Netherlands have carried out a series of investigations of the "perceived naturalness" and "image quality" of color images of natural scenes, and their findings suggest that these two concepts are closely related [5,6]. Their work made use of four different natural scenes, and a number of color-distorted versions of them. The test images were presented one at a time on a video display monitor. The color appearance of each image was manipulated by means of precise distortions of hue, chroma, saturation or lightness, as computed in *CIELUV* color space. The perceived effects of these changes were measured using the assessments made by a number of human subjects on 10-point numerical category scales. It was found that, in general, there appeared to be a linear relationship between image quality and naturalness. In addition, their results showed that both quality and naturalness deteriorated as soon as the image hue angles were deviated from their original values.

Chroma or saturation variations affected the perceived quality and naturalness to a lesser extent than hue variations. The same four images were used, and chroma shifts were applied to every pixel of each image to create a range of new images with chroma error factors ranging from $\times$ 0.5 to $\times$ 2.0. The hue and lightness values in these manipulated images were left unchanged. It was found that both quality and naturalness reached similar peak values, but at different values of chroma, with naturalness peaking at a somewhat lower chroma and declining by the stage at which quality reached its peak. In other words, the subjects displayed a tendency to prefer more colorful images even though they evidently recognized that these images looked somewhat unnatural. It was found that generally similar trends were evident in the data for saturation changes.

A key feature of their results was the finding that the *CIELUV* chroma, and its scatter as measured by the standard deviation of the chroma of all the image pixels, can be combined to give a measure of the image quality. It is not clear whether the IPO work has included any investigation of alternative color models, and it cannot necessarily be concluded from their work that *CIELUV* space is the optimum model for this application. This is, however, one of the conclusions drawn from the following investigations.

The Subjective Experiment

In our own work, we have carried out a series of assessments on a sequence of 30 split-screen video scenes containing a range of semi-random color distortions [3]. Each scene contained one test image and one reference image.

Two test objects were employed, one being a MacBeth ColorChecker® test chart [4], and the other a printed copy of a photographic portrait which contained a large area of facial complexion. Video reproductions of these objects were made under the reference source and under a range of different test sources, many of which were deficient in terms of their color rendering properties. The TV camera controls were used to the full extent available in order to achieve as near correct as possible grey-scale rendition.

The magnitudes of the color shifts in these test images are considered to be of a similar order of magnitude to those likely to occur in most "real life" situations involving degraded image colors.

The color shifts were assessed by a group of 25 observers in a viewing room constructed to conform with CCIR standards [10]. Two five-point grading scales, both based on CCIR recommendations [11], were used: a comparison (or, perceptibility) scale and a quality (or, acceptability) scale. In the analysis of the subjective data, it became evident that the majority of the observers had shown a high degree of consistency between their perceptibility and acceptability judgments, and it was accordingly decided to normalize and combine the two scales to yield a single scale, termed the "mean subjective rating" (MSR) for each linked pair of test scenes.

The permissible values for this rating lie on a scale having a minimum of 10 (signifying a high degree of satisfaction among the observers, and close conformance between the test and reference images) to a maximum of 50 (signifying a high level of dissatisfaction, and very noticeable colour differences). The actual range of the MSR results for all images (and averaged over all observers) was from 23 to 42.

Digital Image Data Collection

Digitized images were acquired from the video tape of the test scenes. Their color content was analysed by using a program that determined the average of the gamma-corrected (R,G,B) pixel values contained within a series of hand-drawn rectangles, one on each color patch of interest, on every image. Some examples of these rectangles are shown in Figure 1. In all cases the reference half of the image was on the right-hand side. Twenty of the pairs of rectangles were drawn on the test-chart images, and three pairs were drawn over easily-identifiable, representative regions of the portrait images, *viz.* a facial complexion area, and sections of the front teeth and lower lip.

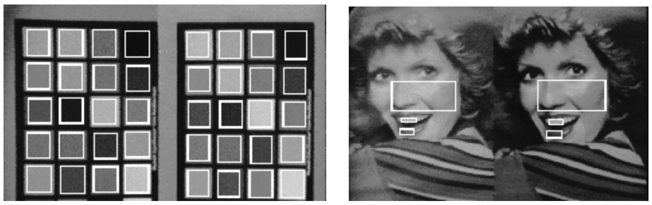

Fig 1 : Sampling Rectangles on Test Chart image (left) and Portrait image (right).

The average (R,G,B) pixel values within each rectangle were transformed into CIE colorimetric data, on the assumption of NTSC primaries and of a white point equivalent to Illuminant C, yielding the color differences between corresponding pairs of rectangles from the two images in each scene. In this way, 23 sets of color-difference data were computed for each linked pair of test scenes, and their average taken, to give an overall average measure of the color degradation.

Assuming NTSC display primaries, the transformation from the display (R,G,B) values to CIE-1931 (X,Y,Z) tristimulus values was as follows:

$$\left.\begin{aligned} X &= 0.607\,R + 0.174\,G + 0.201\,B \\ Y &= 0.299\,R + 0.587\,G + 0.114\,B \\ Z &= \phantom{0.607\,R + {}} 0.066\,G + 1.117\,B \end{aligned}\right) \quad (1)$$

and knowing that each pixel is encoded as 24 bits (*i.e.* 8 bits each for R, G, and B) it is possible to show that screen white is represented by:

$$R = G = B = 255 \quad (2)$$

yielding white-point tristimulus values of:

$$\left.\begin{aligned} X_0 &= 250.4 \\ Y_0 &= 255.0 \\ Z_0 &= 301.7 \end{aligned}\right) \quad (3)$$

The CIE-1976 chromaticity coordinates are defined as:

$$\left.\begin{aligned} u' &= 4\,X / (X + 15\,Y + 3\,Z) \\ v' &= 9\,Y / (X + 15\,Y + 3\,Z) \end{aligned}\right) \quad (4)$$

from which it is deduced that the white point is:

$$(u'_0, v'_0) = (0.2011,\ 0.4608) \quad (5)$$

Derived Data

Color differences were computed in the CIE-1976 (u',v') UCS system, and in the CIE-1976 (L*u*v*) and CIE-1976 (L*a*b*) color spaces, and also included the CIE-1976 saturation difference, as set out below.

The CIE-1976 chromaticity difference is defined as:

$$\Delta F = [(\Delta u')^2 + (\Delta v')^2]^{1/2} \quad (6)$$

and the CIE-1976 ($L^*u^*v^*$) coordinates are given by:

$$\left.\begin{aligned} L^* &= 116\,(Y / 255)^{1/3} - 16 \\ u^* &= 13\,L^*\,(u' - u'_0) \\ v^* &= 13\,L^*\,(v' - v'_0) \end{aligned}\right) \quad (7)$$

so that the CIE-1976 ($L^*u^*v^*$) color difference can be evaluated as :

$$\Delta E(L^*u^*v^*) = [(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2]^{1/2} \quad (8)$$

Similarly, the CIE-1976 ($L^*a^*b^*$) coordinates are given by:

$$\left.\begin{aligned} L^* &= 116\,(Y / Y_0)^{1/3} - 16 \\ a^* &= 500\,[(X / X_0)^{1/3} - (Y / Y_0)^{1/3}] \\ b^* &= 200\,[(Y / Y_0)^{1/3} - (Z / Z_0)^{1/3}] \end{aligned}\right) \quad (9)$$

and the CIE-1976 ($L^*a^*b^*$) color difference by:

$$\Delta E(L^*a^*b^*) = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2} \quad (10)$$

In addition, we derived the CIE-1976 saturation:

$$s_{(uv)} = [(u^*)^2 + (v^*)^2]^{1/2} / L^* \quad (11)$$

Average Color Differences

Briefly, the algorithm for computing the color differences for each linked pair of images was as follows:

For n = 1 ... 23

we computed the following color-difference measures:

ΔF_n ; $\Delta E_n(L^*u^*v^*)$; $\Delta E_n(L^*a^*b^*)$; $\Delta s_{(uv)n}$

and then found the average of each data set for each pair of images:

ΔF_{av} ; $\Delta E_{av}(L^*u^*v^*)$; $\Delta E_{av}(L^*a^*b^*)$; $\Delta s_{(uv)av}$.

To explore the correlations with the subjective data for each test image, we plotted scatter diagrams of these average colorimetric differences against the MSR, and computed the correlation coefficient ***r*** in each instance. Table 1 shows the correlation coefficients for the four different methods of color-difference computation, and Fig. 2 shows the scatter-plot the best-correlated set of data (viz. $\Delta E_{av}(L^*u^*v^*)$ against the MSR).

The results included in Table 1 are those for which we found "useful" levels of correlation (*i.e.* ***r*** > 0.5) with the subjective data. Other colorimetric measures were also investigated, but showed significantly less correlation and were excluded.

Since the *CIELUV* system clearly showed the superior performance, it was decided to standardize on this system in the next phase of the work.

Table 1: Correlation of Average Color Differences with the MSR

Method of Calculation	Correlation Coefficient *r*
$\Delta E_{av}(L^*u^*v^*)$	0.76
$\Delta s_{(uv)av}$	0.67
$\Delta F_{av}(u',v')$	0.58
$\Delta E_{av}(L^*a^*b^*)$	0.51

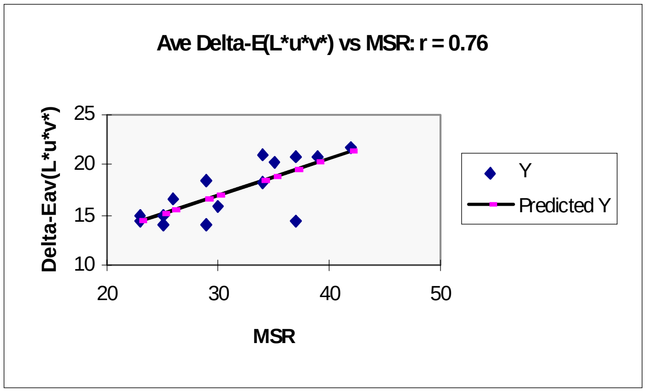

Fig 2 : Scatter-plot of $\Delta E_{av}(L^*u^*v^*)$ vs. MSR

Individual Color Differences

This involved the assessment of which of the colored regions in the test images had had the greatest influence on the MSR, by testing for the correlation of the individual (rather than an average of 23) $\Delta E(L^*u^*v^*)$ values with the MSR. The results are summarized in Table 2, and are presented in full in Table 3.

Table 2: Correlation of Individual *CIELUV* Color Differences with the MSR: Summary

Sample details	Numbers	Percentages
Total	23	100 %
Yielding r > 0	15	65 %
Yielding r > 0.5	9	39 %
Yielding r < 0	8	35 %

It is noteworthy that only nine of the positive correlations gave a value of ***r*** greater than 0.5 (shown in bold type in Table 3). Six of these were color patches from the ColorChecker, including one skin color and five colors of moderate to high chroma. All three of the selected color regions of the portrait image gave values of ***r*** of well over 0.6. This is thought to indicate the greater attention paid by the observers to areas such as these (*i.e.* teeth, lips, and facial complexion) in arriving at their assessments. Of the eight negative correlations for the patches on the ColorChecker, the majority (five) were for the neutral colors and the remaining three were for moderate to low chroma samples in the blue range of hues.

Table 3: Correlation of Individual CIE-1976 (L*u*v*) Color Differences with the MSR

Ref	Color Name	ΔE (L*u*v*) Range	Correlation Coefficient r
01	**Bluish-Green**	**5.1 - 18.2**	**+ 0.764**
02	Blue Flower	4.7 - 13.8	– 0.217
03	Foliage	12.3 - 36.0	+ 0.439
04	Blue Sky	3.9 - 17.6	– 0.538
05	**Light Skin**	**9.4 - 23.8**	**+ 0.636**
06	Orange-Yellow	6.1 - 35.7	+ 0.453
07	**Yellow-Green**	**11.2 - 28.9**	**+ 0.612**
08	Purple	10.8 - 33.9	+ 0.281
09	**Moderate Red**	**17.0 - 47.3**	**+ 0.597**
10	Purplish-Blue	4.2 - 26.1	– 0.504
11	Cyan	9.5 - 23.3	+ 0.182
12	**Magenta**	**3.8 - 30.1**	**+ 0.763**
13	Yellow	8.2 - 27.2	+ 0.365
14	**Red**	**17.3 - 53.9**	**+ 0.736**
15	Green	8.1 - 21.6	+ 0.249
16	Black	3.1 - 34.6	– 0.190
17	Neutral 3.5	7.7 - 22.3	– 0.071
18	Neutral 5.0	2.1 - 14.9	– 0.366
19	Neutral 6.5	2.8 - 10.2	– 0.402
20	Neutral 8.0	2.2 - 8.7	– 0.367
F-1	**Complexion**	**6.2 - 25.3**	**+ 0.764**
F-2	**White Teeth**	**7.2 - 26.7**	**+ 0.694**
F-3	**Red Lips**	**12.2 - 66.2**	**+ 0.636**

Thus, the portrait image was shown to contain specific areas for which consistent correlation could be found between the *CIELUV* color differences and the perceived color quality of the image, while there is some interesting, though less consistent, evidence from the patches of the ColorChecker.

Standard Deviations of Chroma

The color data for the abovementioned rectangles has also been used in a preliminary investigation of De Ridder's hypothesis [6] that the standard deviation of the *CIELUV* chroma for all the pixels in an image has a significant degree of correlation with the perceived image quality.

Chroma is defined in the CIE-1976 systems as:

$$C(u^*v^*) = [(u^*)^2 + (v^*)^2]^{1/2} \quad (12)$$

$$C(a^*b^*) = [(a^*)^2 + (b^*)^2]^{1/2} \quad (13)$$

*C(u*v*)* and *C(a*b*)* have been computed for all of the rectangles in each linked pair of test images. The resultant 23 chromas have been statistically analysed to find their average and standard deviation. For all 15 image pairs the standard deviation has been compared with the MSR to assess whether there is any significant correlation. The correlation of the standard deviation in *C(u*v*)* with the MSR is depicted in Fig. 3, and an overall summary is given in Table 4.

Note that Fig. 3 has plotted the "inverse" MSR along the abscissa. This has been defined as:

$$\text{Inv MSR} = 50 - \text{MSR} \quad (14)$$

in order to yield a regression line with a positive slope. This is justified on the grounds that MSR is defined such that a higher MSR signifies a greater difference between test and reference images (and lower test-image quality); and we are wanting to test here for correlation between greater scatter in the chromas and higher perceived image quality.

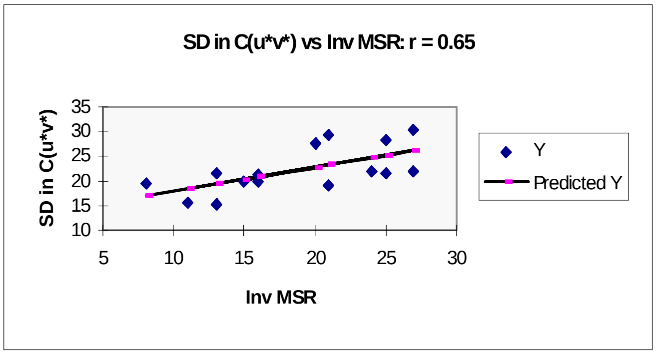

*Fig. 3: Scatter-plot of Std. Dev. in C(u*v*) vs. the Inverse MSR*

In addition to the chroma, the *CIELUV* color vector magnitude:

$$E(L^*u^*v^*) = [(L^*)^2 + (u^*)^2 + (v^*)^2]^{1/2} \quad (15)$$

has been computed for all 23 rectangles, and the standard deviation in the *E* values derived, for each linked pair of test images. It is seen that the effect was to reduce the correlation coefficient below that obtained with *C(u*v*)* - most likely because the incorporation of the *L** data leads to vectors of more nearly equal magnitude.

Table 4: Correlation of Standard Deviations of Chromas with the MSR (preliminary)

Method of Calculation	Correlation Coefficient *r*
SD in *CIELUV* chroma *C(u*v*)*	0.65
SD in *CIELAB* chroma *C(a*b*)*	0.55
SD in *CIELUV* colour *E(L*u*v*)*	0.58

It is clear that the *CIELUV* data gives a significantly superior result by comparison with *CIELAB*. It is stressed that the method used here has ***not*** fully investigated De Ridder's hypothesis since we have worked with averaged (R,G,B) data from within each rectangle, whereas De Ridder has worked with the (R,G,B) data for every individual pixel in the entire image.

The result is encouraging, however, not merely because it gives general support to De Ridder's hypothesis, but chiefly because it holds out hope of defining an objective measure of image color quality that relies solely on the image itself, without recourse to a reference image.

Conclusions

It may be concluded from the foregoing that a measure of color image quality for natural images is within our grasp, but that it evidently remains to be determined what its optimum form may be.

It is quite likely that there is no overall "best solution" and that we must rather consider the prospect of optimising for specific applications. For example, our results, and those achieved by De Ridder *et al.*, suggest that the *CIELUV* color model is acceptable for the classification of quality in color images reproduced on self-luminous displays. The ICC, on the other hand, has up to now tended to favour the use of the *CIELAB* model, and appears to be moving toward the *CIECAM-97* model. Our results do not support the use of *CIELAB*, but our experimental range has been confined to self-luminous displays, and is therefore more restricted than that being addressed by the ICC.

Future Work

There are several extensions to this work planned for the future:

- the collection of complete color data from all pixels in each test image, to fully test the De Ridder hypothesis on our images;
- re-processing of our data in *CIECAM-97* space, for comparison with the data presented here;
- the collection of new subjective data for our images with the intention of strengthening our statistics.

Acknowledgments

The author gratefully acknowledges the support and assistance of the Research Subcommittee of the Manukau Institute of Technology, and of his Head of Department and colleagues in the Department of Electrical and Electronic Engineering.

References

1. A.N. Chalmers, Procs. IVCNZ-95, pp. 203-208 (1995).
2. A.N. Chalmers, Procs. AIC Color-97 Kyoto, Vol. 2, pp. 634-637 (1997).
3. A.N. Chalmers, Trans. SAIEE, Vol. 17, pp. 89-96 (1986).
4. C.S. McCamy, H. Marcus, & J.E. Davidson, Jour. Appl. Phot. Eng., Vol. 2, pp. 95-99 (1976).
5. H. De Ridder, F.J.J. Blommaert & E.A. Fedorovskaya., SPIE Procs., Vol. 2411, pp. 51-61 (1995).
6. E.A. Fedorovskaya, H. De Ridder & F.J.J. Blommaert, Color Res. & Applic., Vol. 22, pp. 96-110 (1997).
7. A. N. Chalmers, personal notes of CIE Division 1 meeting, Kyoto (1997).
8. R.W.G. Hunt, Color Res. & Applic., Vol. 16, pp. 146-165 (1991).
9. Y. Nayatani, K. Takahama and H. Sobagaki, Color Res. & Applic., Vol. 15, pp. 210 - 221 (1990).
10. CCIR, Recommendation 500 (1974).
11. CCIR, Recommendation 405-2 (1974).