https://doi.org/10.2352/ISSN.2169-2629.2018.26.110


# Comparative Evaluation of Color Differences between Color Palettes

***Qianqian Pan*[a]*, Stephen Westland*[a]**
**[a]*School of Design, University of Leeds, Leeds, West Yorkshire, United Kingdom***

## Abstract

*The difference or distance between two color palettes is a metric of interest in color science. It allows a quantified examination of a perception that formerly could only be described with adjectives. Quantification of these properties is of great importance. The objective of this research is to obtain the dataset for perceptual colour difference between two color palettes and develop color difference metric(s) to correspond well with the perceptual color difference. The psychophysical experiment was carried out using Magnitude Estimation method. Three different color difference metrics, namely Single Color Difference Model (Model 1), Mean Color Difference Model (Model 2), and Minimum Color Difference Model (Model 3), respectively, have been proposed and compared. Data analysis include regression analysis, statistical STRESS analysis, and examination of observer variability using coefficient of variance (CV). The results show that the Minimum Color Difference Model (Model 3) outperformed the other two with a coefficient of determination (R-squared) value of 0.603 and an STRESS value of 20.95. In terms of observer variability, the average intra-observer variability is 17.63 while the average inter-observer variability is 53.73.*

*Keywords: color palettes, color difference, magnitude estimation, STRESS, psychophysics*

## Introduction

In color theory, a color palette is the choice of colors used in design for a range of media. A color palette usually contains a number of chromatic and/or monochromatic colors. There are many color palette generators currently available, among which is the well-known Adobe Color CC (previously Adobe Kuler) where a color palette, typically consisting of about 5 individual colors, can be generated either from a color wheel or from an imported image based on a user's choice of preferences (Color Rule), such as 'analogous', 'monochromatic', 'triad', 'complementary', 'compound', 'shades'. Color palette generation and color palette preference can both be subjective based on a designer's color preference and knowledge of aesthetics. When extracting colors from an image using computational methods (for example, clustering) it is common that the colors extracted will depend upon the parameters of the method. For example, with color clustering the way in which the centroids are initially selected will typically affect the final colors that result. This leads to the natural question of what the optimum parameters are. Imagine that we have several color palettes extracted from an image using different computational parameters and that we want to decide which of these palettes best matches a color palette extracted visually. We therefore need a method for estimating the color difference between two palettes. Relatively little research has been carried out directly on comparative evaluation of color palettes. Tokumaru *et al.* [1] published work on the evaluation of a color scheme's harmony in 2000. Besides the application of color schemes in product and interior design, color palettes can also be applied in colour image quantization [2-4] in computer graphics and image processing. Image and video quality is often assessed by image comparison [5]. The image comparison often involves pixel-by-pixel comparison when the images display the same content or scene. Other image comparison metrics include keypoint matching [6], histogram method [7], and keypoint + decision tree [8], etc.

The problem of color-palette difference is analogous to the problem of color-difference prediction of pairs of color patches; the traditional color difference problem can be considered to be a special case of a more general problem where we need to compare a pair of several patches (i.e. palettes). Much research has been conducted on this special case - the evaluation of color difference between homogenous colors [9-11]. For a pair of homogenous color samples or two complex images viewed under specific conditions, color-difference formulae try to predict the visually perceived (subjective) color difference from instrumental (objective) color measurements. Current color difference formulas (e.g. CIEDE2000) for homogenous colors are based on various sets of empirical difference perception data established with different kinds of materials, under different evaluation conditions, and with different observer panels.

This research is mainly focused on comparative evaluation of color difference between color palettes.

## Experiment Design

### Research Methodology

To examine the color difference between two color palettes, a suitable color difference metric needs to be developed in order to calculate the color difference ($\Delta E$) between the two palettes. A psychophysical study also needs to be conducted in order to investigate the visual difference between pairs of color palettes ($\Delta V$). The $\Delta V$ data is then used to test the performance of the color difference metric by examining the correlation between $\Delta E$ and $\Delta V$. This is the same approach that has been used successfully over the past fifty years ago that has led to CIEDE2000 for homogenous colors (of single patches).

In this study, three different methods are proposed for the color difference metric between two color palettes. The strength of the visual color difference between pairs of color palettes is determined using a psychophysical scaling experiment. In general, psychophysical scaling methods are developed to find the relationship between physical stimuli and human sensation.

The psychophysical method used was Magnitude Estimation (ME). Following the wide and successful application of ME in color-difference research, this research also used ME as a research method.

### Color Palettes Preparation

A set of 30 landscape images were used to generate color palettes using a color-based clustering method; more specifically, using k-means clustering [12]. The top 25 colors generated from each image were stored and used to form a color palette (See Figure 1). A set of 30 color palettes were prepared in this way. Note, however, that to a large extent the method of producing pairs of palettes was arbitrary. All that was required was a method to produce a set of color palettes that could be used to form pairs with varying visual difference between them.

FIGURE 1: ILLUSTRATION OF SAMPLE COLOR PALETTE PREPARATION: (A) ORIGINAL LANDSCAPE SAMPLE IMAGE, AND (B) 25 KEY COLORS EXTRACTED FROM THE LANDSCAPE IMAGE ON THE LEFT.

### Psychophysical Experiment

One of the purposes of this study is to find out the visual difference between pairs of color palettes; therefore, the color palettes were presented pairwise. The total number of pairs of color palettes generated from the set of 30 color palettes was 435 pairs ($30 \times 29/2$). It was considered too time-consuming to ask each observer to view 435 pairs during the experiment since observer fatigue could become an issue. Thus, 96 pairs were randomly selected from the 435 pairs. In addition, 20 of the 96 pairs were duplicated (and selected randomly) in order to allow a measure of repeatability. This resulted in 116 pairs ($96 + 20$) to be included in the psychophysical experiment. 30 observers were recruited, each of whom passed the Ishihara color vision deficiency test. Color palettes were then displayed in pairs on an LED computer monitor. Observers were asked to assign a number between 0 and 100 to describe the color difference between the two color palettes in each pair. Each observer went through a training session on the computer with 8 sample pairs before commencing the experiment (observers also viewed all 30 color palettes prior to commencing the experiment in order to familiarize themselves with the range). The psychophysical experiment was encoded in MATLAB. An example of the Graphic User Interface (GUI) is shown in Figure 2. The background color of the GUI had RGB values of [128,128,128]. The psychophysical experiment was conducted in a dark room and observers viewed the display from a distance of about 80cm. The size of each of the palettes on the display was approximately 10cm × 10cm.

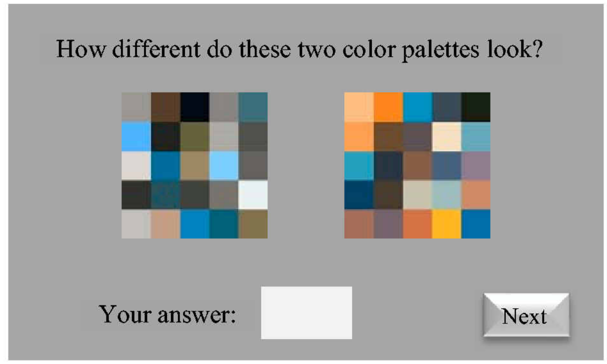


FIGURE 2: AN EXAMPLE OF THE GUI USED IN THE PSYCHOPHYSICAL EXPERIMENT.

### Color Difference Algorithms Design

In this study three different algorithms are proposed for calculating the color difference ($\Delta E$) between two color palettes, namely Single Color Difference Model, Mean Color Difference Model, and Minimum Color Difference Model. These are described in detail below for the case where the palettes each have 25 colors (which is the case in this study).

#### *Single (homogenous) Color Difference Model*

This model represents each color palette as a single color that is the average of all the colors in the palette. This can be achieved by the following steps taking (Figure 3 as an example) thus:

1. Each color patch in the palette (see Figure 3 (a)) can be represented by one set of RGB data since the color in each patch is uniform.
2. The RGB data form a 25 by 3 data matrix.
3. Obtain the mean$\overline{R}$, mean$\overline{G}$ and mean$\overline{B}$ values by averaging the data in each column of the matrix.
4. The mean$\overline{R}$, mean$\overline{G}$ and mean$\overline{B}$ values are used to represent a single color (see Figure 3, (b)).
5. Step 1 to 4 are repeated on a different color palette.
6. The $\Delta E$ is then simply the CIELAB color difference between these two single colors obtained.

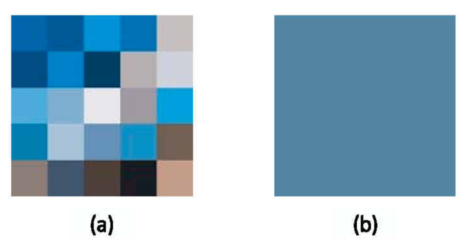

FIGURE 3: AN ILLUSTRATION OF THE SINGLE COLOR DIFFERENCE MODEL, (A) A SAMPLE COLOR PALETTE, AND (B) THE COLOR YIELD BY AVERAGE THE 25 COLORS IN (A).

#### *Mean Color Difference Model*

The concept of this algorithm is to compare the color of each patch in one palette with the color of each patch in another palette and to take an average of the color differences. The details are described thus:

1. Each color in one palette can form a color pair with each color in another palette. Thus, each color in one palette will form 25 color pairs with the 25 colors in another palette. This will result in $25 \times 25 = 625$ color pairs.
2. The CIELAB color difference is calculated between each color pair resulting in a set of 625 color difference values.
3. The mean color difference is obtained by averaging the 625 color difference data to represent the color difference between the two color palettes.

*Minimum Color Difference Model*

The concept of this algorithm is that for each color in one palette, there will be a corresponding color (or even more than one color) in another palette that it most closely matches. Thus:

1. For each color in one palette, the CIELAB color difference between this color and each of the colors in the second palette are calculated. The minimum color difference is recorded.
2. Step 1 is repeated for all the colors in the first palette, finding their closest corresponding colors in the second palette, resulting in 25 color differences.
3. The 25 minimum color difference values are averaged and the mean value symbolized as $m_1$.
4. Steps 1-3 are repeated, but this time for each of the color in the second palette. In other words, for each of these colors the closest corresponding color in the first palette is found. The mean value of these 25 color differences is symbolized as $m_2$.
5. The values of $m_1$ and $m_2$ are averaged to obtain the color difference between the two palettes.

## Results

For 30 color palettes, each containing 25 individual colors, there are in total 750 colors used in this study. The CIE tristimulus values of each of the 750 colors were measured on the screen using the Konica Minolta CS-2000 spectroradiometer. The XYZ values were converted to CIELAB values using the white of the display as the white point. All color differences in this study were calculated using the CIELAB $\Delta E^*_{ab}$ formula shown in equation (1).

$$\Delta E^*_{ab} = \sqrt{(L^*_2 - L^*_1)^2 + (a^*_2 - a^*_1)^2 + (b^*_2 - b^*_1)^2} \qquad (1)$$

**Regression Analysis between Proposed Color Difference Metrics and Perceptual Color Difference**

With respect to color difference, it is usual to distinguish between the visual color difference (ΔV) and the computed color difference (ΔE) for a pair of color palettes. ΔV is the visual color difference between two color palettes perceived by human observers – that is, the answer from of the human visual system obtained from the psychophysical experiment [13]. The computed color difference ΔE is the result provided by a color-difference formula, i.e., each of the three proposed color difference algorithms in our study. Ideally, ΔE should approach ΔV as close as possible. It would be desirable to have a simple mathematical function between ΔV and ΔE in such a way that ΔV would be accurately predicted from ΔE for any pair of color palettes visually assessed under any fixed set of experimental conditions. In this study, ΔV was obtained by taking the geometric mean of all participants assessment on each pair of color palettes. The performance of each of the three color difference algorithms was further discussed in this section.

*Single (homogenous) Color Difference Model*

The performance of this model is evaluated in Figure 4 and Table 1. From Table 1 the correlation coefficient (R) value is close to 0.60 while the coefficient of determination (R-squared) is 0.35.

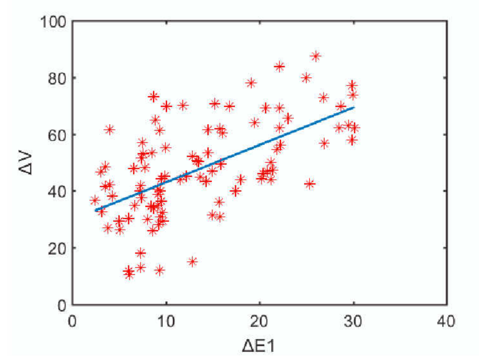


**FIGURE 4: CORRELATION BETWEEN ΔE1 OBTAINED FROM THE SINGLE COLOR DIFFERENCE MODEL AND THE VISUAL COLOR DIFFERENCE ΔV USING A LINEAR FITTING.**

Table 1 lists the estimated coefficients $b_1$, $b_2$ and the Goodness-of-Fit statistics root-mean-square error (RMS Error) and coefficient of determination (R-squared) values from three different curve-fitting models, including linear regression model, non-linear log fit model, and non-linear power fit model. It shows that the linear fit model presents the best performance among all.

| | Linear fit | Log fit | Power fit |
|---|---|---|---|
| Function | $y = b_1+b_2x$ | $y= b_1+b_2*\log(x)$ | $y = b_1+x^{b2}$ |
| $b_1$ | 1.316 | 10.487 | 19.402 |
| $b_2$ | 29.779 | 15.254 | 0.358 |
| RMS Error | 13.8 | 14.3 | 14 |
| R-squared | 0.353 | 0.301 | 0.329 |

**TABLE 1: CURVE-FITTING PARAMETERS FOR LINEAR FIT, LOG FIT AND POWER FIT, RESPECTIVELY (SINGLE COLOR DIFFERENCE MODEL).**

*Mean Color Difference Model*

The performance of the mean color difference model is evaluated in Figure 5 and Table 2. From Table 2 the correlation coefficient (R) value is close to 0.35 while the coefficient of determination (R-squared) is 0.12.

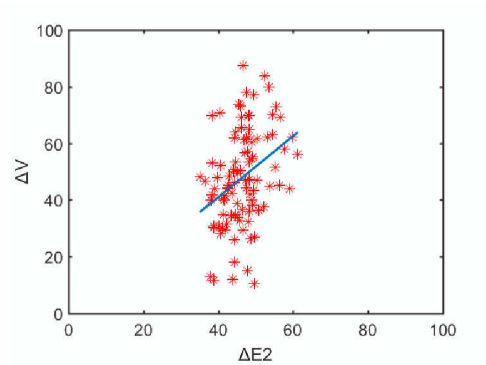


**Figure 5: Correlation between ΔE2 obtained from the mean color difference model and the visual color difference ΔV using a linear fitting.**

| | Linear fit | Log fit | Power fit |
|---|---|---|---|
| Function | $y = b_1+b_2x$ | $y= b_1+b_2*\log(x)$ | $y = b_1+x^{b2}$ |
| $b_1$ | 1.078 | -145.630 | 0.887 |
| $b_2$ | -2.171 | 50.511 | 1.039 |
| RMS Error | 16.10 | 16.10 | 16.1 |
| R-squared | 0.119 | 0.119 | 0.119 |

**Table 2: Curve-fitting parameters for linear fit, log fit and power fit, respectively (mean color difference model).**

*Minimum Color Difference Model*

The performance of the minimum color difference model is evaluated in Figure 6 and Table 3. From Table 3 the correlation coefficient (R) value is close to 0.78 while the coefficient of determination (R-squared) is 0.60.

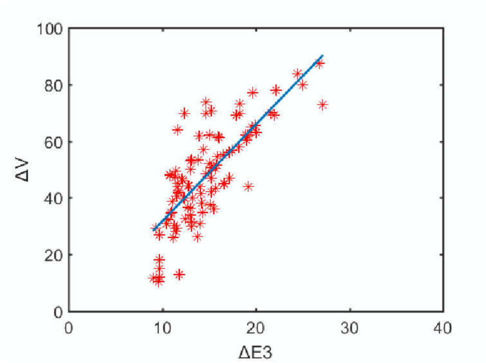


**Figure 6: Correlation between ΔE3 obtained from the minimum color difference model and the visual color difference ΔV using a linear fitting.**

| | Linear fit | Log fit | Power fit |
|---|---|---|---|
| Function | $y = b_1+b_2x$ | $y= b_1+b_2*\log(x)$ | $y = b_1+x^{b2}$ |
| $b_1$ | 3.412 | -97.475 | 3.090 |
| $b_2$ | -2.483 | 54.636 | 1.019 |
| RMS Error | 10.70 | 10.40 | 10.7 |
| R-squared | 0.603 | 0.630 | 0.606 |

**Table 3: Curve-fitting parameters for linear fit, log fit and power fit, respectively (minimum color difference model).**

The linear regression analysis indicates that the minimum color difference model is the best-performing model with $r^2$ of 0.60. However, the long literature on the development of color-difference equations has tended to prefer other measures of fit such as PF/3 and STRESS. In this study, STRESS is utilized and it is described in the next section.

**Standardized Residual Sum of Squares (STRESS)**

To test the performance of different color-difference formulae, the STRESS index employed in multidimensional scaling (MDS) techniques has been found particularly useful and is used. STRESS is defined in Equations 2 and 3.

$$STRESS = 100 \times \left(\frac{\sum(\Delta E_i - F_1 \Delta V_i)^2}{\sum F_1^2 \Delta V_i^2}\right)^{1/2} \quad (2)$$

$$F_1 = \frac{\sum \Delta E_i^2}{\sum \Delta E_i \Delta V_i} \quad (3)$$

where $\Delta V_i$ and $\Delta E_i$ are the visual and computed colour differences for the $i = 1, \dots, n$ pairs of colour palettes, respectively. $F_1$ is a factor adjusting the scales of $\Delta V_i$ and $\Delta E_i$. STRESS is always in the range 0 – 100. Greater values mean worse agreement between visual and computed color differences. An idea color difference formula would produce a STRESS of zero [14]. The STRESS values are 39.16 (Single Color Difference Model), 31.33 (Mean Color Difference Model) and 20.95 (Minimum Color Difference Model). This indicates that the Minimum Color Difference Model has the best agreement between visual and computed color difference.

The squared ratio of the STRESS values from two color-difference formulas follows an F-distribution. This is used to determine whether different color-difference formulae are statistically significantly different at any confidence level (usually 95%). The squared ratio of STRESS are:

Single Color Difference Model v Mean Color Difference Model = 1.562.

Mean Color Difference Model v Minimum Color Difference Model = 2.236.

Single Color Difference Model v Minimum Color Difference Model = 3.494.

The $F_C(0.975, 95, 95) = 0.667$, where the number 95 comes from the number of pairs of color palettes in this dataset, 96, minus one. Because all the three squared ratio values are outside the confidence interval $[F_C; 1/F_C] = [0.667; 1.449]$, it is concluded that these three formulae are significantly different from each other at a 95% confidence level.

**Observer Variance**

Observer variance is examined using Coefficient of Variance (CV). The corresponding equation is as follows:

$$CV = 100\left[\sum_{i=1}^{n}(x_i - y_i)^2/n\right]^{1/2}/\bar{y} \quad (4)$$

*(I) Intra-observer Variance*

There were 20 pair of color palettes that observers evaluated twice during the psychophysical experiment. Intra-observer variance examines variances between each observer's first trial and his/her second trial. For each observer, the CV for each pair of stimuli is calculated; $x_i$ is the ME data for each observer for the $i$'s pair of stimuli from the first trial, where as $y_i$ is the ME data for each observer for the $i$'s pair of stimuli from the second trial. $\overline{y}$ is the mean of $x_i$ and $y_i$. The CV is then averaged across all pairs of stimuli to represent the inter-variance for each observer. The average CV among all observers is then obtained to represent the average CV of intra-observer variance. The average intra-observer variance is 17.63.

*(II) Inter-observer Variance*

Inter observer variance examines the variance across observers. Here the CV equation can be simplified as the ratio of the standard deviation $\sigma$ to the mean $\mu$ in equation (5)

$$CV = \frac{\sigma}{\mu} \quad (5)$$

where $\sigma$ is the standard deviation of the ME data obtained from all 30 observers for each pair of stimuli and $\mu$ is the mean of the ME data from all 30 observers for each pair of stimuli. Equation (5) can also be written in the form in equation (6)

$$CV = 100[\sum_{i=1}^{n}(x_i - \overline{x})^2/n]^{1/2}/\overline{x} \quad (6)$$

where $x_i$ is the ME data from the $i$'s observer for each pair of stimuli, $\overline{x}$ is the mean ME data across all 30 observers for each pair of stimuli. The average inter-observer variance is 53.73. The fairly great inter-observer variance value indicates a great variability of psychophysical scale of each observer without the presence of the anchoring (reference) points. Setting anchoring is commonly used in the ME experiment for the examination of colour difference between single color patch [15]. However, there is very little literature on setting anchoring for stimuli such as color palettes. The lack of anchoring led to a fairly great inter-observer variance compared to other researcher's work where anchoring was set in various fashion. Figure 6 indicates that even without anchoring there is still a good correlation between the color difference metric (ΔE) and the perceptual color difference (ΔV). This suggests that anchoring is not a compulsory, but it may help eliminating the cognitive bias from the observers during the psychophysical experiment. It is now well-known that observers can differ significantly in how they assess differences and as a result the number of observers participating in an experiment becomes a factor also. It has recently become quite clear that observers likely present by far the largest component of total variability [16]. It appears that individual observers have a personally set and relatively reliable relationship between achromatic and chromatic differences that varies widely. This raises serious issues in obtaining reliable data for a world-average observer. Individuals differ genetically in many respects (including the workings of their colour vision apparatus) and have widely different life histories of visual experiences makes it unsurprising that stimuli and stimuli differences are interpreted differently.

## Conclusions

The comparative evaluation of color differences between color palettes is a metric of interest in color science and in design. This study has generated some psychophysical data that has been used in this study but could be used by other researchers (and is available on request). However, in this study we proposed three different color difference metrics for the evaluation of color difference between color palettes, namely the Single Colour Difference Model, the Mean Colour Difference Model, and the Minimum Colour Difference Model. The performance of different color metrics has been compared using various analysis methods including regression analysis, statistical analysis (STRESS). In the regression analysis, the relationship between the ΔE and ΔV was investigated using curve-fitting with linear function, log function and power function. The coefficient of determination (R-squared) for three different metrics are 0.35 (Single Colour Difference Model), 0.12 (Mean Color Difference Model) and 0.60 (Minimum Color Difference Model). The STRESS results for the performance of the three models are 39.16, 31.33 and 20.95, respectively. The three models are also significantly different from each other. Overall, the Minimum Colour Difference Model outperformed the other two metrics based on all analysis. The observer variability has also been examined using coefficient of variance (CV) for intra-observer and inter-observer variance. The average intra-observer CV is 17.63 while the average inter-observer CV is 53.73 due to the fact that the observers can differ significantly in how they assess differences. This research work can be served as the fundamental work for quantifying the visual difference between two color palettes.

Of course, this work has made a first step towards the quantification of color differences between palettes but a number of restrictions should be noted. This work was carried out with color palettes consisting of 25 color patches arrange in a 5 x 5 grid. Other spatial arrangements or different ordering of the patches could affect the visual color difference as could a range of other factors such as the number of patches in the palette and the surround color against which the palettes are viewed.

## References

[1] M. Tokumaru, N. Muranaka, S. Imanishi. Color design support system considering color harmony. InFuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference, vol. 1, pp. 378-383, 2002.

[2] V. Oleg, and J. W. Buchanan. "The local K-means algorithm for colour image quantization." In Graphics Interface, CANADIAN INFORMATION PROCESSING SOCIETY, pp. 128-128, 1995.

[3] P.Scheunders. A comparison of clustering algorithms applied to color image quantization. Pattern Recognition Letters. Nov 1;18(11-13), pp. 1379-84, 1997.

[4] X Wan, CC Kuo. A new approach to image retrieval with hierarchical color clustering. IEEE transactions on circuits and systems for video technology, Sep;8(5), pp. 628-43, 1998.

[5] N Ponomarenko, F Battisti, K Egiazarian, J Astola, V Lukin. Metrics performance comparison for color image database. In Fourth international workshop on video processing and quality metrics for consumer electronics vol. 27, 2009.

[6] DG Lowe. Distinctive image features from scale-invariant keypoints. International journal of computer vision, Nov 1;60(2), pp. 91-110, 2004.

[7] G Pass, R Zabih, J Miller. Comparing images using color coherence vectors. In Proceedings of the fourth ACM international conference on Multimedia, pp. 65-73, 1997.

[8] D Nister, H Stewenius. Scalable recognition with a vocabulary tree. In Computer vision and pattern recognition, 2006 IEEE computer society conference, vol. 2, pp. 2161-2168), 2006.

[9] M. Melgosa, A.Trémeau, and G Cui, Colour difference evaluation. In Advanced Color Image Processing and Analysis, pp. 59-79. Springer, New York, NY, 2013

[10] K. Witt, CIE guidelines for coordinated future work on industrial colour-difference evaluation. Color Research & Application. Dec 1;20(6), pp. 399-403, 1995.

[11] G Cui, MR Luo, B Rigg, W Li. Colour-difference evaluation using CRT colours. Part I: Data gathering and testing colour difference formulae. Color Research & Application. Oct 1;26(5), pp. 394-402, 2001.

[12] S Ray, RH Turi. Determination of number of clusters in k-means clustering and application in colour image segmentation. In Proceedings of the 4th international conference on advances in pattern recognition and digital techniques, Dec 27, pp: 137-143, 1999.

[13] E. D. Montag and D. C. Wilber, A comparison of constant stimuli and gray-scale methods of color difference scaling, Color Res Appl. 28, pp. 36-44, 2003.

[14] M Melgosa, R Huertas, RS Berns. Performance of recent advanced color-difference formulas using the standardized residual sum of squares index. JOSA A. Jul 1;25(7), pp. 1828-34, 2008.

[15] G Sharma, W Wu, EN Dalal. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. Color Research & Application. Feb 1;30(1), pp. 21-30, 2005.

[16] RG Kuehni. Variability in estimation of supra-threshold small color differences. Color Res Appl, Oct 1: 34(5), pp. 367-74, 2009.

## Author Biography

***Qianqian Pan*** *received her BSc in Textile Engineering from Jiangnan University and her PhD in Colour and Imaging from University of Leeds. Since then she has worked as a Research Fellow in the Colour and Imaging group at University of Leeds. Her work has focused on colour appearance, colour modelling and image processing.*

***Stephen Westland*** *received his BSc in Colour Chemistry and PhD in Colour Physics both from University of Leed, where he is now Professor of Colour Science and Technology. His research interests include colour measurement, colour design, colour management, sustainable coloration and colour vision.*