Article:
Yang, J, Chen, Y, Westland, S et al. (2020) Predicting visual similarity between colour palettes. Color Research and Application, 45 (3). pp. 401-408. ISSN: 0361-2317

https://doi.org/10.1002/col.22492

Predicting Visual Similarity Between Colour Palettes

Jie Yang, Yun Chen, Kaida Xiao and Stephen Westland
School of Design, University of Leeds (UK)

Correspondence:
Yun Chen
sdyc@leeds.ac.uk

Abstract

This work has is concerned with the prediction of visual colour difference between pairs of palettes. In this study, the palettes contained 5 colours arranged in a horizontal row. A total of 95 pairs of palettes were rated for visual difference by 20 participants. The colour difference between the palettes was predicted using two algorithms, each based on one of six colour-difference formulae. The best performance (r^2 = 0.86 and STRESS = 16.9) was obtained using the minimum colour difference algorithm (MICDM) using the CIEDE2000 equation with a lightness weighting of 2. There was some evidence that the order (or arrangement) of the colours in the palettes was a factor affecting the visual colour differences although the MICDM algorithm does not take order into account. Application of this algorithm are intended for digital design workflows where colour palettes are generated automatically using machine learning and for comparing palettes obtained from psychophysical studies to explore, for example, the effect of culture, age or gender on colour associations.

1. Introduction

Data is increasingly being used to optimise design. Recent advances in hardware processing speeds, software and connectivity are together leading to the availability of huge amounts of data and the ability to process those data in order to generate deep and meaningful insights.[1] As a consequence, we are seeing the emergence of design tools that are driven by machine learning and powered by vast amounts of data. Machine learning is being used to develop, for example, fashion retail forecasting systems.[2-3] However, recent advances in machine learning offer the potential for software to generate original designs. For example, the recent discovery of Generative Adversarial Networks (GANs) have recently received attention[4] for their application to generating fake (but original) images and videos knows as 'deepfakes' which has some worrying implications.[5] More positive applications of GANs, however, may be the ability to create new original designs and this may pave the way for generative design systems in colour, fashion and design.[6] Rapid and digital workflows in the operation of these systems will require automatic methods for evaluating and/or comparing colours and designs.[7-8] The work described in this study is concerned with methods for automatically predicting the visual similarity between colour palettes.

Colour palettes are ubiquitous in colour design. A colour palette is a collection of colour patches that represents the colours used in a design or an image. Interest in developing and sharing colour palettes has grown considerably in recent years and has led to a proliferation of digital tools that can generate and share colour palettes. For example, the Adobe Color website[9] allows users to create colour palettes according to different rules for colour harmony; for example, colour palettes that can be described as analogous or complementary[10]. The Adobe website also allows users to automatically generate a colour palette from an image and this can be achieved using a machine-learning technique known as cluster analysis. PaletteGenerator also uses a k-means cluster algorithm to allow users to extract a colour palette from a source image[11]. Other digital resources have focussed on allowing users to share their colour palettes with others. For example, the colourlovers website currently incorporates more than 4.6 million palettes that have been generated and shared by users[12]. Colormind is a digital on-line tool that generates colour palettes using deep learning and which allows can automatically generate colour palettes from an upload digital image[13]. Colour Tool allows users to generate colour palettes and incorporates a visualisation tool to allow the user to see how the palette may appear in a particular graphical user interface[14]. Many of these websites have associated apps that can be used on smart devices. Typically, these tools generate colour palettes that consist of 5 colours; however, some allow a variable number of colours to be produced (for example. Palette Generator allows between 2 and 10 colours to be generated automatically from an image). The proliferation of these tools demonstrates a growing demand for tools that can help users to develop harmonious colour palettes quickly and efficiently.

Different colour palettes can communicate different colour meanings or induce different colour emotions.[15] Often colour palettes are generated by designers for practical application based on their knowledge of aesthetics with respect to either a design brief or the designer's own colour preferences.[16-17] Colour palettes may be extracted automatically from an image or a set of images[18] or even generated from a word.[19-20] There are currently no established methods for predicting the visual similarity or visual difference between two colour palettes; the automated design systems that are being developed based on data and machine learning will require the ability to be able to predict similarity between colour palettes.[21-22]

The prediction of colour differences between a pair of colours has of course been of paramount interest to the colour community for nearly a century.[23-25] The last 50 years has seen a number of colour difference formula being published; the majority of these are based on CIELAB colour space, including CIE94, CMC and most recently CIEDE2000 which is the current CIE recommendation for small colour differences.[26] CIELAB remains the CIE recommendation for use with large colour differences although there has recently been interest in predict large colour differences and in using colour difference metric based on colour-appearance spaces other than CIELAB such as CIECAM02.[27-28] Although evaluating colour differences between a pair of spatially homogenous colours has been widely explored there has been relatively little research about colour palette difference evaluation. In fact, it can be argued that the case for a pair of spatially homogenous colours is a special case of a more general problem. The more general problem is when we compare a palette of N colours with another palette of M colours). At some point, of course, this problem becomes the problem of image-difference metrics. A number of studies have explored the prediction of image difference. One method, for example, is to calculate the average colour difference on a pixel-by-pixel basis[29] but that assumes that there are corresponding pixels. This will often be the case for pairs of images (where there is, for example, an original image and a distorted version of it); however, this is most often not the case of pairs of colour palettes. More sophisticated algorithms have considered the spatial properties of the images[30] or strongly weight areas in the images that are deemed to be more significant or salient.[29] It is not clear, however, whether such methods could apply to the prediction of palette differences. However, it is clear that the colour difference formulae that have been developed and tested for use on solid colour patches should serve as the building blocks for image- or palette-difference metrics.

In one recent study, a psychophysical experiment was conducted with pairs of colour palettes, each containing 25 different colour patches.[22] In that study, three different palette-difference metrics were tested using the psychophysical data. The metrics tested were: (1) Single colour difference model (where the RGB values of each palette were averaged and a single colour difference was calculated between the palettes); (2) Mean colour difference model (where each patch in one palette was compared to each patch in the second palette and the mean of all of these colour differences was calculated); and (3) Minimum colour difference model (where each patch in a palette was compared with its closest colour in the other model and the mean of these colour differences was calculated). The coefficient of determination r^2 between each of these metrics and the psychophysical data was 0.35, 0.12 and 0.60 respectively. The Minimum Colour Difference Model (MICDM) was also the best performing of the three metrics using STRESS which has become preferred over the coefficient of determination in many colour-difference studies. All of the metrics tested were based on the CIELAB colour-difference formula. Although this finding suggests that MICDM might be useful for predicting visual differences between palettes, there are several questions that emerge. Firstly, how well does the model work when the number of colour in the palettes are much smaller or larger than the 25-colour palettes that were tested? Secondly, since optimised colour difference formulae such as CIEDE2000 often outperform CIELAB for solid colours, will this also be the case for the MICDM model? Therefore, in this study, a new psychophysical experiment was conducted to obtain the human visual colour differences (ΔV) between colour-palette pairs that each contain only 5 colours. Note earlier that many of the digital tools that allow users to generate and share colour palettes use 5-colour palettes and therefore it is important to see whether the early work (using 25-colour palettes) applies to these smaller palettes. The MICDM model will be evaluated but will also be extended by testing different colour-difference formulae to compute the colour difference (ΔE) between the palettes.

2. Experimental

A psychophysical experiment was conducted to quantify the visual difference between pairs of colour palettes that each contain 5 colours. A semantic differential scale was used as the method to collect the psychophysical data.

Participants

A total of 20 participants (9 males and 11 females) with normal colour vision according to the Ishihara test volunteered to take part in the psychophysical experiment. Their age ranged from 25 to 56 years. The purpose of the experiment was briefly explained to all participants when they were recruited.

Colour-palette pairs selection

A total of 180 colour palettes were obtained from a previous study.[31] Each of the palettes contained 5 colours that were chosen by participants in that study to represent landscape images. The palettes were displayed with the five colours in a horizontal row (Figure 1 illustrates 40 of the 5-colour palettes). There are 100!/(98!2!)=4950 possible pairs of these 180 colour palettes but to include all of these pairs would make the experiment too onerous for the participants. Therefore, 90 pairs of palettes were selected from the possible 4950 pairs to ensure that there was a range of differences from small to large (determined informally by one of the authors).

The metrics that are being considered to predict palette colour difference do not consider the order (or the arrangement) of the colours in the palettes. However, 8 of the pairs were selected so that the two palettes in each case contained the same colours but in a different order so that the effect of order on the psychophysical data could be explored (see Figure 2).

A further two pairs were selected where the colours and the order of the colours for each pair were identical. For the other 80 pairs of palettes the colours for each palette in the pair were different to each other.

Five of the 90 pairs were selected randomly and were duplicated in the experiment to assess intra-observer reliability so that in total each participant evaluated 95 pairs of palettes.

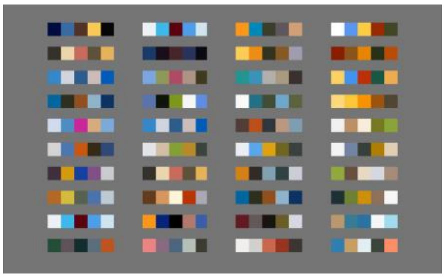

Figure 1: A representation of 40 of the 5-colour palettes used in the study.

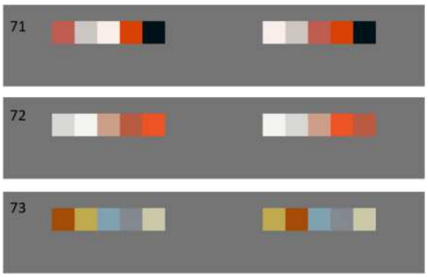

Figure 2: Three of the 8 pairs where the colours were the same but the order was changed.

Display

Colour-palette pairs were displayed on an LED computer monitor (HP DreamColor LP2480zx – a 24-inch LCD Backlit Monitor) and viewed, from a distance of about 1m, against a uniform grey background colour (CIELAB L*=50 approximately). The display of the pairs and the collection of psychophysical data was performed using a computer program written in MATLAB.

Experiment design

The experiment was conducted in a darkened room. Each participant was asked to sit in the dark environment for about 5 minutes to adapt to the environment. In the experiment, each participant was requested to view pairs of colour palettes and to indicate for each pair the degree of similarity or difference using a bi-polar scale with 10 points (see Figure 2) with one extreme (-5) representing most different and the other extreme (+5) representing most similar. At the beginning of the experiment, participants were shown all of the palettes being used in the experiment (Figure 1 shows one of three screens that the participants viewed) to give them an overall impression of the range. The 95 pairs were presented in a different random order for each participant. There was no time limit for each participant to finish the experiment, but the experiment overall took about 20-30 minutes for each participant. The results were automatically saved as the data of visual colour difference (ΔV) by the software. The results of the 5 duplicated pairs were excluded from the main results but were instead used to assess intra-participant repeatability. The ΔV values for each participant were later treated as interval data and were averaged to produce a ΔV value for each pair of palettes.

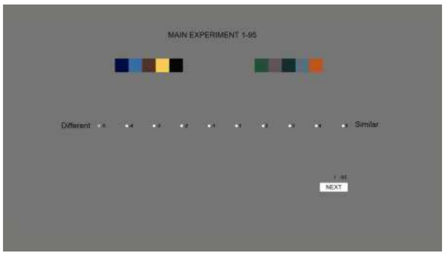

Figure 3: Screenshot of the experimental design.

Analysis

After the experiment the colours of all 95 pairs of colour palettes (in total 95×10=950 colours) were measured on the screen using a Konica Minolta CS-2000 spectroradiometer. Measuring the colours actually used on the screen in this instance is more accurate than using colour management to predict those colours (since what matters is not that we display specific colours but rather that we know the exact colorimetric specifications of the colours that were displayed). The measured spectral data were converted to CIELAB values using the display white point (CIE x = 0.3116, y = 0.3184). The CIELAB values were used for subsequent colour difference calculations. For each pair, the computed colour differences between two colour palettes were examined using one of two metrics that were also used in a previous study (Pan and Westland, 2018). The mean colour difference model MECDM (referred to as ΔE_M) is simply the average of the colour differences that result by comparing each patch in one palette with every other patch in the second palette. Since there are 5 colour patches in each palette this is the average 25 colour differences. The minimum colour difference model MICDP (referred to as ΔE_P) is more complex. The algorithm for calculating ΔE_P is according to the following 5 steps:

1. For each colour in one palette, the 5 colour differences between this colour and each of the colours in the second palette are calculated. The minimum of these colour differences is recorded.
2. Step 1 is repeated for all 5 colours in the first palette, for each finding their closest corresponding colours in the second palette, resulting in 25 colour 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 colours in the second palette. In other words, for each of these colors the closest corresponding colour 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 visual colour difference ΔE_P between the two palettes

Both ΔE_M and ΔE_P can be implemented using various colour difference formulae and in this study the following formulae were used: CIELAB, CMC(2;1), CMC(1;1), CIE94, CIEDE2000(1,1,1) and CIEDE2000(2,1,1). Subsequently, two measures were used to analyse the performance of different formulae. The first is regression analysis and the value of coefficient of determination r^2 was reported. The second is the Standardized Residual Sum of Squares (STRESS) measure[32] as shown in Equation 1.

$$STRESS = \sqrt{\frac{\sum(\Delta E_i - f\Delta V_i)^2}{\sum \Delta {E_i}^2}} \times 100\%$$

Eqn 1

where $f = \frac{\sum \Delta E_i \Delta V_i}{\sum \Delta {V_i}^2}$

3. Results

Figure 4 shows the distribution of mean rating scores for the 90 pairs of palettes. Note that mean ratings covered almost the entire range (from -4.55 to 4.85).

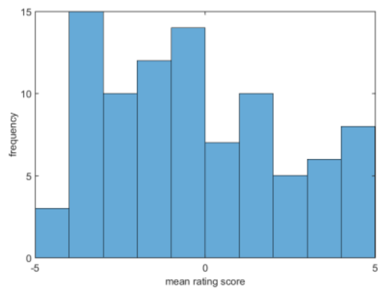

Figure 4: Frequency distribution of mean rating scores for the 90 pairs of palettes.

Figure 5 and 6 shows 10 of the palette pairs and their mean visual ratings. In Figure 5, the pairs have relatively large colour difference whereas in Figure 6 the pairs have much smaller colour difference. These figures are included to give readers an impression of the magnitude and type of colour differences that existed in the palette pairs in this study.

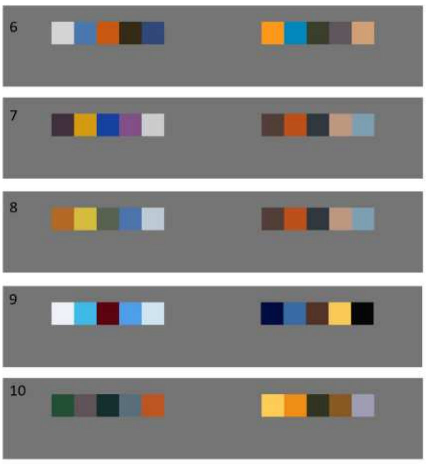

Figure 5: Pairs 6-10 which had mean visual ratings of -2.4, -3.7, -2.7, -3.45 and -3.2.

Of the 100 repeat assessments (20 participants × 5 palettes), 36% were exactly the same rating between the first and second attempt. A total of 73% of the repeat assessments were within 1 rating unit of the original assessment. When pooled over all observers the mean absolute difference between the first and second assessment was 0.45 units.

The two pairs of palettes that had exactly the same palettes being compared (the same colours in the same order) had average ratings of 4.85 and 4.55 which is close to the maximum rating of 5.0 (when the intra-participant variability of 0.45 is considered). There were 8 pairs that had the same colours but arranged in a different order and the average visual rating was 4.08. Given the intra-participant variability of 0.45 this does suggest that order or arrangement of colours in the palettes affects the visual differences that were reported.

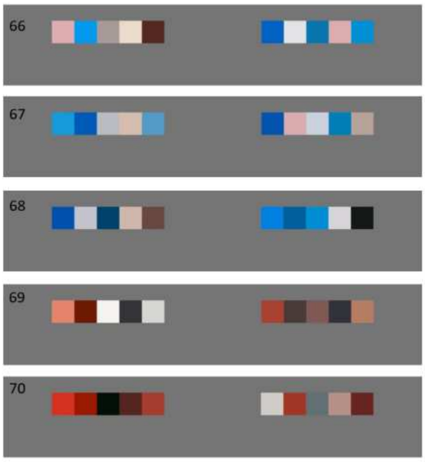

Figure 6: Pairs 66-70 which had mean visual ratings of 0.3, 3.9, -1.0, -1.45 and -1.45.

The models to predict visual difference between the pairs were implemented and the values of r2 and STRESS calculated between the model predictions and the visual data. Table 1 shows the r^2 and STRESS values for the two models (ΔE_M and ΔE_P) using each of the 6 colour-difference equations. Note that better performance is indicated by higher r^2 and lower STRESS values.

The order of performance for the two metrics (r^2 and STRESS) is very similar. It is evident that the minimum colour difference model MICDM performs better than the mean colour difference model (MECDM). However, according to both r^2 and STRESS the CIEDE2000 colour-difference equation (with a weighting of 2 of Lightness) gives the best performance when all six colour difference equations are considered.

Table 1: Calculation of r^2 and STRESS values for the 6 colour-difference equations and the two algorithms (MICDM and MECDM).

	r^2		STRESS	
	ΔE_M - MECDM	ΔE_P - MICDM	ΔE_M - MECDM	ΔE_P - MICDM
CIELAB	0.345	0.821	43.10	19.33
CMC(1,1)	0.213	0.768	41.87	22.99
CMC(2,1)	0.256	0.812	44.06	19.31
CIE94	0.281	0.768	45.54	22.69
CIE2000(1,1,1)	0.325	0.832	40.34	18.92
CIE2000(2,1,1)	0.392	0.864	39.29	16.93

4. Discussion

This work has confirmed earlier work[22] that the minimum colour difference model algorithm is able to make good predictions of the visual difference between colour palettes. In that earlier study the palettes contained 25 colours arranged in a 5 × 5 block and the r^2 and STRESS values obtained for the MICDM using CIELAB were 0.60 and 21.0 respectively. In this study, with palettes consisting of 5 colours in a horizontal row the r^2 and STRESS values obtained for the MICDM using CIELAB were 0.82 and 19.3 respectively. In other words, slightly better performance is reported in this work (with 5-colour palettes) than was reported in the earlier study with 25-colour palettes. We have unpublished data for a similar study with 45-colour palettes where the r^2 and STRESS values obtained for the MICDM using CIELAB were 0.38 and 33.51 respectively[33]. It appears that the MICDM method works better the fewer colours are in the colour palette. Most applications of colour palettes in design will have relatively few colours in them (typically less than 10) and therefore this work suggests that there can be practical applications of the MICDM algorithm. Note that when there is only one colour in each palette the well-known colour-difference problem that has been widely studied. The MICDM algorithm in this case simplifies to being the colour difference between the two colour patches. In one analysis of classical colour difference the STRESS values that have been reported for CIEDE2000 range from about 19-30.[32]

The STRESS value that is reported in this study for the MIDCM algorithm using the CIEDE2000 equation is 16.9 and this was obtained using a weighting of 2 for the Lightness parameter. This study found some evidence that the order or arrangement of the colours in the palette could affect the visual difference and this makes intuitive sense. However, we suggest that the order effect may be relatively unimportant for two reasons. Firstly, the visual colour differences reported in this study for the samples where the order was changed (but the colours in the palettes remained the same) was 4.08 on a scale of -5 to +5 (where 5 indicates maximum similarity). Secondly, the performance of the MICDM algorithm is high ($r^2 = 0.86$ and STRESS = 16.3) even though this algorithm does not consider the order of the colours in the palettes. Nevertheless, future work may well improve upon the performance of the MICDM algorithm if the order of colours in the palettes can be included in the algorithm.

The likely application of this work is in digital workflows where colour palettes are automatically generated. For example, colour palettes have been generated for words or sentences based on semantic knowledge[14] and from automated internet search methods.[21] Other potential applications are in the comparison of colour palettes that are selected by different groups defined by, for example, culture, age or gender.[34] The method could be used to compare palettes that are generated from various websites (such as Adobe Colour and Colormind) and could help to answer questions about how similar the outputs from these different website are. For many applications there is no ‘gold standard’; that is, different methods may generate different palettes (given the same input), each of which are equally valid (perhaps as inspiration for, or application to, a design process). However, in other applications there will be a ’correct’ or target colour palette (usually derived psychophysically) and in those cases the metric developed in this work could be used to measure how closely automatically generated colour palettes match the target.

5. References

1. Pereira AC and Romero F (2017). A review of the meanings and the implications of the Industry 4.0 concept, *Procedia Manufacturing*, **13**, 1206-1214.
2. Yu Y, Choi TM and Hui CL (2011). An intelligent fast sales forecasting model for fashion products, *Expert Systems with Applications*, **38** (6), 7373-7379.

3. Loureiro AL, Miguéis VL and da Silva LF (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail, *Decision Support Systems*, **114**, 81-93.
4. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B and Bharath AA (2018). Generative adversarial networks: An overview, *IEEE Signal Processing Magazine*, **35** (1), 53-65.
5. Chesney R and Citron D (2019). Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics, *Foreign Aff.*, **98**, 147.
6. Sbai O, Elhoseiny M, Bordes A, LeCun Y and Couprie C (2018). Design: Design inspiration from generative networks, in *Proceedings of the European Conference on Computer Vision* (ECCV) (pp. 0-0).
7. Anderson D, Perez KB, Xuereb Z, Otto K and Wood K (2018). Design processes of design automation practitioners, *Proceedings of the ASME Design Engineering Technical Conference*, Volume 7.
8. Gu Z and Lou J (2016), Data driven webpage color design, *Computer-Aided Design*, **77**, 46-59.
9. Adobe Color: https://color.adobe.com/create (last accessed 7th January 2020).
10. Westland S, Laycock K, Cheung V, Henry PM and Mahyar F (2007). Colour Harmony, *Colour: Design & Creativity*, 1 (1), 1-15.
11. PaletteGenerator: https://palettegenerator.com/ (last accessed 7th January 2020).
12. Colourlovers: https://www.colourlovers.com/palettes (last accessed 7th January 2020).
13. Colormind: http://colormind.io/ (last accessed 7th January 2020).
14. Colour Tool: https://material.io/resources/color/ (last accessed 7th January 2020).
15. Bartram L, Patra A and Stone M (2017). Affective color in visualization, in *Proceedings of the 2017 CHI conference on human factors in computing systems*, 1364-1374. ACM.
16. Moretti G and Lyons P (2002). Tools for the selection of colour palettes, In *Proceedings of the SIGCHI-NZ Symposium on Computer-Human Interaction*, 13-18. ACM.
17. Luo MR (2006). Applying colour science in colour design, *Optics & laser technology*, **38** (4-6), 392-398.
18. Morse BS, Thornton D, Xia Q and Uibel J (2007). Image-based color schemes, in *IEEE International Conference on Image Processing*, 3, *III-497*. IEEE.
19. Jahanian A, Vishwanathan SVN and Allebach JP (2015). March. Autonomous color theme extraction from images using saliency, in *Imaging and Multimedia Analytics in a Web and Mobile World 2015* (**9408**, 940807). International Society for Optics and Photonics.

20. Havasi C, Speer R and Holmgren J (2010). Automated color selection using semantic knowledge. In *AAAI Fall Symposium Series*, Nov 3.
21. Lindner A and Süsstrunk S (2013). Automatic color palette creation from words, in *Color and Imaging Conference*, 69-74. Society for Imaging Science and Technology.
22. Pan Q and Westland S (2018). Comparative Evaluation of Color Differences between Color Palettes, Proceedings of the IS&T *Color and Imaging Conference*, 110–115.
23. McDonald R (1990). European practices and philosophy in industrial colour-difference evaluation, *Color Research & Application*, **15** (5), 249-260.
24. Witt K (1995). CIE guidelines for coordinated future work on industrial colour-difference evaluation, *Color Research & Application*, **20** (6), 399–403.
25. Melgosa M (2000). Testing CIELAB-based color-difference formulas, *Color Research & Application*, **25** (1), 49-55.
26. Luo MR, Cui G and Rigg B (2001). The development of the CIE 2000 colour-difference formula: CIEDE2000, *Color Research & Application*, **26** (5), 340-350.
27. Kuehni RG (2009). Variability in estimation of supra-threshold small color differences, *Color Research and Application*, **34** (5), 367-374.
28. Wang H, Cui G, Luo MR and Xu H (2012). Evaluation of colour-difference formulae for different colour-difference magnitudes, *Color Research & Application*, **37** (5), 316–325.
29. Hong G and Luo MR (2006). New algorithm for calculating perceived colour difference of images, *The Imaging Science Journal*, **54** (2), 86-91.
30. Zhang X and Wandell BA (1997). A spatial extension of CIELAB for digital color-image reproduction, *Journal of the Society for Information Display*, **5** (1), 61-63.
31. Yang J, Westland S and Xiao K (2017). Developing a Method for Generating Colour Palette from Landscape Images, *Proceedings of the 13th Congress of the International Color Association*, OS03-1, Jeju, Korea.
32. Melgosa M, Huertas R and Berns RS (2008). Performance of recent advanced color-difference formulas using the standardized residual sum of squares index, *JOSA A.*, **25** (7), 1828-1834.
33. Chen Y (2019). *Private Communication*.
34. Chen Y, Yang J, Pan Q, Vazirian M and Westland S (2019). A method for exploring word-colour associations, *Color Research & Application*, (under review).