Experimental Filters for Estimating Image Differences

Zhaohui Wang and M. Ronnier Luo; Department of Colour Science, University of Leeds; Leeds, UK

Abstract

Evaluation of image difference including colour and spatial aspects is often required in colour reproduction industry. The objective of a successful image difference model should have a good agreement with the image difference perceived by observers. The perceived image difference is affected by many factors, such as image characteristics, human vision system, rendering methods and viewing conditions. The current conventional colour difference formulae can not well predict the perceived difference. In this paper, a new experimental method was designed to investigate how image difference is identified by human observers.

Introduction

In imaging industry, colour difference formulae are typically used on pixel-by-pixel calculation for evaluating differences between two images. The conventional colour difference formulae, such as CIELAB [1], CMC [2], CIE94 [3], CIEDE2000 [4] have been successfully applied to estimate colour difference between large size uniform colour patches. However, when dealing with the image based colour difference computation, these formulae could introduce wrong impressions.

Although conventional colour difference formulae may not be directly applied to complex images, its spatial extension introduced by Zhang et al. [5] made some significant advances for the spatially altered images. This extension was based on the human contrast sensitivity function (CSF) which varies as a function of spatial pattern [6, 7]. Since then researches [8-12] have been focused on the optimisation of CSF functions in the evaluation of image difference. The application of spatial filter matched the criteria of which the conventional colour difference formula was defined as the details or the high frequency part of the image was filtered by the low pass filter. However, the successful application is only tested using the compressed or half-toned images. Hong and Luo [13] investigated the image areas of significance and assigned higher weights to which were perceptually important. They proposed that both larger area of the similar colour and larger colour difference between the edges of two objects should be weighted higher.

This paper introduces a new experimental method intended to reveal how image difference was evaluated by observers directly. The colour differences between some larger parts of an image are identified. The novel aspect of this experiment is that software was used to locate the objects in an image based on which the observers judged the image difference. Note that the term "object" used in this study including the objects and areas in the image identified by observers.

Experimental Method

Eight images were selected for the experiment. They cover memory colours (such as grass, sky, fruit, and skin tone), complex image scenes, artificial objects, and uniform backgrounds. They were purposely developed to test colour image reproduction systems and coded in Bitmap format. Six image rendering algorithms (the modification of lightness, chroma, lightness & chroma, compression rate, noise amount, and blur or sharpen radius) were applied at different levels. In total, 280 testing image pairs were used in the experiment, which include 179 reproduced image pairs and 101 repeated image pairs for accuracy control.

The experiment was conducted in a darkened room using a CRT monitor having a white point corresponding to the chromaticity of D65 illuminant. Experimental software was developed as shown in Figure 1. For each assessment, observers saw an original and a test image. Each image had a white border against a grey background with an L* value of 50. The positions of each test and original images and the sequences of the pairs displayed randomly to compensate any uniformity of the CRT.

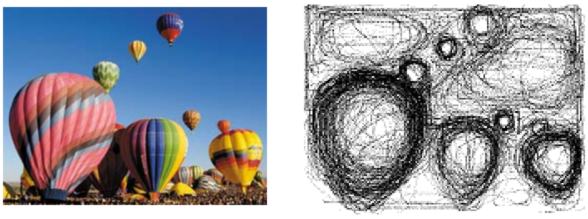
***Figure 1** Experimental interface (The right image shows some objects, in black frame, selected for judging image difference)*

The grey scale method was applied to scale the perceived difference of an image pair. This method was originally used for evaluating colour fastness for assessing change in colour or staining for uniform colour patches. The grey scale consists of a range of neutral samples. Observers select a particular grey sample using the scroll bar located in the bottom of Figure 1 together with a fixed grey standard to form an image pair which shows a colour difference close to that of image pair.

Ten normal colour vision observers were employed according to Ishihara test. Each observer had two tasks: to assess the total image difference and to mark the object or objects which were used for assessing the image difference. Software was also developed to allow observers to select region(s), or object(s), used for judging image difference.

Construction of Experimental Filter

The visual results in terms of grey scale grades were converted to visual difference in terms of CIELAB. As mentioned earlier, the identified object(s) for each image were also collected. Each object judged by the number of observers was added for each attribute of manipulation (e.g. lightness, sharpness). Figure 2 illustrates the process as Figure 2(b) shows the objects selected by all the observers for the 'Balloon' image (Figure 2(a)). The results were used to construct a filter depending on the weight.



(a) Balloon *(b) Objects drawn by observers*
Figure 2 *Demonstration of visual results*

Although the appearance of objects drawn by observers has low legibility, however, the patterns of the objects selected are clear and can be refined according to the image content as shown in Figure 3. For the 'Balloon' image, it was divided into eight objects. These objects include the balloons of different sizes and sky background. The remaining areas of the image were discarded and not be used in the further processing. For the other images, the individual objects were separated using the same method.

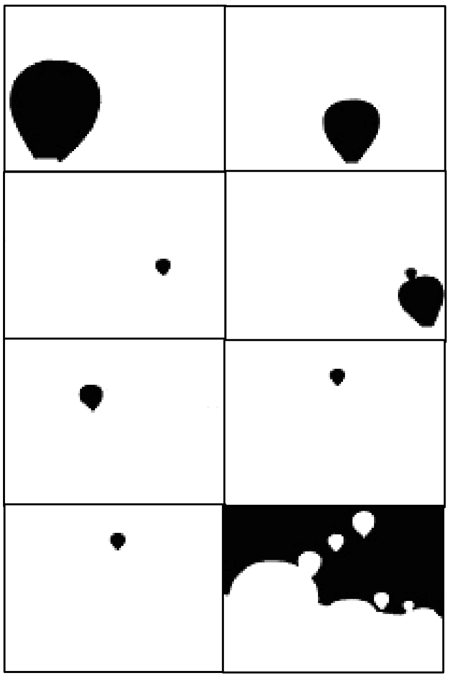


Figure 3 *Objects template of the 'Balloon' image*

After the image analysis, a filter was established for the 'Balloon' image based on the number of counts. Different objects have different weights representing the number of times the object had attracted the observers when the differences were judged between image pairs. The results are summarised in Figure 4. It can be seen that large objects are more important when observers identified image difference. The numbers marked on each object show how many times and percentages observers had focused their attentions on the corresponding objects. The sky background and three larger balloons received much more attentions than the other objects.

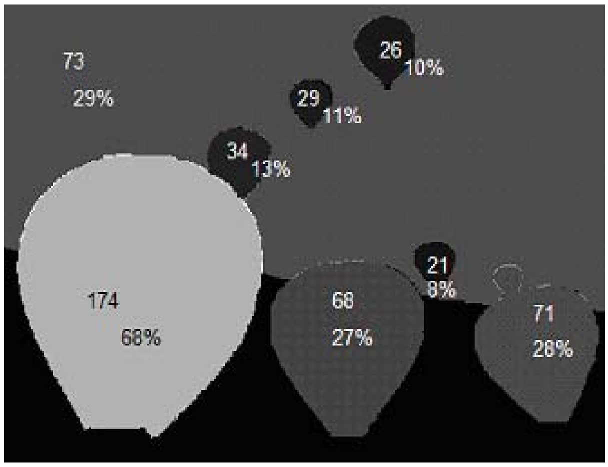


Figure 4 *Experimental filters for the 'Balloon' image based on visual judgements*

The same method of data analysis was applied to other seven original images. Similar results showed that observers tended to focus on certain objects, usually objects of lager size and gave their judgements mainly based on the difference of these objects. This is because the colour differences in the large area are more perceptibly noticeable. For the small objects, especially for the areas which were ignored by observers, the results proved that human eyes tend to be more tolerant towards colour difference of smaller image areas. When a colour difference formula was used to calculate image difference, one general method is to simply average the colour difference in an image. Figure 4(b) shows that the overall perceived image difference could not be treated as a simple average of pixel by pixel.

Figure 5 shows the individual filters constructed for images having the same manipulation function. The numbers in the bracket are the total number of judgements for each image manipulation. The results confirm that observers judged image difference based on large objects is independent of the image manipulation method. The largest balloon had attracted much more attention than any other objects of the image. However, for noise manipulation, the result is different, in which the sky area received twice attentions than the largest balloon did. By comparing with the results of other manipulation methods, it can be concluded that the detection ability of human visual system is higher on a plain background for the noisiness manipulation images. This is not only limited in this manipulation, but also in compression manipulation. It is also found in Figure 5 that the sky area got a higher number of attentions in compression manipulation. By compression algorithms, the uniformity defects were introduced which resulted in a significant appearance on a plain background. For most rendering method, the smaller objects (e.g. marked as 0 in chroma manipulation) played a trivial role in the image difference judgement. In sharpness manipulation, the appearance of details was enhanced and resulted in the smaller objects attracted much more attentions than that in the other manipulation algorithms.

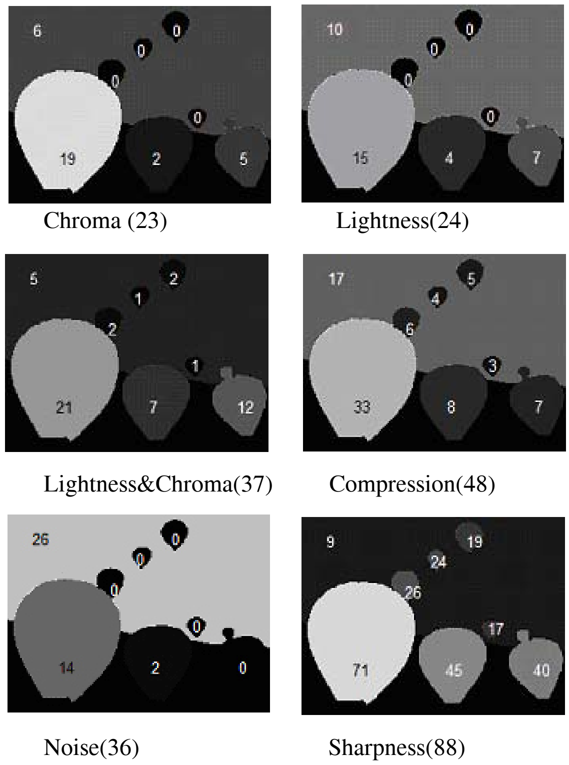


Figure 5 *Experimental filters for each image manipulation method*

Discussion

Observer Accuracy

In any psychophysical experiment, it is important to know the reliability of the visual results which is often represented by observer accuracy. The observer accuracy investigated in this study includes intra-observer (repeatability) and inter-observer variation.

In the experiment, a number of 101 repeated image pairs were used to test the observer's repeatability. Thus, for each observer, the two visual judgements on the same stimuli were used to calculate observer's repeatability in terms of Coefficient of Variation (CV). The results are summarised in Table 1 for each of the 10 observers. The typical performance is about 30% variation.

Observer	1	2	3	4	5	Mean
CV	26	32	25	42	30	
Observer	6	7	8	9	10	29
CV	28	30	25	23	27	

Table 1 *Observer Repeatability Performance*

For inter-observer's variation, the mean visual results are taken as the standard values. Again, the CV was calculated between each observer's and mean results. The results are listed in Table 2. It can be seen that the inter-observer variation was slightly higher than that observer repeatability, i.e. 36 against 29 CV units respectively.

Observer	1	2	3	4	5	Mean
CV	43	54	28	36	46	
Observer	6	7	8	9	10	36
CV	33	24	30	34	32	

Table 2 *Inter-Observer's Variation*

Effect of weighting filter

The CV value calculated between colour difference and visual results was used for evaluating the performance of image difference metrics. Images were first weighted by the experimental filter in CIELAB colour specifications, L*, a*, and b*, and CIELAB colour difference was then calculated pixel by pixel. The median was used to represent image difference. This procedure has the same effect as the weighting coefficients applied on the pixel by pixel colour difference. Finally, CV value was calculated between the original and weighted CIELAB colour differences, and visual results. The CV values of each of the 8 images and all images are listed in Table 3. It is somewhat disappointing that the two methods gave quite similar performances. This implies that the weight filter does not improve the CV values much.

Images	CV	
	CIELAB	Weighted
1	58	58
2	59	50
3	58	63
4	65	60
5	51	37
6	63	56
7	57	63
8	39	39
Average	56	53

Table 3 *Comparison of CV values between normal CIELAB and weighted CIELAB*

The experimental filters in Figure 4 show that main objects have a lager effect on the determination of image difference. To investigate the effect of the main objects, the numbers of objects were involved to calculate image difference in terms of CIELAB formula.

According to the percentage of the attention that objects obtained in each image, it was divided into two levels. The first level included the main object which received more than 50% attention in each image. This represents only one main object at this level for each image. The second level is 20% level which involved objects received the attentions more than 20%. This level covered most objects of image. The colour difference was calculated by the same procedure as before. The results are given in Table 4.

Obviously, the average and overall performance of main objects is quite similar with the performance of CIELAB, which means the colour difference between image pairs can be represented by the difference between main objects in the image. However, for each individual image, the performance of the weighted main objects is not consistent, as in some cases, it improved the performance but worse for some others. The results again show that all three methods gave very similar formula performance, i.e. the filter method does not improve the formula performance.

Image	CV (CIELAB)	CV(weighted)	
		>50%	>20%
1	58	41	58
2	59	57	58
3	58	48	55
4	65	79	72
5	51	59	49
6	63	65	62
7	57	63	63
8	39	49	56
Average	*56*	*58*	*59*
Overall	*62*	*63*	*65*

***Table 4** Effects of main objects in image difference evaluation*

Effect of manipulation methods

The same performance testing method was used to each of the six manipulation methods. Figure 5 shows that different manipulation methods resulted in different weighting filters. For investigating the performance of weighting filters according to the rendering method, two levels, 50%, and 20% were used again. The results are summarized in Table 5 in terms of CV values.

Manipulation Methods	CV (CIELAB)	CV (weighted)		
		>50%	*20%*	*all*
Chroma (L)	62	24	51	80
Lightness (C)	32	30	37	67
L&C	35	38	48	60
Compression	38	43	31	71
Noise	26	25	23	53
Sharpness	47	47	37	66
Overall	*62*	*63*	*65*	*71*

***Table 5** Comparison of performance by manipulation methods*

It can be found that the CV value was improved much by main object (50% level) weighting filter under the Chroma rendering method. For the other methods, 50% and 20% levels gave a similar performance. From Table 5, it is clear that the rendering methods affect very little to the final results.

Conclusion

An experiment was conducted for assessing the image difference between image pairs. Different from the other research, this study focused on investigation of how image difference was determined by human vision system according to spatial aspect.

It was found that observers assessed the image difference based on the main objects in the image. According to the visual judgements, weighting filters were developed in which the main objects weighted higher and the trivial objects weighted lower or zero. However, the evaluation of image difference by weighting filter does not improve the formula performance. This implies that the image difference may not be represented by the weighted sum of colour differences of individual objects.

However, some limitations in the experiment may affect the results in this study. First of all, the ranges of image differences from different manipulation methods are quite different. This could leads to quite different filters and results. Second, the method used to identify observers' attention for judging image difference was not accurate. All above could be further improved in the future study. An eye tracking device could be useful in the future study.

References

[1] CIE, Colorimetry, CIE Publ. 15:2004, Central Bureau of the CIE, Vienna, (2004).
[2] F. J. J., Clarke, R. McDonald, B. Rigg., "Modification to the JPC79 colour-difference formula", J. Soc. Dyers Colour, 100, pg.128-132 and 281-282, (1984).
[3] CIE, Technical Report, Industrial Colour-Difference Evaluation, CIE Publ. 116-1995, Central Bureau of the CIE, Vienna, (1995).
[4] M. R. Luo, G. Cui and B. Rigg., "The development of the CIE 2000 Colour-Difference Formula: CIEDE2000", Color Res. Appl., 26, 5, pg. 340-350, (2001).
[5] X. Zhang, and B. A. Wandell, "A Spatial Extension of CIELAB for digital Color-image Reproduction", Journal of the Society for Information Display, 5, 1, pg. 61-63, (1997).
[6] S. Anderson, K. Mullen, R. Hess, "Human peripheral spatial resolution for achromatic and chromatic stimuli: Limits imposed by optical and retinal factors", J. Physiol., 442, pg. 47-64, (1991).
[7] A. B. Poirson, B. A. Wandell, "The appearance of colored patterns: Pattern-color separability", J. Opt. Soc. Am. A, 10, 12, pg. 2458-2470, (1993).
[8] G. M. Johnson, and M. D. Fairchild, "A Top Down Description of S-CIELAB and CIEDE2000", Color Res. Appl., 28, 6, pg.425-435, (2003).
[9] M. D. Fairchild and G. M. Johnson, "The iCAM framework for image appearance, image difference, and image quality", J. of Electronic Imaging, 13, pg. 126-138, (2004).
[10] G. M. Johnson and M. D. Fairchild, Drawinsim of Color Image Difference Models, Proc. IS&T/SID 9th CIC, pg.108-112, (2001).
[11] G. M. Johnson and M. D. Fairchild, On Contrast Sensitivity in an Image Difference Model, proc. IS&T PICS, pg. 18-23, (2002).
[12] J. Morovič and P. L. Sun, "Predicting Image Differences in Colour Reproduction from Their Colorimetric Correlates", J. Imaging Sci. and Technol., 47, 6, pg. 509-516, (2003).
[13] G. Hong,l M. R. Luo and P. A. Rhodes, "A new algorithm for calculating perceived colour difference of images", J. Imaging Sci. and Technol., 54, 2, pg. 86-91, (2006).

Author Biography

Zhaohui Wang received his MSc. Degree in Imaging Science from the University of Derby (2004). He is currently a research student at the Department of Colour Science in the University of Leeds. He is working on the field of image categorization, scene referred profile, and image difference evaluation.