Visual Differences In Colour Reproduction And Their Colorimetric Correlates

Ján Morovic and Pei–Li Sun
Colour & Imaging Institute, University of Derby
Derby, United Kingdom

Abstract

This paper first presents a summary of a psychophysical experiment in which observers made judgements about the types of differences they perceived between originals and reproductions in a cross–media colour image reproduction. The results of the observer–reported visual data from that experiment are then compared with analogous metrics extracted from colorimetric data of the corresponding originals and reproductions. While there is a good agreement in terms of the most general findings, looking at more detailed results shows significant differences between visual and colorimetrically–based data. The paper then proceeds to describe a colorimetrically based metric that takes into account some aspects of the visual system and using information both about the statistics of colour differences, of the original images and of changes to spatial characteristics is able to give a close prediction of observer responses. The final metric is proposed for further testing as a means of predicting observer responses of image difference in colour reproduction as well as in other applications.

Introduction

In cross–media colour image reproduction rendering originals on other than their native imaging media almost invariably introduces changes to their appearance. These changes can, amongst others, be in terms of the image's spatial detail (e.g., when a reflective original is rendered on a display), in terms of the image's colours (e.g. when the reproduction medium has a more limited gamut than the original image) or in terms of the image's contrast. Understanding what these changes are is then the first step in being able to improve the performance of the image reproduction system that resulted in them.

To this end a study was conducted to find out what differences observers see in cross–media image reproduction experiments and the results of those observer responses have been reported previously.[1] While knowing what kinds of differences were noted by observers is an important first step in being able to improve an image reproduction system, it is also necessary to know how observer judgements can be predicted from colorimetric data about the original and its reproductions. That is because it is this data from which parameters can potentially be extracted for an improved system.

The aim of the present paper is therefore to analyse metrics derived from the colorimetric image data available for the originals and their reproductions that were evaluated in the previous study and to determine what relationship there is between them and the observers' visual results. If a strong relationship is found then observer judgements could also be predicted for other images and image reproduction systems and the behaviour of the latter could be adjusted accordingly.

The following sections will first describe the experimental setup in which the present data was obtained, then summarise the results of the visual differences previously reported by observers and finally extract analogous metrics from colorimetric data available for the experiment. The relationship between visual and colorimetric analyses of the same cross–media reproduction system's performance will then be made and an improvement of their relationship will be attempted.

Psychophysical Experiment

The visual differences that will be modelled here were obtained in a psychophysical experiment where observers were shown a series of eight test images (Fig. 1) on a CRT display alongside pairs of their printed reproductions.

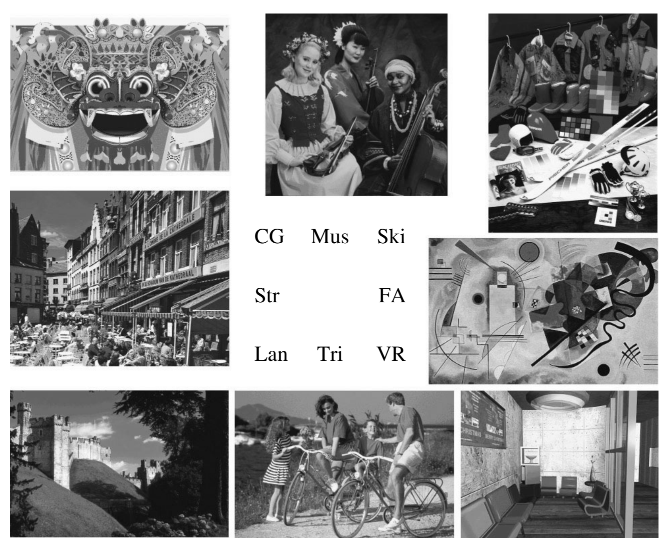


Figure 1. Test images used in psychophysical experiment.

The reproductions were made using four gamut mapping algorithms (CARISMA,[2] GCUSP,[2] SKNEE[3] and WCLIP[1]) and observers were first asked to list, in their own words, all the differences they could see between an original and its reproduction. Here observers reported differences like changes in lightness (L), colourfulness (C) and hue (H) as well as differences in detail and contrast. Observers also reported that some reproductions were pale, faded or blurred when compared with the corresponding original and all the judgements were made either for an entire reproduction or only for some part of it.

Next observers were asked to judge the importance of each of the differences they reported and finally a category judgement experiment was performed in which the accuracy of each of the reproductions was evaluated.

The raw experimental results for four reproductions of each of the eight originals were then processed to unify the terminology used by different observers, to balance the contribution made by different observers and to convert the image–relative judgements to a single scale. Details of how the experiment was set up, conducted and analysed can be found elsewhere.[1]

Summary of Visual Results

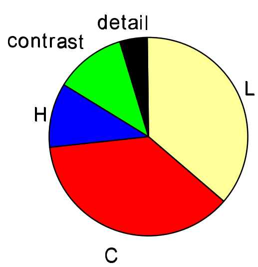


Figure 2. Relative importance of types of visual observer–reported differences.

The overall results of the above experiment are shown in Fig. 2 and it can be clearly seen from them that over 80% of differences were reported to be due to differences in colour attributes between originals and reproductions and only less than 20% were due to differences in contrast or spatial detail. Furthermore the ratio of L:C:H is 1.0:1.0:0.3 – in other words lightness and chroma changed to a similar extent and hue only had less than a third of that change. Note also that these overall results agree very well with two previous studies[4,5] that looked at the relative importance of various types of differences. Even though they using very different approaches, they also found that colour differences are responsible for a very large proportion of image differences in colour reproduction.

The proportion of these differences was also looked at on a GMA–by–GMA basis and those results are shown in Fig. 3. It can there be seen what kinds of differences were reported for the reproductions made with different algorithms. For example, the figure shows that CARISMA caused greater lightness and hue differences than the other algorithms but did well in terns of chroma. SKNEE on the other hand did well for most attributes except for detail.

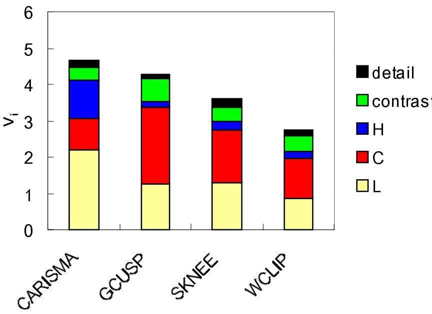


Figure 3. Visual differences for individual GMAs.

Finally, the results also showed where in the images observers perceived greatest differences (Fig. 4) and here it can be seen that this was the case for areas of sky as well as light–green objects.

Figure 4. Local visual differences (the lighter the shade, the greater a difference it represents).

Comparing Visual and Colorimetric Data

Given the above findings of the psychophysical experiment, one can look at analogous metrics extracted from the colorimetric data of the originals and reproductions and see how the two relate to each other.

This, however, presents the first difficulty in that a whole range of metrics can be extracted for each of the types of differences reported by observers. Lightness can be extracted in terms of CIELAB $L*$,[6] CIECAM97s J,[7] RLAB L^R,[8] etc. and chroma and hue present an analogous situation. Turning to the other two types of differences the problem becomes even more complex whereby contrast can be considered both globally and locally[9] and metrics ranging from the standard deviation of lightness values to statistics of spatially filtered versions of the reproductions can be used. Detail can also be quantified in terms of many metrics including the reproductions' power spectra and statistics of high–pass filtered versions of the images. The challenge then is to find such colorimetrically based metrics that relate most closely to the visual differences reported by observers.

As there are many metrics that can be extracted from the basic colorimetric data in an attempt to match either overall visual observer responses or some aspect of them, it is not practical to look at all combinations of their combinations and the following analysis is only an exploration of a subset of the many possible approaches.

Summary of Colorimetric Data

To give an initial view of the relationship between visual observer–reported differences and the properties of colorimetric data of original and reproduced images, analogous summaries of the colorimetric data to those extracted from the visual results will be computed first. Figure 5 therefore shows the relative magnitude of lightness, chroma, hue, contrast and detail metrics obtained from the colorimetric data and Figure 6 does so individually for the four algorithms. Lightness, chroma and hue differences were here predicted using CAM97s2.[17] Contrast difference was predicted using the difference between the standard deviations of original and reproduction image lightnesses. Detail difference was modelled using the difference between mean values of high–pass filtered versions of original and reproduction lightness images.

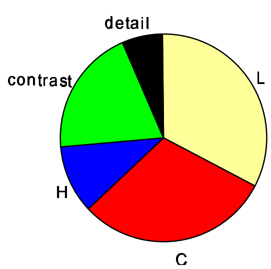


Figure 5. Relative magnitude of types of colorimetrically–based differences.

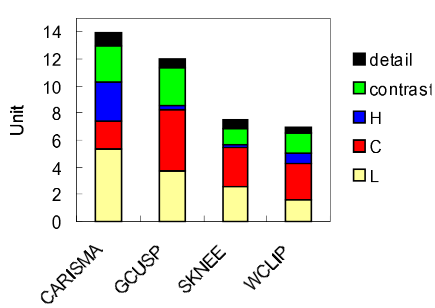


Figure 6. Colorimetrically–based differences for individual GMAs.

In the above figures it can be seen that in general the colorimetric data does show a similar picture to the visual results – the ratio of colour differences to contrast and detail differences is similar to that for the visual results and so is L:C:H. The results for individual GMAs, however, begin to exhibit greater differences. This is so most notably for contrast and detail but also for lightness, chroma and hue.

Finally, Figure 7 contains images showing pixel–by–pixel CAM97s2 Euclidean distances (ΔE_{97s}) between originals and their corresponding reproductions and while there are clear differences with the visual results, there are also significant similarities.

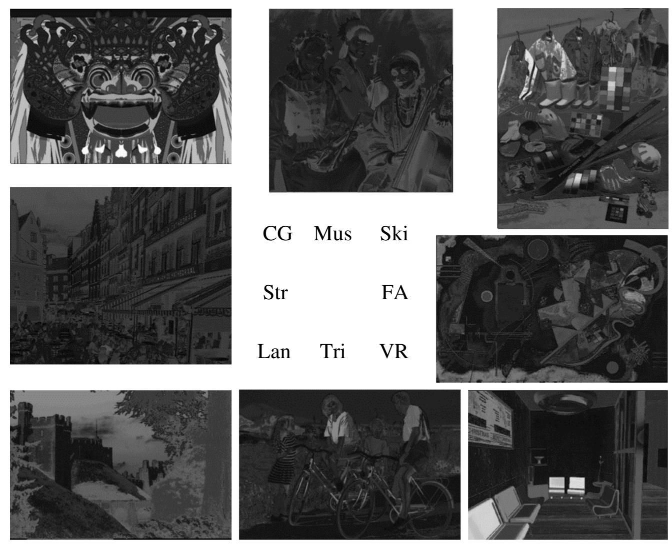


Figure 7. Local ΔE97s differences (black represents zero and white represents 35.2)

Overall it can be seen that the most general results of both visual and colorimetric analyses (i.e. the relative magnitude of difference types) agree very well but that differences between the two become more and more pronounced as greater levels of detail are taken into account. This can also be seen by looking at the values of determination coefficients (R^2) between visual differences and corresponding colorimetrically–based metrics (Table 1) which show very low levels of correlation between the two.

Seeing such a weak relationship between colorimetry and observer results should not come as a surprise, given that the former is fundamentally based on data obtained for uniformly coloured patches seen against uniform backgrounds whereas the latter are judgements made about properties of complex images. To improve the relationship between colorimetrically–based data and even just visual results that are in terms of colour attributes it will be necessary to take into account a number of factors beyond just the colour attributes of image pixels. The aim of the remainder of this paper will therefore be to find such a combination of colorimetrically–based metrics that correlate better with the visual results and that can be better used for predicting them.

Table 1. R^2 Values Between Visual and Colorimetric Differences.

Visual	L	C	H	LCH
Colorimetric	$\|\Delta J\|$	$\|\Delta C\|$	$\|\Delta H\|$	ΔE
R^2	0.393	0.480	0.555	0.378

Improving Correlation Between Colorimetric and Visual Results

As it is the colour differences that were found to be by far the most important ones both in the present study and in previous work,[4,5] the focus here will be to predict only this part of the visual results and to leave contrast and detail for future work. Table 1 already showed that R^2 was only 0.378 between the combined LCH observer–reported visual differences and the mean ΔE_{97s} values for the 32 reproductions used here (Fig. 8).

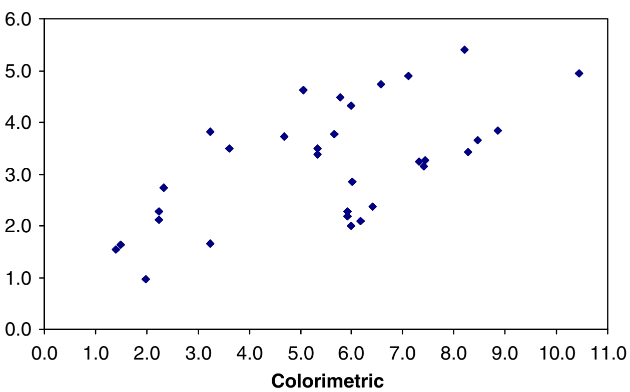


Figure 8. Visual versus mean ΔE_{97s} values for 32 reproductions.

99^{th} Percentile

A first step for improving this correlation is to look at the 99^{th} percentile of ΔE_{97s} distributions as high percentiles have previously been reported to better correlate with perceived differences between complex images.[10] In this study too the correlation between the visual LCH data and the 99^{th} percentile of ΔE_{97s} values is higher than for the mean and has R^2=0.438.

CSF Filter

Second, it has previously been proposed that differences between images should take into account the human visual system's contrast sensitivity by filtering luminance and chrominance channels to reflect higher sensitivity to the former. Therefore the filters proposed in the SCIELAB[11] model were applied to the present colorimetric data with an improved correlation between the 99^{th} percentile of ΔE_{97s} values between spatially filtered images of R^2=0.469.

Weighted ΔE

Third, as weighted ΔE equations have been shown to give better results in colour reproduction than unweighted equations,[12,13] the 99^{th} percentile of weighted colour differences (ΔE_{97sWT}) between filtered images was used. ΔE_{97sWT} had weights that divide ΔJ, ΔC and ΔH in a ratio of 1:2:1, giving chroma half the weight given to lightness and hue. This further improved correlation with the visual data to R^2=0.584.

Proportion of Unacceptable Differences

Fourth, to take into account the proportion of reproduction image pixels that had large differences from the original, the percentage of pixels with colour differences smaller than 6 ΔE units was used as a weight to the results of step 3. The reason for using 6 ΔE as the threshold is that this is close to what has been suggested as an acceptability threshold of pixel–by–pixel differences between complex images.[10,14] As such this weight takes into account the proportion of pixels that had unacceptable differences and doing so increased correlation to R^2=0.604.

Lightness Differences

Fifth, taking into account the distribution of lightness differences by including the median and standard deviation of absolute lightness differences between original and reproduction further improved correlation to R^2=0.679. This suggests that visual judgements are influenced both by the magnitude of lightness changes as well as to the variation of these changes.

Lightness and Chroma of Originals

Sixth, in addition to looking at the differences between originals and reproductions, it is also beneficial to incorporate factors determined solely by the originals. To this end the mean lightness and mean chroma values of originals were taken into account whereby differences in reproductions of originals that had more dark colours were given more weight. This was done as such images are subject to most noticeable change since gamut differences are typically greatest at lower lightnesses where original and reproduction gamuts differ most significantly. The mean chroma of the original, on the other hand, has negative correlation with the relationship of visual and colorimetric results as observers seem to be more sensitive to changes of images with lower chromas than to highly chromatic images. Taking mean lightnesses and chromas of originals into account in addition to the parameters mentioned in the previous steps resulted in an improved correlation of R^2=0.696.

Spatial Detail

Seventh, looking at how spatial detail changed between original and reproduction and giving more weight to images where this difference was great further improved correlation to R^2=0.707. Differences in spatial detail were quantified by taking high–pass filtered versions of originals and reproductions and looking at the difference between the mean values of these. Taking this factor into account using a Sobel filter[15] has also previously been found to enhance colour image difference prediction.[16]

Final Colorimetrically Based Metric

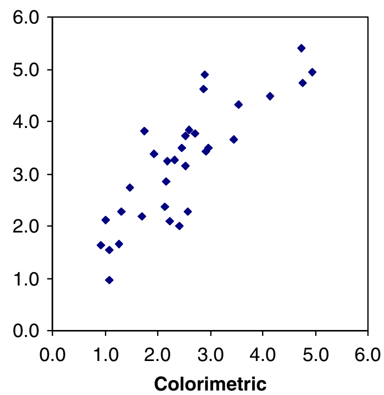


Figure 9. Visual versus final colorimetrically–based values for 32 reproductions.

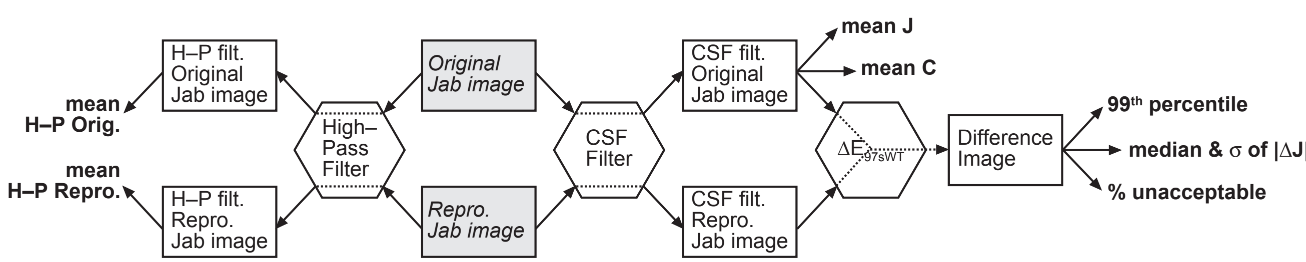
The previous sections have shown how taking into account the colour statistics of differences between an original and a reproduction, the colour statistics of the original itself, changes in spatial detail as well as some aspects of the visual system can improve the ability to predict observer judgements (Fig. 9). In summary the colorimetrically–based visual image difference importance metric (ΔI_{CM}) proposed here consists of the stages shown in Fig. 10.

In the equation shown there all data is based on CSF–filtered versions of the original and reproduction, k is a scaling constant, ΔE is ΔE_{97sWT}, σ_{J} is the standard deviation of lightness differences, $p_{6\bullet E}$ is the percentage of pixels that have ΔEs no greater than 6 units, h is a high–pass filtered image and the O and R subscripts denote the original and reproduction respectively. The over–bar in the equation denotes the mean and *med*() is a function providing the median value of its inputs. As can be seen the final metric shown here has not had optimised weights added to each of its factors as it was felt that the data–set available here was not large enough to allow for this. The principal aim of this metric is to show which factors can improve the relationship between colorimetrically based metrics and observer–reported judgements on the importance of visual differences.

The values of the parameters used in this metric and the ΔI_{CM} values for the 32 reproductions on which it is based can be seen at http://colour.derby.ac.uk/~jan/dicm/.

Conclusions

This paper has attempted to show how observer responses about colour differences in colour reproduction can be predicted by taking into account a range of colorimetrically–based parameters from an original and its reproduction. Knowledge gained from previous studies of colour difference in complex images as well as work on gamut clipping is applied and shown to result in improvements even under the present conditions. The final metric arrived at here is not the complete answer to predicting observer responses in the type of experiment dealt with here but it does show how the agreement between colorimetric data and visual responses can be considerably strengthened if parameters about the distribution of colour differences and original colours are taken into account. Finally, the metric presented here is proposed for further verification and extension and is part of an ongoing effort to improve the understanding of colour differences between complex images. While the metric was derived on the basis of reproductions obtained using gamut mapping algorithms, it is also intended for use in other applications where image differences need to be modelled.



$$\Delta I_{CM} = k\left(\Delta \bar{E} + med(|\Delta J|) - \sigma_{\Delta J}\right)\left(1 + p_{6\Delta E}\right)\left(1 + \frac{100 - \bar{J}_O}{100}\right)\left(1 + \frac{200 - \bar{C}_O}{200}\right)\left(1 + \left|\bar{h}_O - \bar{h}_R\right|\right)$$

Figure 10. Final colorimetrically–based visual image difference importance metric and flow–chart of its parameters.

Acknowledgements

The authors would like to thank Tadashi Yoshida and Todd Newman (Canon Development Americas, Inc.) for generously supporting this research.

References

1. P. L. Sun and J. Morovic (2002) What Differences Do Observers See In Colour Image Reproduction Experiments? *CGIV 2002*, Poitiers, France.
2. J. Morovic (1998) *To Develop A Universal Gamut Mapping Algorithm*, PhD. Thesis, University of Derby, Derby, UK.
3. G. J. Braun (1999) *A Paradigm for Color Gamut Mapping of Pictorial Images,* PhD. Thesis, Rochester Institute of Technology, Rochester, NY.
4. E. D. Montag and H. Kasahara (2001) Multidimensional Analysis Reveals Importance of Color for Image Quality, *Proc. 9th IS&T/SID Color Imaging Conf.*, 17-21.
5. P. L. Sun and J. M. Morovic (2002) 3D Histograms in Colour Image Reproduction, *Proc. of SPIE*, **4663**/9.
6. CIE (1986) CIE Publication 15.2, *Colorimetry*, 2nd Edition
7. M. R. Luo and R. W. G. Hunt (1998) The Structure of the CIE 1997 Colour Appearance Model (CIECAM97s), *Color Research and Application*, **23**:138–146.
8. M. D. Fairchild (1998) *Color Appearance Models*, Addison–Wesley
9. E. Peli (1990). Contrast in complex images. *J. Optical Soc. Am. – A*, **7**(10):2032–2040.
10. J. Uroz, M. R. Luo and J. Morovic (2000) Colour Difference Perceptibility for Large–size Printed Images, *Proc. Colour Image Science 2000 Conf.*, University of Derby, UK, 138–151.
11. X. M. Zhang and B. A. Wandell (1996) A spatial extension to CIELAB for digital color image reproduction, *Proc. SID Symp.*
12. N. Katoh and M. Ito (1996) Gamut Mapping for Computer Generated Images (II), *Proc. 4th IS&T/SID Color Imaging Conf.*, 126–129.
13. F. Ebner and M. D. Fairchild (1997) Gamut Mapping from Below: Finding the Minimum Perceptual Distances for Colors Outside the Gamut Volume, *Color Research and Application*, **22**:402–413.
14. M. Stokes, M. D. Fairchild and R. S. Berns (1992) Colorimetrically Quantified Visual Tolerances for Pictorial Images, *TAGA/ISCC Proc.*, **2**:757–777.
15. R. C. Gonzalez and R. E. Woods (1992) *Digital Imaging Processing*, Addison Wesley, 199–200.
16. G. M. Johnson and M. D. Fairchild (2001) Darwinism of Color Image Difference Models, *Proc. 9th IS&T/SID Color Imaging Conf.*, 108–112
17. C. J. Li, M. R. Luo and R. W. G. Hunt (1999) The CAM97s2 Model, *IS&T/SID 7th Color Imaging Conference: Color Science, Systems and Applications*, 262–263.

Biographies

Dr. Ján Morovic received a PhD in Colour Science at the Colour & Imaging Institute (CII) where the topic of his research was “To Develop a Universal Gamut Mapping Algorithm.” He now works at the CII as lecturer in Digital Colour Reproduction and is module leader for three modules on the MSc in Imaging Science. Further he also serves as chairman of the CIE's technical committee 8–03 on Gamut Mapping.

Mr. Pei–Li Sun received his bachelor and master degrees from the Department of Graphic Communications of the Chinese Culture University in Taipei, Taiwan, 1994 and 1996 respectively. Currently he is working at the Colour & Imaging Institute of Derby University towards his PhD entitled “The Influence of Image Characteristics on Colour Gamut Mapping.” He is about to commence work as a lecturer at the Department of Graphic Communications and Technology of the Shih Hsin University, Taiwan.