Predicting Image Differences in Color Reproduction from their Colorimetric Correlates

Ján Morovič

Color & Imaging Institute, University of Derby, Derby, United Kingdom and Hewlett-Packard Company, Barcelona, Spain

Pei-Li Sun

Department of Graphic Communications and Technology, Shih Hsin University, Taiwan, R.O.C.

This paper describes work carried out to predict the result of a psychophysical experiment in which observers made judgments about the types of differences they perceived between originals and reproductions rendered on different media. The results of these observer–reported visual data are compared with analogous metrics extracted from colorimetric data of the corresponding originals and reproductions. While there is good agreement in terms of the most general findings, looking at more detailed results shows significant differences between visual and colorimetrically–based information. The paper then introduces a colorimetrically based image difference metric that takes into account some aspects of the human visual system. Using information both about the statistics of color differences, of the original images and of changes to spatial characteristics, the metric is able to give a close prediction of observer responses. The final delta-ICM metric is proposed for further testing as a means of predicting observer responses of image difference in cross–media color image reproduction as well as other applications.

Journal of Imaging Science and Technology 47: 509–516 (2003)

Introduction

In cross-media color 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 an image's spatial detail, e.g., when a reflective original is rendered on a display, in terms of the image's colors, 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 could be the first step towards improving the performance of the image reproduction systems that was used for rendering them.

Even though a large number of solutions to cross-media color image reproduction have been proposed,[1] each of them tends to work well only under some circumstances. More specifically it is that part of the cross-media reproduction process that deals with overcoming differences between the original and reproduction media's gamuts that is responsible for variation. The reason for this is that the other parts of the process, i.e., those modeling the nature of the media and of color appearance, are descriptive, whereas the gamut mapping stage is transformative. For example, some gamut mapping algorithms (GMAs) are particularly suited for making printed reproductions of originals present on transparencies, whereas others work well when trying to match images using various printing technologies. Then there are GMAs that work well for certain images but not for others and consequently a fully automatic cross-media color image reproduction system is still a utopia.

The Structure of Gamut Mapping Studies

Looking at the way in which GMAs have been developed shows that in the vast majority of cases researchers start with some idea of how to change an image's colors so as to make them fit a reproduction imaging medium's gamut. In many cases this initial idea is based either on experience from doing color reproduction in a trial and error way or on *a priori* theoretical grounds that the reproduction ought to have certain properties, e.g., that there should be a certain balance between lightness and chroma changes applied to the original.

Some other studies first try to understand how observers would gamut-map images and they then attempt to model the observer behavior. Understanding how observers gamut-map images can either be done by looking at what color reproduction professionals do with the tools used commercially[2,3] or by developing tools that allow images to be modified in terms of some of their appearance attributes and then getting naïve observers to make adjustments to reproductions so as to make them more accurate or pleasant.[4–6]

To extend the bases for GMA development a study was carried out in which observers were simply asked what differences they saw between originals and their

Original manuscript received February 3, 2003

jan.morovic@hp.com; gcsun@cc.shu.edu.tw

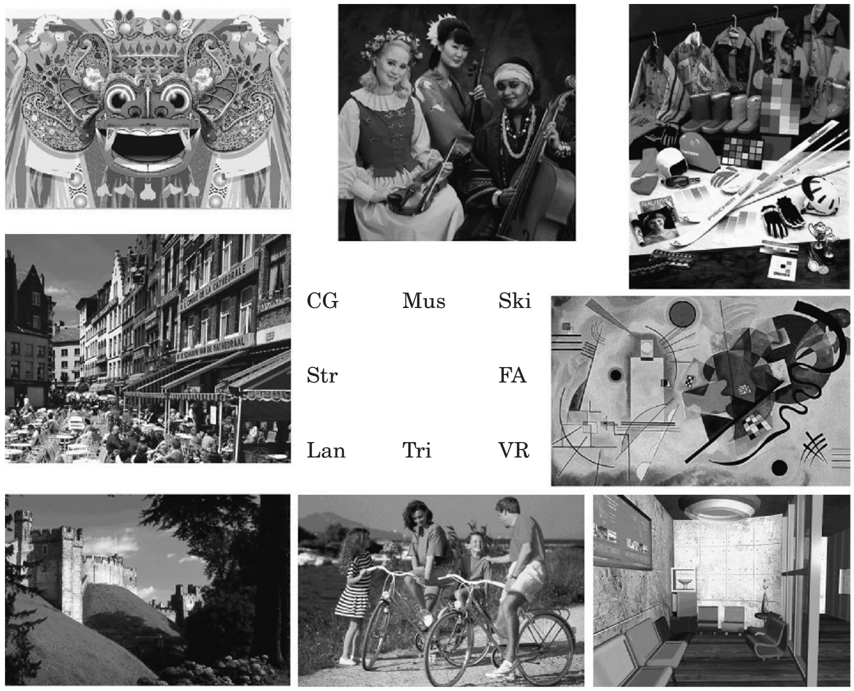


Figure 1. Test images used in psychophysical experiment.

reproductions obtained using different GMAs.[7] Analyzing such observer responses showed first, what kinds of differences were considered important between originals and reproductions, e.g., loss of detail, chroma changes, contrast reduction, and second, how important these differences were compared to each other. Knowing what differences still remain between originals and reproductions can also serve as a basis for developing new methods of gamut mapping that focus on image characteristics that are not dealt with well in existing approaches.

Further, as the results of the above experiment consist of images and difference judgments between them, they can also be used to develop image difference metrics. Such a metric, which predicts observer judgments for differences between images rendered on different media, could then be used for developing new GMAs and for predicting how well they perform without the need for conducting resource intensive psychophysical experiment. Hence, the aim of this paper is to predict observer judgments from colorimetric data about originals and their reproductions.[8]

The following sections will first briefly describe the experimental setup in which the present data was obtained, present the results of the visual differences 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 analyzed and an improvement of its strength will be attempted by developing an image difference metric.

Psychophysical Experiment

To gain an understanding of what observers take into account when judging reproductions, a psychophysical experiment[7] was set up in which observers were shown an original on a CRT display and a number of printed reproductions of it in a viewing booth. These reproductions were always presented in pairs so as to make the conditions similar to a pair-comparison experiment[9] already carried out using the same stimuli.

Eight test images were used as originals in this study (Figure 1) and each of them was reproduced using the following four gamut mapping algorithms: SKNEE,[5] WCLIP,[7] CARISMA and GCUSP.[10] Fifteen color–normal observers then participated in the experiment where they were asked to list, in their own words, all the differences they could see between an original and its reproduction. The group of observers, all of whom passed the Ishihara test, consisted of six females and nine males whereby nine were of Asian and six of European origin. The age range of observers was 26 to 42 years with a mean of 30.5. In terms of experience with judging image appearance two of the observers were naïve, 12 had some experience and one was a professional photographer.

Hence the result of this experiment is a record of 15 sets of observer reported differences for 32 image pairs, i.e., the pairs formed by the eight originals and the four reproductions made of each of them. Observers reported differences like changes in lightness (L), colorfulness (C) and hue (H) as well as differences in detail and contrast. Further they also reported that some reproductions were pale, faded or blurred when compared with the corresponding original and all judgments 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 judgment 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 judgments to a single scale. De-

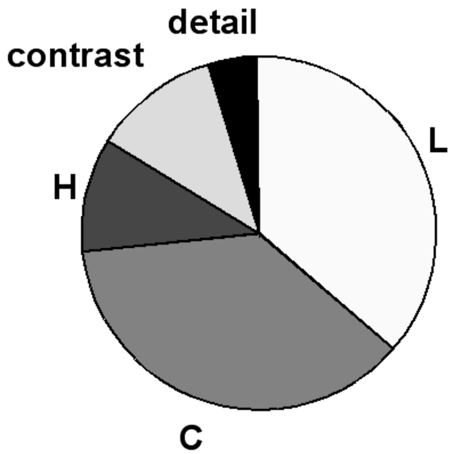


Figure 2. Relative importance of visual observer-reported difference types.

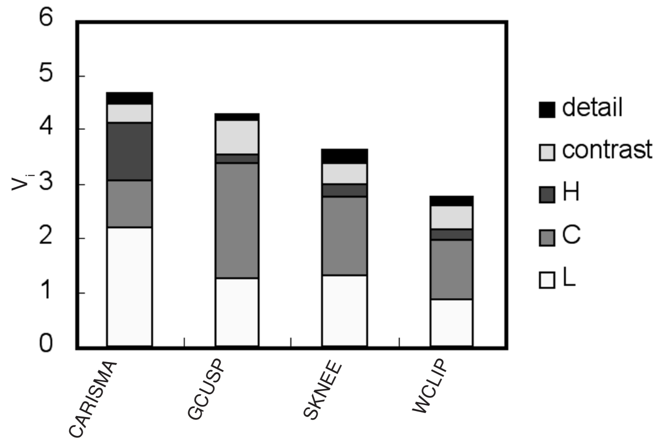


Figure 3. Visual differences for individual GMAs.

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

tails of how the experiment was set up, conducted and analyzed can be found elsewhere.[7]

Experimental Results

The overall results of the above experiment are shown in Fig. 2 and it can be seen there that over 80% of differences were reported to be due to differences in color attributes between originals and reproductions. Only less than 20% were due to differences in contrast or spatial detail. Furthermore the ratio of L:C:H was 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[7,11], that looked at the relative importance of various types of differences. Even though these studies used very different approaches, they also found that color differences are responsible for a very large proportion of differences in color reproduction.

The proportion of these differences was also looked at on a GMA-by-GMA basis (Fig. 3) and this shows what kinds of differences were reported for the reproductions made with the 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.

Finally, the results also showed where in the images observers perceived greatest differences (Fig. 4).

From Fig. 4 it seems that location within an image is not a factor influencing the attention of observers. In-

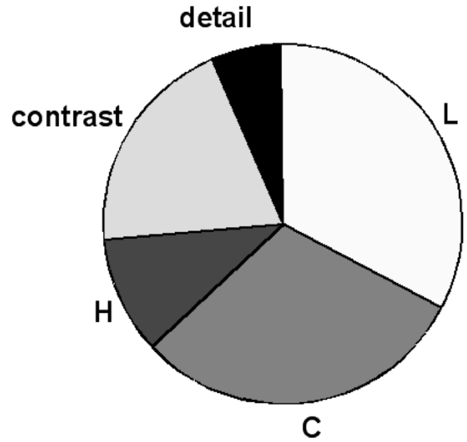


Figure 5. Relative magnitude of types of colorimetrically-based differences.

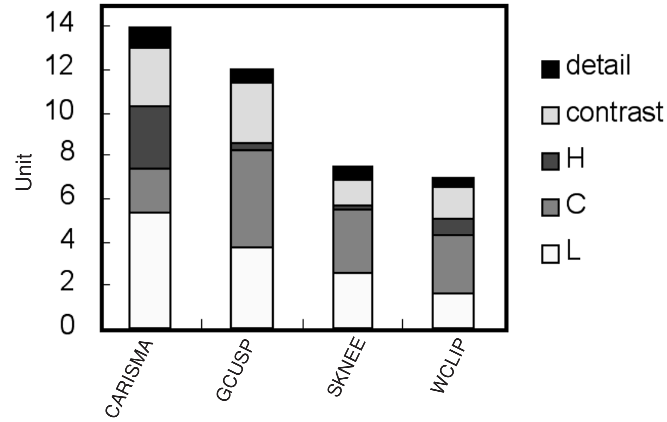


Figure 6. Colorimetrically–based differences for individual GMAs.

stead large uniform backgrounds, especially sky, are more likely to be regions where observers identify important visual differences. Surprisingly areas of flesh tones did not show important differences in the present study and there are at least two possible reasons: first, skin colors did not occupy large areas in the test images used here and second, the colors were not very chromatic and thus changed only slightly as a result of the gamut mapping process. Further analysis of the observer reported differences can be found elsewhere.[12]

At this stage it is also important to note that it might be impossible to reduce some of the differences identified here by observers due to the inherent differences between the original and reproduction media.

Comparing Visual and Colorimetric Data

Next the focus will be on analyzing metrics derived from the colorimetric image data of the eight originals used here and of their reproductions so as to determine what relationship there is between them and the observers' visual judgments. If a strong relationship is found then observer judgments could also be predicted for other images and image reproduction systems and the behavior of the latter could be adjusted to minimize them.

The first difficulty that comes about when attempting such a comparison is 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[13] L^*, CIECAM97s[14] J, RLAB[15] L^R, 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[16] 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 numerous 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 Fig. 6 does so individually for the four algorithms.

Lightness, chroma and hue differences were here predicted using CAM97s2,[17] however CAM02[18] could now be used instead and it is likely that differences between using these two color appearance models would be negligible for this application. Contrast difference was predicted using the difference between the standard deviations of original and reproduction image lightnesses. Detail difference was modeled using the difference between mean values of high-pass filtered versions of original and reproduction lightness images. A pixel's value in these high-pass filtered images was obtained by subtracting the mean of the pixel's 3x3 pixel neighborhood from the pixel's own value in the unfiltered image.

In Fig. 5 it can be seen that the overall colorimetric data does show a similar picture to the visual results – the ratio of color 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, most notably for contrast and detail but also for lightness, chroma and hue.

Finally, Fig. 7 contains images showing pixel-by-pixel CAM97s2 Euclidean distances (ΔE_{97s2}) between originals and their corresponding reproductions and while there are clear differences with the visual results, there are also significant similarities.

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

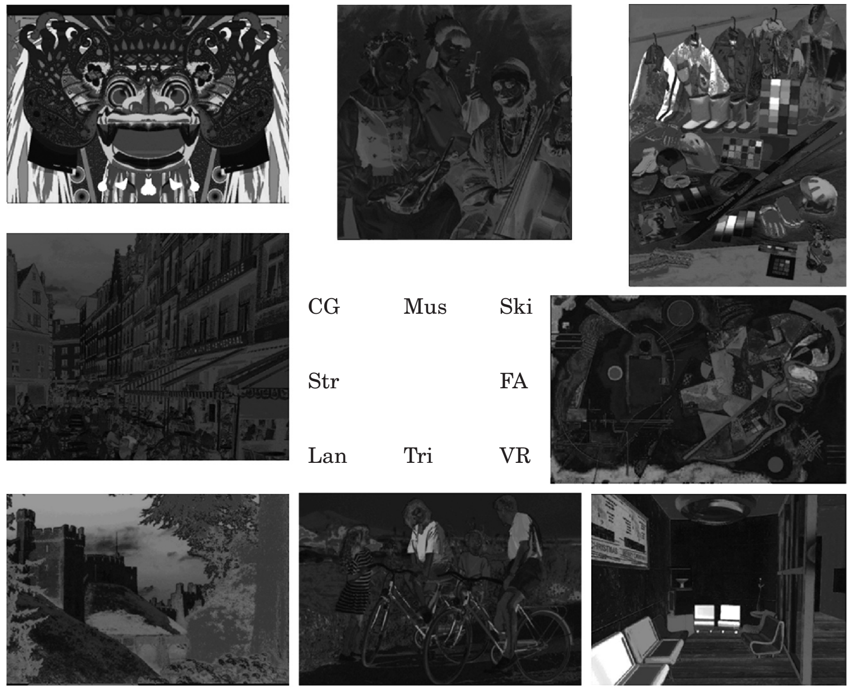


Figure 7. Local ΔE_{97s2} differences (black represents zero and white represents 35.2)

TABLE I. 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

(Table I) which show very low levels of correlation between the two.

The correlation between visual and colorimetric attributes can also be studied using *Principal Components Analysis (PCA)* with *Multidimensional Scaling (MDS)*. A p by n data matrix, i.e., p variables with n observations, can be decomposed into a product of its characteristic vectors, characteristic roots, and principal component scores using PCA. A multidimensional representation of the characteristic vectors for each variable can then display the relationships among the variables. In practice, two-dimensional plots, so-called "biplots", are used because they are easily visualized, but they need to be accompanied by a statement of what proportion of the total variability is accounted for by the first two characteristic roots that are its dimensions.[19]

Figure 8 shows the "biplot" of this study's experimental data where variables p refer to the ten sets of both visual and colorimetric data and the number of observations n refers to the 32 pairs of images. The amount of variability accounted for by the first two characteristic roots was 41.2% and 24.2% respectively and the plot hence accounts for 65.4% of their interrelationship. The characteristic vectors for each variable are also rotated to make the linear functions formed by each characteristic root mutually orthogonal.

In Fig, 8, ΔV_i and ΔC_m represent visual and colorimetric attributes respectively, the size of circles represents relative magnitude and the distance between each of the attributes indicates how close their relationship is.

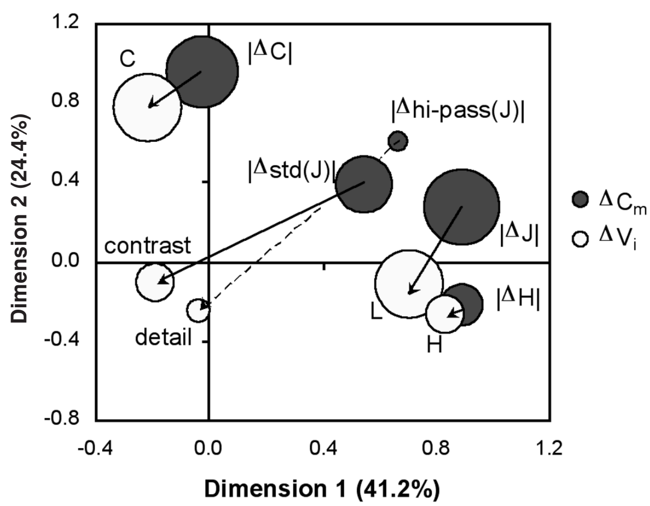


Figure 8. "Biplot" of interrelationship (correlation) between ΔV_i and ΔC_m.

As can be seen, chroma and hue have better correlation between their ΔV_i and ΔC_m values than other factors. Lightness, which dominated 36.4% of overall ΔV_i, shows lower correlation and visual contrast and detail differences are far from our colorimetric predictions too.

Seeing that the relationship between colorimetric and observer results is weak should not come as a surprise, given that the former is fundamentally based on data obtained for uniformly colored patches seen against uniform backgrounds whereas the latter are judgments made about properties of complex images.

To improve the relationship between colorimetrically based data and even just visual results that are in terms

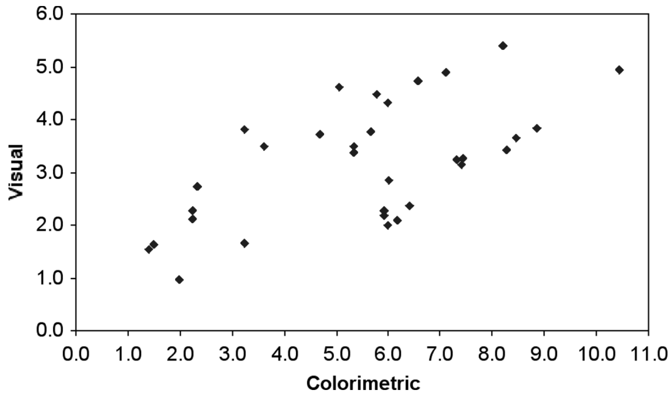


Figure 9. Visual versus mean ΔE_{97s2} values for 32 reproductions.

of color attributes it will be necessary to take into account a number of factors beyond just the color attributes of image pixels. The aim of the next section 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.

Improving Correlation Between Colorimetric and Visual Results

As it is the color differences that were found to be by far the most important ones both in the present study and in previous work,[7,11] the focus here will be on predicting only this part of the visual results and to leave contrast and detail for future work. Table II already showed that R^2 was only 0.378 between the combined LCH observer reported visual differences and the mean ΔE_{97s2} values for the 32 reproductions used here. Looking at the relationship between the two types of data for individual reproductions shows further that the relationship is indeed weak (Fig. 9).

99th Percentile

A first step towards improving this correlation is to look at the 99th percentile of ΔE_{97s2} distributions as high percentiles have previously been reported to better correlate with perceived differences between complex images.[20] In this study too the correlation between the visual LCH data and the 99th percentile of ΔE_{97s2} 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[21] model were applied to the present colorimetric data with an improved correlation between the 99th percentile of ΔE_{97s2} 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 color reproduction than unweighted equations,[4,22] the 99th percentile of weighted color differences (ΔE_{97s2WT}) between filtered images was used. ΔE_{97s2WT} 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 color 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.[20,23] 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 judgments 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 colors 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[24] has also previously been found to enhance color image difference prediction.[25]

As the spatial processing was performed on lightness data this might require some justification. Typically spatial processing of images is performed in luminance and this represents a processing in a domain linearly related to reflectance and hence to the physical properties of an image. On the other hand spatially processing images in an appearance space simply means that the quantities that are being altered refer to perceived magnitudes rather than physical ones. In other words the intention here was to extract differences between the lightness appearances of neighbouring image pixels rather than between their luminances, which are perceptually non-linear.

Final Colorimetrically Based Metric (ΔI_{CM})

The previous sections have shown how taking into account the color statistics of differences between an origi-

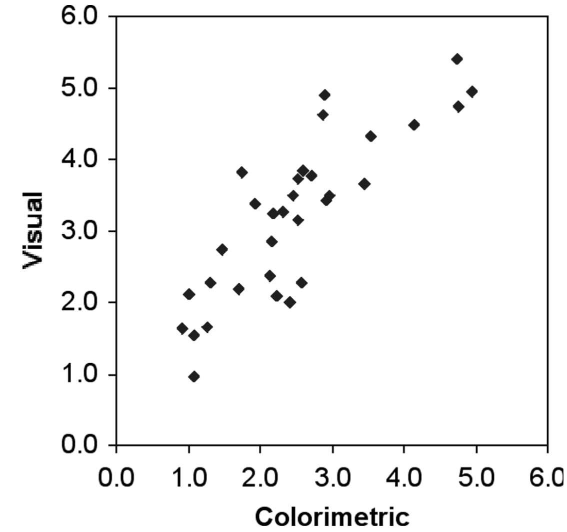
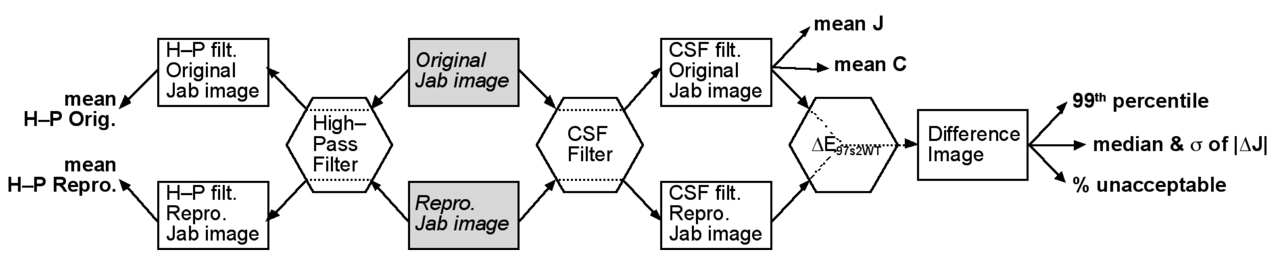


Figure 10. Visual versus final colorimetrically based values for 32 reproductions.

nal and a reproduction, the color 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 judgments (Fig. 10). The predictions of this metric can also be compared to the visual results in terms of difference whereby the mean prediction error is 0.68 ΔVi units (which are on a 0-8 unit accuracy scale) and the maximum error is 2.06. In other words, the metric's predictions are on average within ±1 accuracy category of the actual observer judgment.

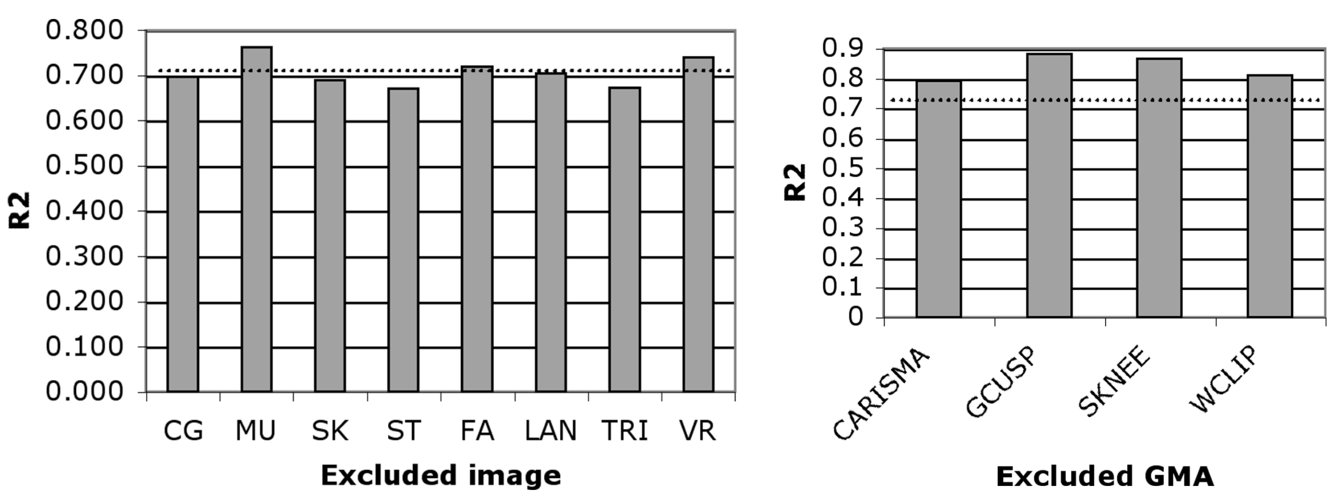
In summary, the colorimetrically based visual image difference importance metric (ΔI_{CM}) proposed here consists of the stages shown in Fig. 11. In the equation shown there all data is based on CSF-filtered versions of the original and reproduction, k is a scaling constant, ΔE is ΔE_{97s2WT}, $\sigma_{\Delta J}$ is the standard deviation of lightness differences, $p_{6\Delta E}$ 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.[26]

As can be seen the final metric shown here has not had optimized 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



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

Figure 11. Final colorimetrically–based visual image difference importance metric and flow–chart of its parameters.



(a) (b)

Figure 12. Effect of excluding (a) and image and (b) a GMA.

colorimetrically based metrics and observer reported judgments on the importance of visual differences rather than to promote a specific image difference model.

Image and GMA Dependence of Results

Even though the number of images used in this study is large in the context of image difference evaluation, it is nonetheless small in absolute terms. In a similar way there is also an issue with the number of change types, i.e., GMAs, applied to the originals and it is therefore of value to see how much the chosen images and algorithms affect the final result.

To do this, the determination coefficient has been computed between such subsets of the full data that exclude one image or GMA at a time. The results of excluding the four reproductions of each of the eight test images in turn is shown in Fig. 12a and it can be seen that the agreement between the statistics of each of the seven image subsets and the overall R^2 is very good. Excluding the MUS image has the largest R2 of 0.762 among the seven image subsets, the smallest R^2 of 0.671 is obtained when the Str image is not included, and the standard deviation of the eight R^2s is 0.032. As such it can be seen that using only seven of the eight images would have resulted in similar results regardless of which image is excluded.

Applying the same procedure to the GMAs used here shows a different result (Fig. 12b) in that the relationship between results of just three GMAs is stronger than for the full set of four and that this is so for all three-GMA subsets. In other words, if only the data from three reproductions of all the eight originals is taken into account then the coefficient of determination is larger than for all four reproductions. As such it can be seen that the number of GMAs plays a stronger role in the current results than the number of images. Furthermore this analysis shows that if only a smaller number of GMAs are of interest then the current approach can lead to even greater accuracy than for the four GMAs used here.

Conclusions

A new approach to understanding the cross-media color image reproduction process was the basis for the work described this paper. In particular an experiment was described that resulted in information about what differences observers see between originals and reproductions. These experimental results showed that color differences were more important than contrast and detail differences in cross-media reproduction.

The main aim of this study was then an attempt to show how these observer responses could be predicted by taking into account a range of colorimetrically based parameters from an original and its reproduction. Knowledge gained from previous studies of color difference in complex images as well as work on gamut clipping was applied and shown to result in improvements even under the present conditions.

While the final metric arrived at here is by no means a complete answer to predicting observer responses in the type of experiment dealt with here, it does show how the agreement between colorimetric data and visual responses can be considerably strengthened if parameters about the distribution of color differences and original colors are taken into account. Furthermore the observer dependence of the target visual difference results could also be analyzed in the way in which image and GMA differences were analyzed here.

Finally, the ΔI_{CM} metric presented here is proposed for further verification and extension and is part of an ongoing effort to improve the understanding of color 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 modeled.

Acknowledgment. The authors would like to thank Tadashi Yoshida and Todd Newman (Canon Development Americas, Inc.) for generously supporting this research and the reviewers of the first draft of this manuscript for their invaluable recommendations.

References

1. J. Morovic and M. R. Luo, The Fundamentals of Gamut Mapping: A Survey, *J. Imaging Sci. Technol.* **45,** 283–290 (2001).
2. CARISMA (1992) *Color Appearance Research for Interactive System Management and Application - CARISMA, Work Package 2 - Device Characterisation, Report WP2-19 Color Gamut Compression*.
3. P. Green and M. R. Luo, Developing the CARISMA Gamut Mapping Model, *Proc. Color Image Science 2000 Conference*, University of Derby, Derby, UK, 2000, pp. 244–256.
4. F. Ebner and M. D. Fairchild, Gamut Mapping from Below: Finding the Minimum Perceptual Distances for Colors Outside the Gamut Volume, *Color Res. Appl.* **22,** 402–413 (1997).
5. G. J. Braun, *A Paradigm for Color Gamut Mapping of Pictorial Images*, Ph.D. Thesis, Rochester Institute of Technology, Rochester, NY, 1999.
6. B. H. Kang, M. S. Cho, J. Morovic, and M. R. Luo, Gamut Compression Analysis Based on Observer Experimental Data, in *Proc. 7th IS&T/SID Color Imaging Conf.*, IS&T, Springfield, VA, 1999, 295–300.
7. P. L. Sun and J. Morovic, What Differences Do Observers See In Color Image Reproduction Experiments? *Proc. of IS&T CGIV 2002*, IS&T, Springfield, VA, 2002, pp. 181–186.
8. J. Morovic and P. L. Sun, Visual Differences In Color Reproduction And Their Colorimetric Correlates, *Proc. 10th IS&T/SID Color Imaging Conf.*, IS&T, Springfield, VA, 2002, pp. 292–297.
9. P. L. Sun and J. M. Morovic 3D Histograms in Color Image Reproduction, *Proc. SPIE* **4663,** 51–62 (2002).
10. J. Morovic, *To Develop A Universal Gamut Mapping Algorithm*, Ph.D. Thesis, University of Derby, Derby, UK, 1998 .
11. E. D. Montag and H. Kasahara, Multidimensional Analysis Reveals Importance of Color for Image Quality, *Proc. 9th IS&T/SID Color Imaging Conf.*, IS&T, Springfield, VA, 2001, pp. 17–21.
12. P. L. Sun, *Influence of Image Characteristics on Color Gamut Mapping for Accurate Reproduction*, Ph.D. Thesis, Color & Imaging Institute, University of Derby, 2002.
13. CIE Publication 15.2, *Colorimetry*, 2nd ed., CIE, Vienna, Austria (1986).
14. M. R. Luo and R. W. G. Hunt, The Structure of the CIE 1997 Color Appearance Model (CIECAM97s), *Color Res. Appl.* **23,** 138–146 (1998).
15. M. D. Fairchild, *Color Appearance Models*, Addison-Wesley, New York, 1998.
16. E. Peli, Contrast in complex images, *J. Opt. Soc. Amer. A*, **7**, 2032–2040 (1990).
17. C. J. Li, M. R. Luo and R. W. G. Hunt, The CAM97s2 Model, *Proc. 7th IS&T/SID Color Imaging Conf.*, IS&T, Springfield, VA, 1999, pp. 262–263.
18. N. Moroney, M. D. Fairchild, R. W. G. Hunt, C. Li, M. R. Luo, and T. Newman, The CIECAM02 Colour Appearance Model, *Proc. 10th IS&T/SID Color Imaging Conf.*, IS&T, Springfield, VA, 2002, pp. 23–27.
19. J. E. Jackson, *A User Guide to Principal Components*, John Wiley and Sons, Inc., New York, 1991.
20. J. Uroz, J. Morovic and M. R. Luo, Perceptibility Thresholds of Color Differences in Large Printed Images, in *Color Image Science: Exploiting Digital Media*, L. W. MacDonald and M. R. Luo, Eds., John Wiley and Sons, 2002, pp. 49–73.
21. X. M. Zhang and B. A. Wandell, A spatial extension to CIELAB for digital color image reproduction, *SID Symposiums Technical Digest* **27,** 731–734 (1996).
22. N. Katoh and M. Ito, Gamut Mapping for Computer Generated Images (II), *Proc. 4th IS&T/SID Color Imaging Conf.,* IS&T, Springfield, VA, 1996, PP. 126–129.
23. M. Stokes, M. D. Fairchild and R. S. Berns, Colorimetrically Quantified Visual Tolerances for Pictorial Images, *TAGA/ISCC Proc.* **2,** 757–777 (1992).
24. R. C. Gonzalez and R. E. Woods, *Digital Imaging Processing*, Addison Wesley, New York, 1992, pp. 199–200.
25. G. M. Johnson and M. D. Fairchild, Darwinism of Color Image Difference Models, *Proc. 9th IS&T/SID Color Imaging Conf.,* IS&T, Springfield, VA, 2001, pp. 108–112
26. 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 found at http://colour.derby.ac.uk/~jan/dicm/.