Verification of Gamut Mapping Algorithms in CIECAM97s Using Various Printed Media

Ján Morovic and M. Ronnier Luo
Colour & Imaging Institute, University of Derby
Derby, United Kingdom

Abstract

The aim of this paper is to describe a study carried out to investigate how the performance of five selected gamut mapping algorithms (GMAs)[1] is influenced by the difference in gamut size between two printed media. In addition to this, the GMAs were implemented in CIECAM97s[2] so as to verify their performance in a different colour space and also to see what effect the new colour appearance model's colour space has compared with CIELAB. Based on a pair comparison experiment, it was found that the algorithms which performed best in CIELAB again performed best in the new colour space and that the magnitude of gamut difference is indeed of importance.

Introduction

The present paper is the latest in a series describing the development of **universal gamut mapping algorithms** at the Colour & Imaging Institute. This project started by first comparing a number of existing algorithms[3] in terms of their accuracy, then developing new second–generation GMAs on the basis of their evaluation and testing them in comparison with the initial GMAs.[4] The results of the second experiment then served as a basis for the development of further new third–generation algorithms.[1]

Throughout this project, the aim was to find such universal algorithms, which give the most accurate reproductions. However, as the pleasantness of reproductions in isolation is also of great importance, it was studied too and found to be strongly and positively correlated with accuracy.[5]

In this context, the aims of the present study were the following four: (a) to evaluate the third–generation UniGMA and LCUSPH algorithms, (b) to investigate the effect of changing the gamut mapping colour space from CIELAB to **CIECAM97s** on CARISMA, GCUSP[4] and LLIN[3], (c) to obtain additional information about the performance of CARISMA and GCUSP (which performed best in the previous experiment) and (d) to study the influence of the magnitude of **gamut difference** on the performance of the selected algorithms.

Experimental Setup

To achieve the above aims, a colour reproduction system was implemented, which consisted of a CRT and a printer on which reproductions were made on two media—a glossy synthetic substrate and an uncoated inkjet paper.

In this colour reproduction system the PLCC model[6] was used for characterising the Barco Calibrator CRT, which was measured using a Bentham telespectroradiometer and the third order masking equations[7] with greyscale correction[4] were used for characterising the HP DeskJet 850C printer on the basis of measurements made with an X–Rite 938 spectrophotometer.

Table 1. Overview of Evaluated GMAs.

Mnemonic	Description
GCUSP	C dependent J compression + compression to J of cusp on J axis
CARISMA	J compression + hue shift + relative, gamut shape dependent mapping
UniGMA	hue shift + relative, gamut shape dependent J compression and mapping
LCUSPH	J compression + hue shift + compression to J of cusp on J axis
LLIN	J compression + linear C compression

Gamut mapping between these media was carried out using five GMAs (Table 1) in a colour space determined by the lightness (*J*), chroma (*C*) and hue (*h*) predictors of CIECAM97s whereby its parameters were those for an average surround (i.e. F=1.0, c=0.69, F_{LL}=1.0 and N_c=1.0). For both media the adopted white was the medium white (i.e., R=G=B=100% for the CRT and the substrate for the printer) and the background (Y_b) had the same chromaticity as the adapting white and 20% of its luminance.

The luminance of the adapting field (L_A) was mistakenly set to 90 cd/m^2, instead of being set to 20% of the adapting white's luminance (i.e., 17 cd/m^2), which would have more closely described the conditions used in the experiment and which is also recommended in the model. Fortunately, the consequences of this mistake resulted in little change to the evaluated reproductions both in terms of colour gamut and contrast (Figure 1), which means that choosing the correct L_A value would have given very similar reproductions.

Reproductions of five test images were then made using the five GMAs and their accuracy was evaluated by 10 observers in a pair comparison experiment. Note, that judgements were made for individual image regions of characteristic colour as well as for the overall appearance of reproductions. Throughout these experiments it was assumed that the appearance of the image on the original device (CRT) is what needs to be reproduced (i.e. the algorithms had no image enhancing intents).

Both the experimental procedure and data analysis used here were the same as described in a previous paper[5] with the difference that once z–score matrices were calculated, the accuracy scores were calculated by averaging the z–scores in each GMA's column and that the 95% confidence interval for these scores was $\pm\sigma/(N1/2)$. Here σ is the standard deviation and based on Case V of Thurstone's paper on pair comparison[8] it equals $1/(2^{1/2})$ and N is the sample–size on which the accuracy score is based.

Influence of CIECAM97s on GMAs

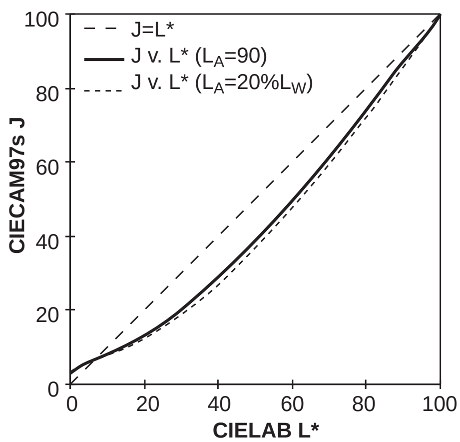

Figure 1. J versus L plot for XYZ values of colours with equal RGB values (showing two L_A settings).*

The first effect of using CIECAM97s is a result of the uniformity of its **hue predictor**, which was previously shown[9] to be better in the blue region and worse in the red–yellow region than the hue predictor of CIELAB. Both these results could be observed in the reproductions made for this experiment—red colours tended to be bluer than the original and yellows greener; the hue of blue colours, on the other hand, was maintained after lightness and/or chroma changes.

The second effect of using CIECAM97s was a change in the contrast of images due to a difference in the lightness **predictors** of the two colour spaces. To understand this difference, Figure 1 shows the plot of J versus L* values of 86 achromatic colours. These were obtained by first calculating the XYZ tristimulus values from the device–dependent coordinates of colours which had equal RGB values and were equally–spaced and then calculating L* and J from these. It can be seen from this plot that the contrast of colours with an L* of above approximately 40 is increased, that of between approximately 20 and 40 is kept the same and that of below 20 is decreased.

The change of the lightness predictor also resulted in changes of lightness ranges for the media used here and hence also of the ratios between the lightness ranges of media pairs (Table 2).

Table 2. Media Lightness Ranges

medium	CRT	glossy print	glossy/ CRT	plain print	plain/ CRT
L* range	100.0	87.0	87.0%	76.6	76.6%
J range	97.3	90.9	93.4%	84.9	87.3%

As can be seen, the lightness–range ratios in CIECAM97s are larger than those in CIELAB, which means that the gamut differences are smaller. The extent of this is such that the CIECAM97s ratio of the plain paper medium's lightness range to that of the CRT is almost the same as the CIELAB ratio of the glossy paper medium's lightness range to that of the CRT.

Experimental Results

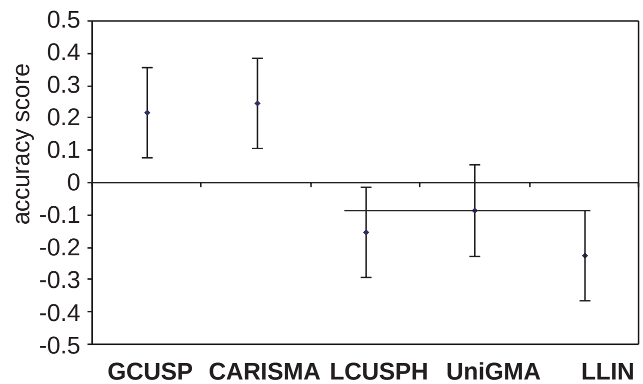

Figure 2. Overall results (based on 100 observations).

The overall results in Figure 2 (obtained by combining the results for five test images reproduced on two media) show that CARISMA and GCUSP gave significantly more accurate reproductions than the other three algorithms. Within the group of the bottom three algorithms the UniGMA algorithm performed best, while being on the boundary of the LLIN algorithm's 95% confidence interval.

Table 3. Variances of GMAs

	gcusp	carisma	lcusph	unigma	llin
variance	0.341	0.901	0.329	0.192	1.186

From these results it would seem that the two third–generation algorithms have failed completely, as they are significantly outperformed by CARISMA and GCUSP. However, a look at all the judgements made for the GMAs evaluated here shows that the new algorithms (in particular UniGMA) had lower variances (Table 3), which means that their performance was less influenced by colour region and test image. In addition to this, it is important to note that the accuracy scores dealt with here are relative within a given group (i.e., average scores on a relative scale do not imply average scores on an absolute scale). Nonetheless, the CARISMA and GCUSP algorithms gave a better performance overall. The variance scores also suggest a notable advantage of GCUSP over CARISMA, as the former has a much lower variance.

To have a better understanding of the five algorithms, 30 colours were chosen (Figure 3), gamut–mapped and their original and gamut–mapped *JCh* values were compared. Note, that the gamut–mapped values are not the *JCh* values of the reproduced colours, but are the values which would be transformed via the printing medium's characterisation model and then printed. That is, they do not include characterisation and printer–variation errors.

The statistics of the differences made to these colours by the five GMAs—combined for both printed media—are shown in Table 4. It is encouraging to see that the Pearson correlation coefficient between the median $|\Delta C|/|\Delta J|$ ratio and the GMAs' ranking is 0.94, which suggests the importance of maintaining more chroma at the expense of lightness (in relative terms). However, it needs to be noted that this correlation is much lower when the two printed media are considered separately (0.60 and 0.39 for the glossy and plain paper media respectively).

Table 4. Statistics of Changes Made to 30 Test Colours

GMA	Median ΔE_{97s}	Median $\|\Delta J\|$	Median $\|\Delta C\|$	Median $\|\Delta C\|/\|\Delta J\|$
gcusp	11.20	4.20	9.94	1.52
carisma	13.51	6.19	1.81	0.32
lcusph	15.44	4.57	6.05	1.82
unigma	15.27	3.27	8.32	1.67
llin	14.93	5.57	12.78	2.68

At the same time, it is interesting to note the strong negative correlation between the accuracy score and the median ΔE_{97s} colour difference, which is –0.82, –0.72 and –0.92 for the combined, glossy and plain media respectively (Note, that the ΔE_{97s} label has been used for the Euclidean distance in CIECAM97s as there is no colour difference formula specified in the model.). This suggests that (within the group of five GMAs considered here) the algorithms, which make the smallest change give the most accurate reproduction. It is also encouraging to see that the correlation between these two parameters is strong for the printed media both individually and collectively and that this was also the case for the results of a previous experiment.[4]

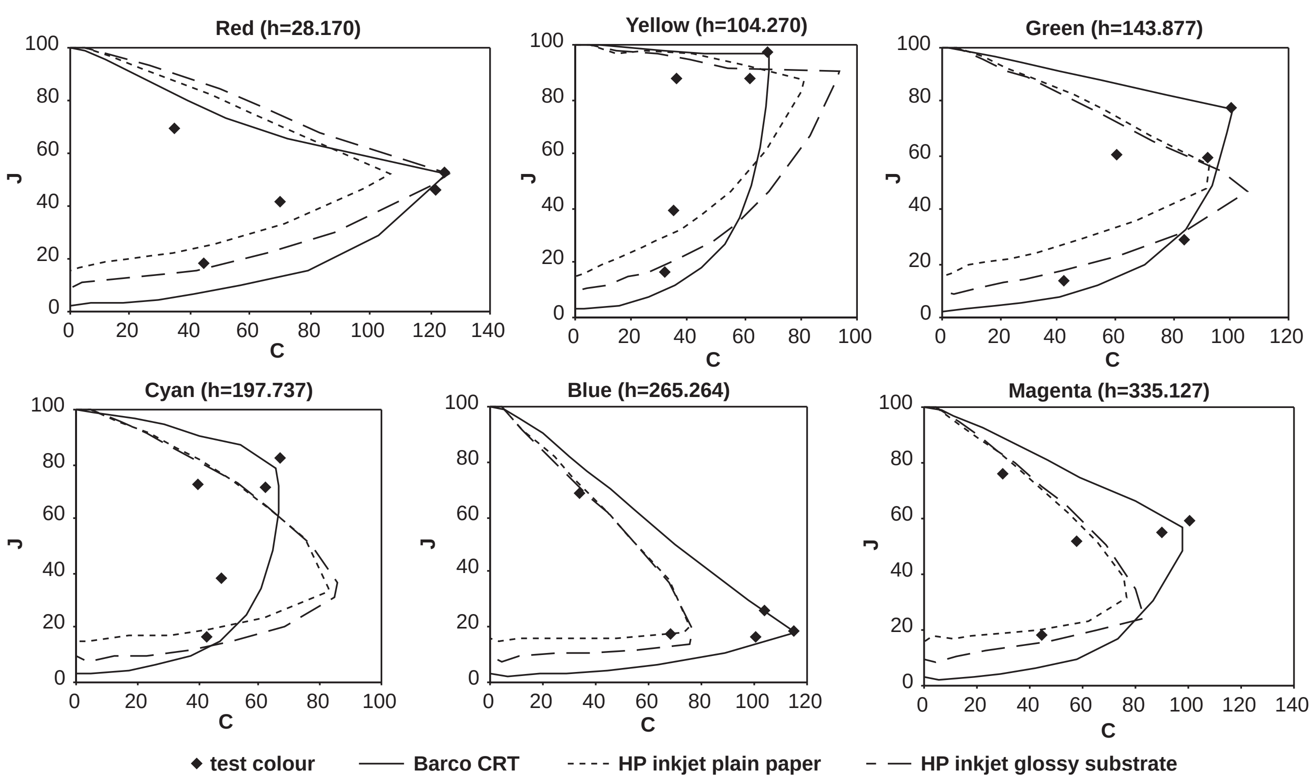

Figure 3. Gamut boundaries at CRT's primary and secondary hue angles and test colours used for investigating changes made by GMAs.

To sum up, it can be said that the accuracy of the CARISMA and GCUSP algorithms was significantly higher than that of the other results, whereby the GCUSP algorithm performed more stably and made a smaller overall change to the colours of the five test images used here.

Comparison of Results for Plain and Glossy Paper

For both media, the CARISMA and GCUSP algorithms were ranked as the top two and the LLIN algorithm was always ranked among the bottom two. Further, the LLIN performed better for the glossy substrate—i.e. where the gamut difference was smaller—and GCUSP performed better when the gamut difference was larger.

Looking at the individual accuracy scores obtained in this experiment shows that the range of accuracy scores was larger for plain paper than for glossy paper. This was the case for 80% of all judgements made for overall images and colour regions within them and the ratio of plain paper accuracy–score range to glossy paper accuracy–score range was 1.6. This means that there are larger differences between the reproductions made on plain paper than between the reproductions on glossy paper, which implies that the choice of GMA is more critical when the gamut difference is larger.

The relationship between the lightness range difference of the two printed media and the resulting accuracy–score range difference is also of interest. Here it can be seen that the plain printed medium has a lightness range which is 7% smaller than that of the glossy substrate (Table 2) and that this difference results in an accuracy–score range which is 60% larger. This clearly suggests the importance of gamut difference for the evaluation of gamut mapping algorithms and differences in this parameter might well have been the causes of differences in the results of previous experimental studies from different sources.

Table 5. Ranking of GMA Groups for Five Test Images

	BUS	DOL	MUS	NAT	SKI	mean
gcusp	1	1	1	2	2	1.4
carisma	2	2	2	1	1	1.6
lcusph	3	2	1	3	2	2.2
unigma	2	2	2	2	2	2.0
llin	4	1	1	2	3	2.2

Results for Individual Test Images

A useful way of looking at the performance for individual images is by considering the ranking of groups of GMAs whereby the grouping is done in terms of the 95% confidence interval (i.e. algorithms in the same group are not significantly different from each other). It can be seen here that the GCUSP and CARISMA algorithms are in the top two groups for each of the five test images, which is further evidence for their suitability as universal GMAs (Table 5).

Conclusions

The results of the experiment described here suggest that the magnitude of original and reproduction gamut difference has a significant influence on the range of GMA performance, which means that the choice of GMA for the reproduction of images between two media is more critical when there are larger differences between these.

Further, this experiment also shows that the GCUSP and CARISMA algorithms performed best in terms of overall results as well as the results for individual test images and colour regions in them. Of these, the GCUSP algorithms in particular also had a low variance of accuracy scores and is much simpler to implement and calculate. These characteristics, together with its good performance in previous experiments[4,5] makes GCUSP a promising candidate for a universal gamut mapping algorithm.

Acknowledgements

The authors would like to thank Prof. Tony Johnson and Dr. Peter Rhodes for their helpful suggestions.

References

1. J. Morovic and M. R. Luo, *Proc. CIM '98 Conf.*, Derby, UK, (1998).
2. M. R. Luo and R. W. G. Hunt, *Col. Res. & Appl.*, **23**, 138–146, (1998).
3. J. Morovic and M. R. Luo, *Proc. AIC Color 97 Kyoto*, **2**, 594–597, (1997).
4. J. Morovic and M. R. Luo, *Proc. 5th IS&T/SID Col. Img. Conf.*, 44–49, (1997).
5. J. Morovic and M. R. Luo, *Proc. ICPS Conf.*, Antwerp, Belgium, (1998).
6. D. L. Post and C. S. Calhoun, *Col. Res. & Appl.*, **14**, 172–186, (1989).
7. J. Morovic and M. R. Luo, *Proc. 4th IS&T/SID Col. Img. Conf.*, 70–74, (1996).
8. L. L. Thurstone, *Psychological Review*, **34**, 273–286, (1927).
9. F. Ebner and M. D. Fairchild, *SPIE Proc.*, **3300**, 107–117, (1998).