An Evaluation of Colour Models' Performance Using Image Compression Algorithms

Shao Ying Zhu and M. Ronnier Luo
Colour & Imaging Institute, University of Derby
Derby, United Kingdom

Abstract

Five colour models were evaluated using the JPEG and uniform quantization algorithms. They form two categories: linear models including XYZ and YIQ, and non-linear models including CIELAB and CIELUV uniform colour spaces, and CIECAM97s color appearance model. The results show that all non-linear models performed better than those linear models, and not much different between non-linear models.

Introduction

In colour imaging industry, colour spaces and appearance models are used for processing, transmitting and storing colour images for colour management purposes. Image compression algorithms are used for reducing the sizes of images for practical applications. It has been well established in the previous studies[1,2,3,4] that image compression algorithms are colour model dependent. The choice of colour models has a large impact on the performance of image compression techniques.[4] This research work is aimed to test colour models' performance in terms of colour fidelity and image quality using two image compression methods: the Joint Photographic Experts Group's (JPEG) image compression algorithm[5,6] and quantization technique. Five colour models were investigated in this study: two linear models including XYZ[7] and YIQ,[8] and CIELAB,[7] CIELUV[7] and CIECAM97s[9] for non-linear models. The CIECAM97s is the colour appearance model adopted by the CIE in 1997, which can accurately predict a number of colour attributes over a wide range of viewing conditions. Psychophysical experiments were conducted to evaluate colour models' performance using the category judgement method based on a panel of ten observers. The results show that the non-linear colour models performed better than those of linear ones. There is no large difference between non-linear colour models.

Experimental Setup

The images used in this study were divided into two groups: 1) photo-realistic and 2) colour patch images. These are shown in Fig. 1.

Group 1 photo-realistic images:

Doll (top left), Musician (top right), Nature (bottom left), Ski (bottom right)

Group 2 colour patch images

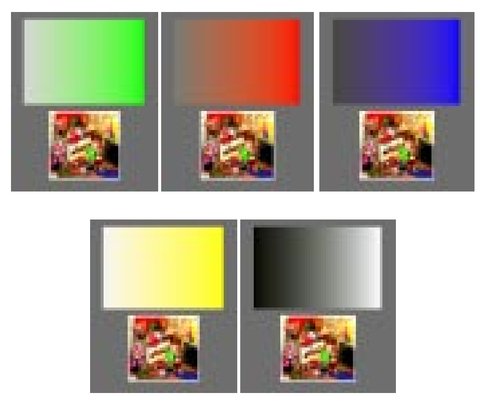

Figure 1. Experimental Images

Each image had a constant physical size, 16 by 16 cm square (580x580 pixels) with a viewing distant about 60 cm. All images were 24-bit RGB images. Group1 images were the same as those used by Morovic[14]. The colour patch images include two parts: a colour patch and a 'Doll' image (the same as that used in Group 1 except that it is smaller). This makes the overall appearance of Group 2 images appeared as a complex image. All colour patch images had a grey background with an L* value of 50. The colour patch areas are the CIELAB chroma scales ranged from zero to the maximum having the same L* and h values of four colours: red, yellow, green and blue. They were the monitor's three primaries and a mixture of the red and green (yellow). The other Group 2 image had a grey scale (neutral colours) instead of a chroma scale. There were over 450 colours in the colour patch area with equally spaced C* or L* values. Group 2 images were intended to investigate each colour model's performance in different colour areas along a fixed direction.

The colour models selected in this study were divided into two groups: linear and non-linear models. Note that the definitions of these are based on the transformation from the CIE tristimulus values (XYZ). The former group includes XYZ and YIQ models. The latter group includes CIELAB and CIELUV uniform colour spaces and the CIECAM97s colour appearance model. The CIECAM97s is the colour appearance model adopted by CIE in 1997. The CIELAB and CIELUV colour spaces are most widely used uniform colour spaces. The YIQ colour space is used for transmitting TV signals and is a linear transform of XYZ space[8].

The images were displayed on a Barco Reference Calibrator with the white point set to D65. The GOG model [10,11] was implemented to transform between the monitor digital counts and luminance. This procedure characterised the monitor's RGB data, a device dependent space, to the CIE tristimulus values, a device independent space. A total of 10 observers took part in the experiment. All observers passed the Ishihara Test [12] for colour blindness. They had some experience for performing psychophysical experiments. The whole experiment was divided into 3 phases. The experimental conditions are summarised in Table 1.

In Phase 1, four Group 1 images were used as shown in Fig. 1. Each image was compressed to 4 different levels for each of the 5 colour models using JPEG algorithm. This resulted in a total of 80 images. These were judged by a panel of 10 observers. Each observer repeated the same experiment twice.

Each image was processed according to the procedure described in Fig. 2. Each pixel of an image was first transformed to the XYZ colour space via a forward monitor model. For all models except the XYZ model, the XYZ values were then transformed to L, C_1 and C_2 attributes, such as lightness, redness-greenness and yellowness-blueness. (The L used here represents the luminance channel and two chrominance channels represented by C_1 and C_2.) These signals were normalised to the range of 0-255 prior to the JPEG compression, in which the coefficients used in the quantization tables and huffman encoding tables were the same as those suggested by JPEG [6]. The lightness of non-linear colour models was quantized using the luminance quantization table and the other two channels were quantized using the chrominance table. The decompressed visual attributes were transformed to the R'G'B' signals for displaying on the Barco monitor via the reverse colour model and reverse monitor characterization model. All models' transformations were carried out using floating point.

Table 1 Summary of the experimental conditions for each phase.

Phase	Image Group (No. of images)	Compression Methods	No of Compre-ssion Levels	Repe-tition	Obs	Total images	Total observation
1	1 (4)	JPEG	4	2	10	80	1600
2	1 (4)	quantization	4	1	10	80	800
3	2 (5)	quantization	4	1	10	100	1000

(Note: Five levels of quantization levels will be described later for each image.)

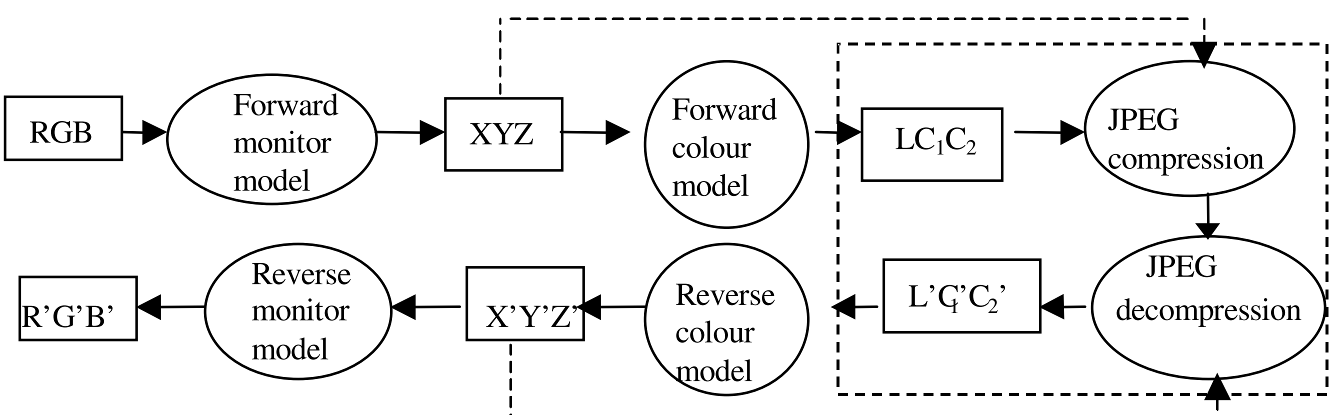

Figure 2. Compression procedure

Table 2 Bit per pixel and compression ratio for each image at each compression levels

	Bit Per Pixel				Compression Ratio			
Quality Factor (applied on CIECAM97s)	**Doll**	**Mus**	**Nat**	**Ski**	**Doll**	**Mus**	**Nat**	**Ski**
15	0.88	0.60	0.58	0.78	27	40	41	31
35	1.24	0.84	0.86	1.07	19	29	28	22
55	1.64	1.10	1.18	1.40	15	22	20	17
75	2.31	1.53	1.72	1.94	10	16	14	12

For each colour model, each image was compressed to 4 different levels, or quality factors. For a lower quality factor, a higher compression ratio was applied. The 4 levels of compressed images using CIECAM97s were first chosen as standard images. These had quality factors of 75 (the default value for JPEG), 55, 35 and 15. The image with a quality factor of 15 was the lowest quality image acceptable by the experimenter. The other two levels were selected having equal interval of 20 between 15 and 75. The images of the other colour models were set the same physical file size as those of the corresponding CIECAM97s images. The bit per pixel for each image in 4 compression level is given in Table 2.

The category judgement method was used for evaluating models' performance. A pair of images was displayed on the Barco monitor in a darkened room. The experimental condition is shown in Fig. 3. One image is the original (uncompressed) image and another is the test image. Observers were asked to make judgements using two scales: colour fidelity and image quality by applying the selection bars on the bottom left and right of display respectively (see Fig. 3). The colour fidelity judgement considered the closeness of colour match between the test and original images. The image quality judgement considers the overall image quality including colour fidelity, artifact, sharpness, etc. Each scale had 7 categories from 1 (the worst) to 7 (the best). The middle category, 4, represents the acceptable level.

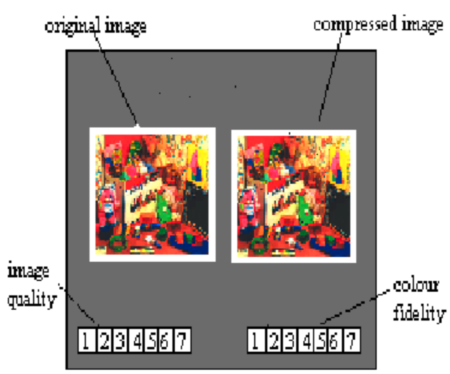

Figure 3. Experiment viewing condition

Phase 2 experiment was the same as that of Phase 1 except that the uniform quantization technique was used for image compression instead of the JPEG algorithm. The attributes for each colour model were normalised within the range of 0 to 255 before quantization. Each image was again quantized to 4 levels. The degree of quantization for each channel was different for each image, but the same for each model. Different quantization levels applied to different images as shown in Table 3. Since human eyes are less sensitive to chrominance channel than luminance channel[13], more bits for luminance channel were preserved than for the chrominance channels. Two chrominance channels were quantized to the same bits.

For selecting quantization levels for each model, the lowest bit rate that the image can be quantized without obvious artefacts was first chosen. The other levels had a higher bit rate than the first chosen one. The selected quantization levels for each image is given in Table 3. For example. the lowest quantization level for the 'Doll' image is 6_5_5 which represents 6, 5 and 5 bits were used for the luminance and two chrominance channels respectively.

Table 3. Quantization levels for each image

Image	Quantization levels (Luminance_chrominance_chrominance)
Doll	6 _5_5 , 6_6_6, 7_5_5, 7_6_6
Nature	6_4_4, 6_6_6, 7_4_4, 7_6_6
Musician	6_6_6, 6_7_7, 7_6_6, 7_7_7
Ski	6_5_5, 6_6_6, 7_5_5, 7_6_6

Group 2 images were used in Phase 3. The experimental conditions were the same as those of the other two phases. Four quantization levels were selected: 7_6_6, 7_7_7, 8_6_6 and 8_7_7 for each image across all colour models.

Results

Observer Variations

Observer variations were investigated using each observer's raw category data. The CV measure given in eq. (1) was used. For examining observer accuracy, the CV value was calculated between the mean and each individual observer's results. For examining observer repeatability, only Phase 1 data were used. (Each observer was repeated twice only in Phase 1 experiment). The CV value was calculated between each individual observer's two sessions.

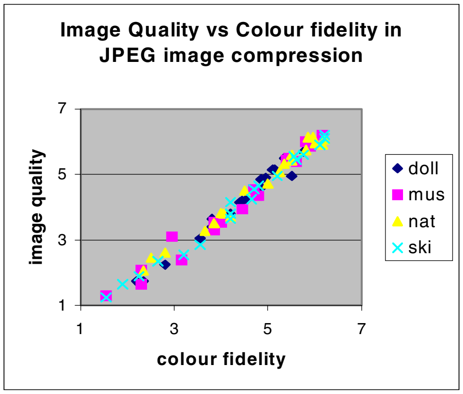

Figure 4. Image quality vs. colour fidelity

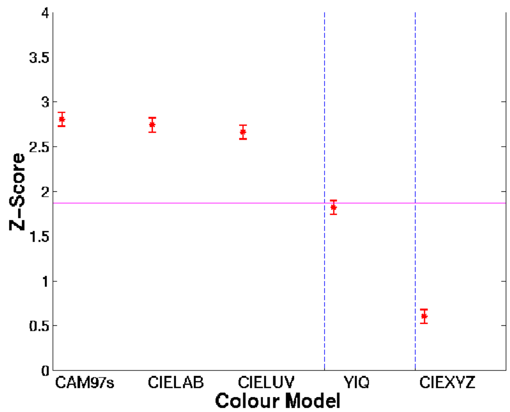

Figure 5. Summary of Phase 1 results for each colour model.

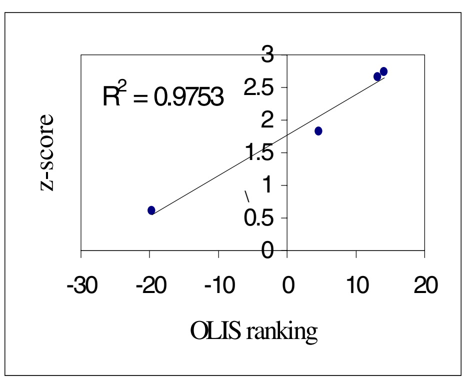

Figure 6. The present finding against Moroney's finding.

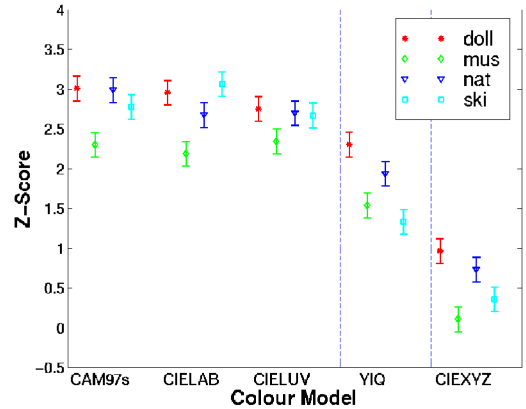

Figure 7. Summary of Phase 1 results for each model/image.

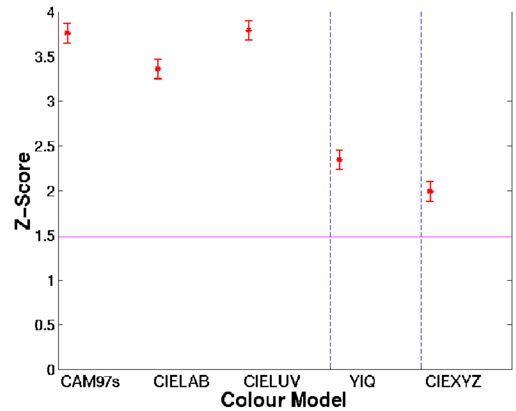

Figure 8. Summary of Phase 2 results for each colour model.

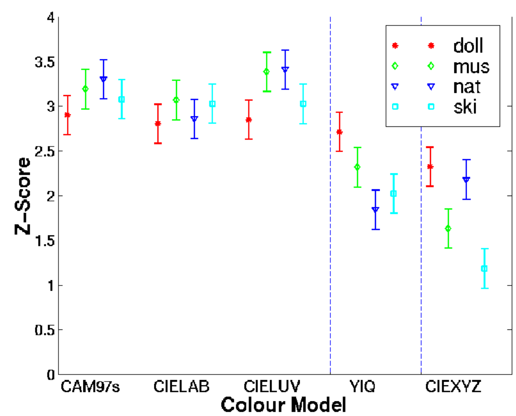

Figure 9. Summary of Phase 2 results for each model/image.

$$CV = \frac{100\sqrt{\frac{\sum (x_i - y_i)^2}{n}}}{\bar{y}} \quad (1)$$

The CV value can be considered as the percentage error between two sets of data. For perfect agreement, the value should be zero. A CV value of 30 means a 30% disagreement between two sets of data. The results show that the mean CV value for repeatability study was 6%, which is considered to be highly repeatable. The CV values for accuracy were 17%, 24% and 25% for Phases 1, 2 and 3 respectively. It was calculated between each individual observer's and mean visual data. These represent the typical accuracy performance in this kind of experiments. The accuracy for Phase 1 is higher than the other two phases. This is because the differences in terms of colour fidelity (or image quality) for individual images are much larger in Phase 1 than the other two phases.

Comparing between the Colour Fidelity and Total Image Quality Results

The mean category results from the colour fidelity and from the image quality judgements for Phase 1 are plotted in Fig. 4 to represent the typical results from all experimental phases. It can be seen that there is a good agreement between the colour fidelity and image quality results. This implies that colour fidelity plays an important role for determining the image quality. The image quality results are used in the following data analysis.

Phase 1

The data analysis was carried out to transform raw category data to z-score. A higher z-score means a better colour fidelity or image quality for a particular colour model. The results are summarised in Figs. 5 and 7 for each colour model. The 95% confidence intervals are also drawn to compare the significant difference amongst colour models. In addition, a horizontal line is plotted (see Fig. 5) to indicate the level of acceptance, which corresponds to Category 4. A point in Fig. 5 represents the performance for each colour model by combining the responses from all four images, i.e. 320 responses for each model. The results show that CIECAM97s and CIELAB performed the best, followed by CIELUV, YIQ with XYZ the worst. This indicates that non-linear models outperformed the linear model by a large margin, and the linear models are not suitable for using JPEG image compression (see acceptance level in Fig. 5). The CIECAM97s recommended by CIE for image applications and CIELAB gave an overall best performance in terms of mean z-score.

The present results were compared with those found by Moroney.[4] Fig. 6 shows that the present z-score results for CIELAB, CIELUV, YIQ and XYZ are plotted with the overall logistic interval scale (OLIS) rankings in Moroney's study. The R^2 value between these two results were 0.975 which means the present results are in good agreement with those found by Moroney.

In Fig. 7, the results of the four images for each colour model are summarized. It shows a consistent trend that all models performed the worst for the 'Musician' image. The linear models gave a larger range of z-score than those of the non-linear models. This implies that the linear models are more image dependent.

Phase 2

The same data analysis was carried out for Phase 2 results. These are summarised in Figs. 8 and 9 for each color model and each model/image respectively. The results are very similar to those of Phase 1, i.e. the non-linear models outperformed the linear models. The CIELUV and CIECAM97s performed the best, followed by CIELAB, YIQ with XYZ the worst. There is no significant difference between CIECAM97s and CIELUV. All models are above acceptance level in this experiment. This is mainly because the reduction of bit rates using the quantization technique is significantly less than those using JPEG compression (Phase 1).

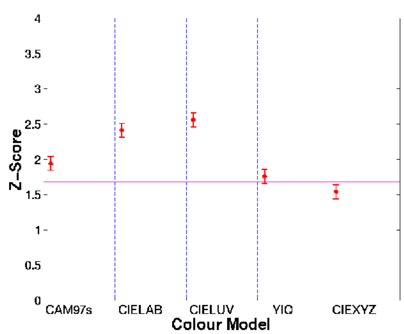

Figure 10. Summary of Phase 3 results for each colour model.

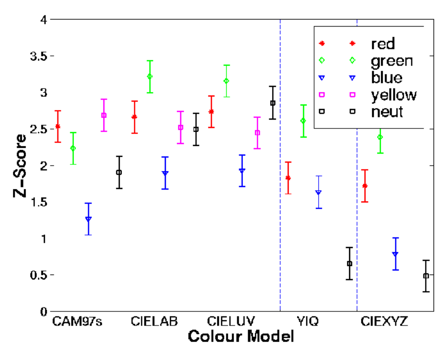

Figure 11. Summary of Phase 3 results for each model/image.

Figure 9 shows the performance of colour models using each individual image. Again, the results confirmed with those found in Phase 1 that the linear models are more image dependent than the non-linear models. It can be seen that the CIECAM97s and CIELAB are the least image dependent models than the other models.

Phase 3

The results from Phase 3 are summarised in Figs. 10 and 11. It can be seen that CIELUV perform the best, followed by CIELAB, CIECAM97s, YIQ with XYZ the worst. There is no significant difference between the CIELAB and CIELUV. However they both performed better than CIECAM97s.

Again, the results in Fig.11 show that the linear models are more image dependent than those non-linear models. I.e. the results for images in each linear model have a much wider spread than in the non-linear models. The CIECAM97s gave a disappointing performance in applying this kind of images. Further investigation will be carried out to investigate the causes of the discrepancy.

Conclusion

This experiment was conducted to investigate the performance of five colour models in terms of colour fidelity and overall image quality. Two image compression techniques were used: JPEG and quantization. Ten normal colour vision observers took part in the experiment using a categorical judgement method. Each observer was asked to judge the compressed image against the original image using a seven-category scale.

The observer accuracy and repeatability were examined. The visual results are considered to be quite reliable. The results from three phases show that the non-linear models (CIECAM97s, CIELAB and CIELUV) outperformed linear models (YIQ and XYZ). In addition, the former models are less image dependent than the latter models. In comparison of three non-linear models, the results show that CIECAM97s performed slightly better than CIELAB and CIELUV for Phases 1 and 2 results. However, it did not perform well for Phase 3 results. This will be further investigated. The present finding shows that the image compression techniques can be used to test colour models' performance.

References

1. A. Drukarev, Compression-related Properties of Color Spaces, *SPIE*, vol. **3024**, (1997)
2. R. E. Van Dyck and S. A. Rajala, Sensitivity to Color Errors Itroduced by Processing in Different Color Spaces, *IEEE on VSPC'91*, (1991)
3. R. S. Gentile, J. P.Allebach and E. Walowit, Quantization of Color Images Based on Uniform Color Spaces, *Journal of Imaging Technology*, vol. **16**, no. 1, pp11-20, (1990)
4. N. Moroney and M. Fairchild, Color Space selection for JPEG Image Compression, *Journal of Electronic Imaging*, vol.**4**, pp373-381, (1995)
5. G. K.Wallace, The JPEG Still Picture Compression Standard, *Communications of the ACM*, vol.**24**, no.4,pp.30-44, (1991)
6. W. B. Pennebaker, J. L. Mitchell, *JPEG Still Image Data Compression Standard*, Van Nostrand Reinhold, (1993)
7. *Colorimetry*, Second edition, CIE Publ. No.15.2, Central Bureau of the CIE, Paris, (1986)
8. G. Buchsbaum, Colour Signal Coding: Color Vision and Color Television, Color Res. Appl., vol **12**, no.5, October , pp266-269, (1987)
9. M. R. Luo, R. W. G. Hunt, The Structure of the CIE 1997 Colour Appearance Model (CIECAM97s), *Col. Res. App*. **23**, pp.138-146, (1998)
10. R. S. Berns, M. E. Gorzynski, and R.J.Motta. CRT colorimetry. Part I: Theory and Practice, *Col. Res. Appl*., vol. **18**, 299-314, (1993)
11. R. S. Berns, M. E. Gorzynski, and R. J. Motta. CRT colorimetry. Part II: Metrology, *Col. Res. Appl*., vol. **18**, 315-325, (1993)
12. S. Ishihara, The Series of Plates Designed as a Test for Colour-blindness, Kanehara Shuppan, Tokyo, (1985)
13. R. W. G. Hunt, Why is Black and White so Important in Colour, *CIM'98 conference proceeding*, 26-27 March, University of Derby, (1998)
14. J. Morovic, To develop a Universal Gamut Mapping Algorithm, *PhD. Thesis*, University of Derby, UK, (1998)