



Colour difference acceptability for calibrated digital images

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Yu, D. K. C., & Oulton, D. P. (2000). Colour difference acceptability for calibrated digital images. *Imaging Science Journal*, 48(4), 165-174.

Published in:

Imaging Science Journal

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.





Colour difference acceptability for calibrated digital images

D K C Yu & D P Oulton

To cite this article: D K C Yu & D P Oulton (2000) Colour difference acceptability for calibrated digital images, The Imaging Science Journal, 48:4, 165-176, DOI: [10.1080/13682199.2000.11784358](https://doi.org/10.1080/13682199.2000.11784358)

To link to this article: <http://dx.doi.org/10.1080/13682199.2000.11784358>



Published online: 06 Oct 2016.



Submit your article to this journal [↗](#)



View related articles [↗](#)

Colour difference acceptability for calibrated digital images

D K C Yu* and D P Oulton

Department of Textiles, University of Manchester Institute of Science and Technology, Manchester, UK

Abstract: The aim of this research work is to establish what level of numeric colour difference between two digital images is visually acceptable by average human observers on a calibrated monitor under a fixed set of viewing conditions. A visual ordinal category method is introduced and a description of the experimental design is provided. Results based on over 4000 visual and numeric comparisons are reported. All results are analysed by statistical methods. The correlation between visual assessments and numerical assessments are found by means of the Pearson product moment coefficient. Overall colorimetry accuracy and metamerism are discussed.

A very high level of correlation is found to exist between the visual ordinal categories assigned by observers and the equivalent numeric (ΔE CMC 2:1) colour difference. Individual observers, within the group of 20 observers tested, were found to be consistent in their ordinal categories. However, some observers were consistently severe in their judgement and others significantly more lenient.

Keywords: colour difference, observer, metamerism, visual category

1 INTRODUCTION

The rapid development of digital camera technology enables a wide range of imaging applications, such as traffic control, remote surveillance of industrial areas, biomedical studies, tracking a customer's movement in a department store, fish selection and processing, and food appearance measurement in agriculture [1]. The information generated by these new imaging applications depends on the quality of the colour reproduction systems used to display the output. Digital cameras are developing in their capability to handle increasingly large image size and a wider range of file formats, with higher throughput and image quality. The need for better image quality leads to, but is not limited to, higher requirements of colour accuracy and repeatability.

Digital colour imaging is a complex process, which

is affected by many factors such as human vision, colour appearance phenomena, imaging technology, device characteristics and media properties. The criterion of image quality addressed in this paper is colour consistency, based on colorimetric equivalence of imaged object and reproduced colour. This requires a device-independent encoding standard, properly calibrated colour-imaging devices and a colour conversion engine to convert device input to device output. Although the criterion may not be met in all situations, this restriction is thought necessary in order to achieve a generally higher standard of image reproduction.

Research on perception of colour differences in images has been most often based on the use of CRT monitors as the colour reproduction medium. Stokes related the nature of the changes to alterations that are likely to be introduced by CRT displays [2]. The mathematical operations of addition or multiplication can be related to common characteristics of devices, such as colour shifts or gain. CRT displays are well simulated by analytical models, in which shift, gain, gamma and contrast are relevant parameters [3]. The

The MS was received on 28 September 2000 and was accepted after revision for publication on 18 April 2001.

** Corresponding author: Department of Textiles, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK.*

acceptability and/or perceptibility of colour difference depend on the size of the objects, on the geometry and colour of the environment, and on the spectral composition and intensity of the light. Human visual response characteristics are also a key factor. The properties of the human response also introduce the complicating dimension of conditional visual equivalence or metamerism.

The selection of test colours for validating any given calibration is very important. They should be representative of the range of colours that are being investigated, but should also include standard reference colours. For this reason, all calibration methods are initially tested with the *de facto* standard Macbeth ColorChecker Chart [4]. They are further tested with customized colour cards, which contain 45 colours, created by the UMIST colour research group in order to classify their metameric effect. More details of metameric definition can be found in reference [5].

Visual experiments have been carried out to establish what level of numeric colour difference (single patch, and average) between two digital images is visually acceptable by average human observers on a calibrated monitor under a fixed set of viewing conditions. An image shown on a calibrated monitor, based on measured colour coordinates, is compared with a calibrated image derived from camera RGB (red–green–blue), converted to a CIE coordinate response by the respective calibration method [6]. All colour-matching experiments are based on a D65 illumination environment, which is created by ImageMaster [7].

A total of 20 observers with normal colour vision have completed the experiments based on the Macbeth ColorChecker (24 colours). In addition, 15 observers finished the experiments based on the custom colour cards (45 colours), as given in Appendix 1.

2 MONITOR SCREEN CALIBRATION

Without calibration, colours displayed by computer monitors are indeterminate and device specific. Typically, the colours seen relate approximately to the CRT (cathode ray tube) RGB values. Because of the intrinsic variation, two monitors (even samples of an identical model) will rarely have the same visual colour when used to reproduce a given set of RGB coordinates.

Colour calibration is handled by the UMIST Adaptive Driver system [8], which maintains a

dynamic mapping between screen RGB drive values and CIE XYZ coordinates, based on a feedback measurement of screen colour using a Minolta CA100 colour analyser under system control. The CRT analyser is used to feed displayed CIE coordinates back to the calibration software, allowing a unique mapping to be built between independent CIE coordinates and device RGB coordinates. A typical calibration process takes approximately 15 min and uses 18 main calibration points out of the infinite CIE gamut.

Long-term exhaustive trials have shown that across the monitor gamut the system is capable of reproducing CIE colour specifications to within an average of 0.5 ΔE CMC (2:1). The largest errors occur at the gamut limits, where gun quantization is at its greatest. Table 1 shows typical calibration results of the monitor used for all visual experiments. Results are given in both CIE Lab and CMC (2:1) colour difference measures.

3 VISUAL EXPERIMENTS

All colour-matching experiments are based on a D65 illumination environment, which is created by ImageMaster. The screen layout is shown in Fig. 1.

Observers were given an experiment brief as given in Appendix 2 and confirmed they understood the classification by ordinal category before doing the experiment. Each observer had to give a category of 4 (perfect match), 3 (near match), 2 (poor match), 1 (bad match) or 0 (no visual correspondence) for each pair under assessment. The experiment was conducted in a darkened room on a recently recalibrated monitor, after checking the numeric accuracy of screen reproduction. All observers were requested to perform the colour deficiency test [9] before doing any experiment.

No speed limit was applied during the visual test.

Table 1 Monitor screen calibration results

Colour groups	CIE Lab		CMC (2:1)	
	Maximum difference	Average difference	Maximum difference	Average difference
1. Munsell greys	1.090	0.466	1.679	0.643
2. Pale shades	1.611	0.960	1.221	0.652
3. Medium shades	0.938	0.569	0.700	0.377
4. Dark shades	1.140	0.656	0.960	0.517
Overall results	1.611	0.655	1.679	0.545

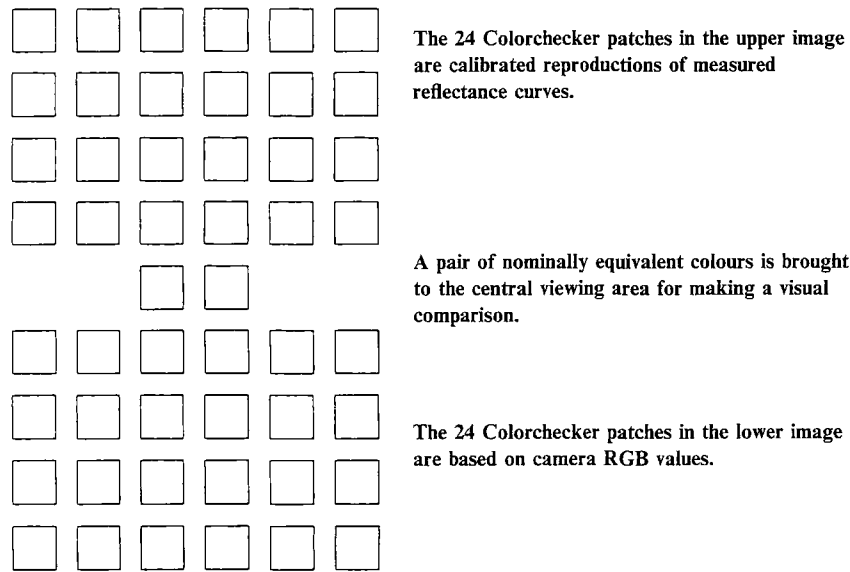


Fig. 1 Screen layout for ImageMaster

It was, however, found necessary to allow a 5 min break for every half an hour looking at the screen. The standard procedure is to do the 24 colours of the Macbeth ColorChecker first and then the 45 colours of the self-design colour cards. However, the observer was required to finish each group of test each time.

3.1 Visual ordinal category results based on Macbeth ColorChecker (24 colours)

An example of overall ranking distribution and cumulative frequency distribution can be found at the end of this section. The full data set is available from the authors. In the following analysis, this large data set is summarized. The correlation between each category and the corresponding average colour difference can be found by means of the Pearson product moment coefficient formula [10]:

Pearson correlation coefficient,

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} \quad (1)$$

where n is the number of input data.

The ordinal categorization is now related to the scalar measure of colour difference. Table 2 is based on 4800 ($24 \times 20 \times 10$) numeric versus visual ordinal categories across all colours. It shows the input data used to calculate the Pearson coefficient. This

Table 2 Pearson coefficient

Category	Average CMC (2:1)
4	1.29
3	2.46
2	4.83
1	6.70
Pearson coefficient	-0.9925

coefficient shows that the category system has a high inverse correlation to the average colour difference CMC (2:1). The negative sign means that the observer gives a higher ranking when the average colour difference is decreasing.

If colour differences X_1, X_2, \dots, X_k occur with frequencies f_1, f_2, \dots, f_k respectively, the standard deviation of colour differences in the CMC (2:1) conformity system of each ranking can be calculated using the following formulae [11]:

$$s = \sqrt{\frac{\sum_{j=1}^k f_j (X_j - \bar{X})^2}{N}} \quad \text{where } N = \sum_{j=1}^k f_j \quad (2)$$

Table 3 shows the standard deviation (SD) of each ranking based on equation (2).

For normal distributions, the 0.95 probability of the measurements are included between $\bar{X} - 2s$ and $\bar{X} + 2s$ [12]. The full set of ordinal category assignments by observers are used to avoid undue influence of outlying parts on a typical assignment. The method

Table 3 Standard deviation

Category	SD CMC (2:1)
4	1.13
3	1.87
2	2.51
1	2.73

used is to reject those assignments that lie outside the 95th percentile of the data set. The histograms in Fig. 2 demonstrate the distribution of assigned categories for category 4, i.e. the perfect match.

3.2 Results based on the UMIST test set of alternative colour cards (45 colours)

An example of the overall ordinal category distribution and cumulative frequency distribution can be found at the end of this section. Similarly, the correlation between each category and the corresponding average colour difference can be found by using the Pearson correlation formula as above. The results are listed in Table 4, based on 2025 ($45 \times 15 \times 3$) numeric versus visual ordinal categories across all colours.

The correlation result for 45 colours is very close to the result for 24 colours. This shows that the category system is also highly correlated as measured against the average colour difference in CMC (2:1) colour difference units. For normal distributions, the 0.95 probability of the measurements occurs between $\bar{X} - 2s$ and $\bar{X} + 2s$. Histograms for the 95th percentile of category are shown in Fig. 3. Table 5 shows the standard deviation of each category based on equation (2).

4 RESULT ANALYSIS

Tests of significance can be used to validate ranking classifications. In practice a level of significance of 0.05 or 0.01 is customary, although other values are used. Table 6 gives critical values of the calculated z score for four levels of confidence [11].

Let \bar{X}_1 and \bar{X}_2 be the sample means obtained in samples of sizes N_1 and N_2 drawn from respective populations having means μ_1 and μ_2 and standard deviations σ_1 and σ_2 . Consider the null hypothesis that there is no difference between the population means. The test statistic z score can be calculated

using the following formulae:

$$\mu_{\bar{x}_1 - \bar{x}_2} = 0 \quad (3)$$

and

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \quad (4)$$

Thus

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{x}_1 - \bar{x}_2}} \quad (5)$$

Based on these equations, the following rule is formulated to test the level of significance:

1. Reject the null hypothesis at a 0.05 level (0.95 probability) of significance if the z score lies outside the range -1.96 – 1.96 . This is equivalent to saying that the observed sample statistic is significant at the 0.05 level.
2. Accept the null hypothesis otherwise.

Table 7 shows the results of tests to find the significance based on data from Tables 2 and 3. Table 8 shows the results of tests of significance based on data from Tables 4 and 5. From Tables 7 and 8, it can be stated at 0.998 probability confidence that the ordinal category classification is significant in all experiments using ColorCheckers (24 colours) and UMIST colour cards (45 colours).

Tables 9 to 11 show the results of additional colours in calibration optimization. These tables demonstrate that even by using more colours for validating a given calibration, no significantly better results are obtained.

5 DISCUSSION OF RESULTS

It is important to emphasize in the analysis of the results that there is inherent variability, due to the existence of both intra- and inter-observer inconsistencies. Cumulative acceptability of observers and colour for single-patch pair comparisons has been addressed in the first part of this paper. Cumulative acceptability of the composite image is now addressed. Image content is standardized and limited to a specific and constant colour set.

An overall high correlation between the ordinal observer category and the scalar measure of colour difference has been previously established. The average category of a composite image is now derived from the set of individual patch comparisons and used

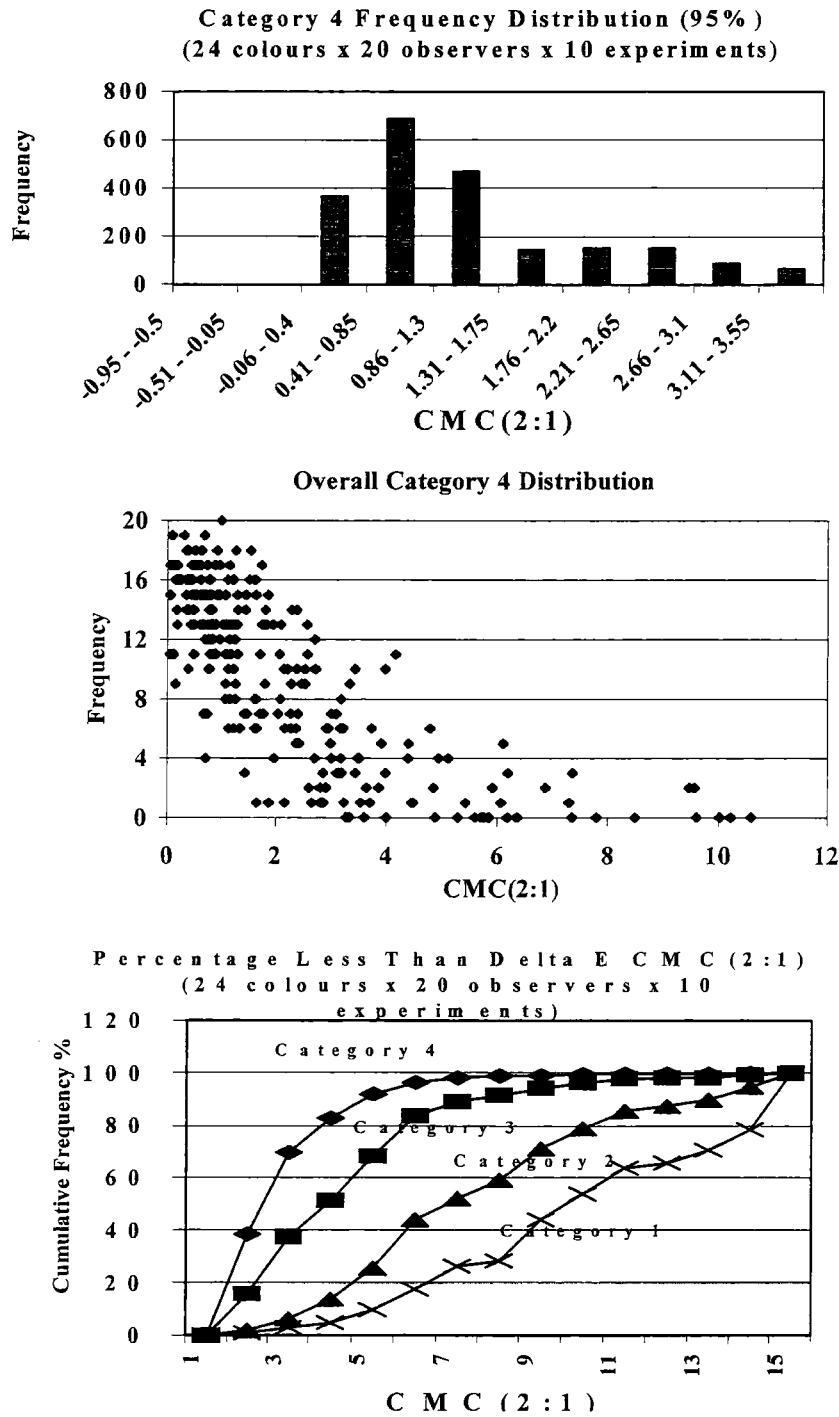


Fig. 2 Results based on Macbeth ColorChecker (24 colours)

as a measure of overall image ‘acceptability’. The chosen measure is a percentage, whereby 100 per cent would indicate a perfect score—a category of 4 for each individual patch colour pairing. It is calculated as follows:

Percentage ordinal category

$$= \frac{\text{total score for composite image components (sum of all colours in category values)}}{\text{maximum score (4 \times total number of colours)}}$$

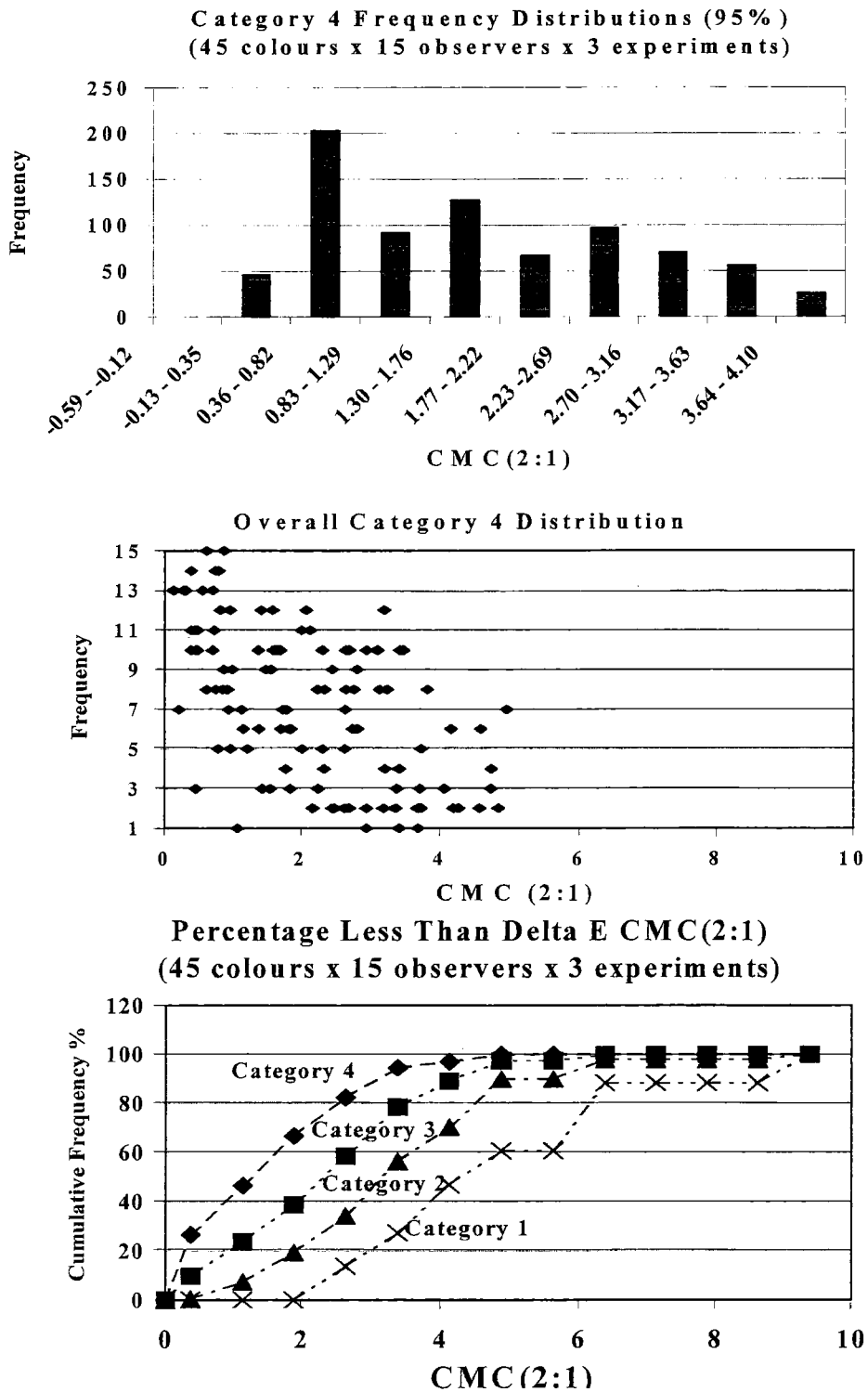


Fig. 3 Results based on UMIST test set (45 colours)

This is expressed as a percentage (e.g. 82.5 per cent) or as a fraction to three decimal places (e.g. 0.825). As a category of 3 equates to a 'near match', the tolerance for visual ordinal acceptability can be set

equal to, or greater than, 75 per cent. That is better than a 'near match' on average.

Observers were limited to making specific judgements of individual patch matching and

Table 4 Pearson coefficient

Category	Average CMC (2:1)
4	1.76
3	2.68
2	3.65
1	5.12
Pearson coefficient	-0.9931

Table 5 Standard deviation

Category	SD CMC (2:1)
4	1.17
3	1.43
2	1.60
1	1.99

were not asked for an overall judgement of visual effect.

5.1 Intra- and inter-observer consistency

As given in Appendix 3, it was found that each observer had his or her own (generally consistent) idea of acceptability. For example, observer 4 in ColorChecker

experiments and observer 3 in UMIST colour card experiments were the same person, who gave the lowest scores across the board for each calibration.

Observer 20 in ColorChecker experiments and observer 15 in self-design colour card experiments were the same person, who gave the highest scores across the board for each calibration. The above two examples prove that consistency in observer ranking could be established based on a personal hierarchy of the visual colour difference. The inter-observer correlation is slightly lower and at an acceptable level.

5.2 Problematic colours

Certain colours are specifically problematic in an instrument-metameric sense, across the majority of calibrations. A given calibration typically corrects a large majority of colours, but instrument metamerism causes certain colours to be 'misinterpreted'. The effect is well known in general imaging, and results, for example, in a specific dress to be reproduced in the 'wrong' colour, although the image is a good overall reproduction.

The main offenders in the ColorChecker set that highlights this type of instrument metamerism are shown in Table 12. Instrument metamerism causes the cameras consistently to produce an incorrect representation of these 'rogue colours'.

Table 6 Critical values of z

Level of significance α	0.05	0.01	0.005	0.002
Critical values of z for two-tailed tests	-1.96 and 1.96	-2.58 and 2.58	-2.81 and 2.81	-3.08 and 3.08

Table 7 Test of significance based on data from Tables 2 and 3

ColorChecker	Compare category 4 and 3	Compare category 3 and 2	Compare category 2 and 1
z score	-23.80	-39.41	-6.39
Decision at 0.05 level	Difference	Difference	Difference
Decision at 0.01 level	Difference	Difference	Difference
Decision at 0.005 level	Difference	Difference	Difference
Decision at 0.002 level	Difference	Difference	Difference

Table 8 Test of significance based on data from Tables 4 and 5

UMIST colour	Compare category 4 and 3	Compare category 3 and 2	Compare category 2 and 1
z score	-22.49	-9.66	-5.04
Decision at 0.05 level	Difference	Difference	Difference
Decision at 0.01 level	Difference	Difference	Difference
Decision at 0.005 level	Difference	Difference	Difference
Decision at 0.002 level	Difference	Difference	Difference

Table 9 Calibration optimization results (I)

Applied to Macbeth ColorChecker			
Calibration based on	Macbeth (24 colours)	UMIST (45 colours)	All 69 colours
Mean CMC (2:1)	1.71	2.78	1.92
SD CMC (2:1)	1.30	2.04	1.34
Compare	24 colours to 45 colours	45 colours to 69 colours	24 colours to 69 colours
z score	-2.651	2.498	-0.676
Decision at 0.05 level	24 colours better	45 colours better	No difference
Decision at 0.01 level	24 colours better	No difference	No difference
Decision at 0.005 level	No difference	No difference	No difference
Decision at 0.002 level	No difference	No difference	No difference

Table 10 Calibration optimization results (II)

Applied to UMIST colours			
Calibration based on	UMIST (45 colours)	Macbeth (24 colours)	All 69 colours
Mean CMC (2:1)	1.70	2.79	2.01
SD CMC (2:1)	1.12	1.84	1.27
Compare	45 colours to 24 colours	24 colours to 69 colours	45 colours to 69 colours
z score	-2.652	1.923	-1.369
Decision at 0.05 level	45 colours better	No difference	No difference
Decision at 0.01 level	45 colours better	No difference	No difference
Decision at 0.005 level	No difference	No difference	No difference
Decision at 0.002 level	No difference	No difference	No difference

6 CONCLUSIONS

In the overall visual experimental results (all samples, all tests), the Pearson product moment coefficient was found to be -0.9925 based on the Macbeth

ColorChecker (24 colours) and -0.9931 based on the UMIST colour cards (45 colours). The ordinal category method is highly inverse-correlated to the average colour difference of the samples in the tested images. Based on the ordinal category system intro-

Table 11 Calibration of optimization results (III)

Applied to 69 colours (Macbeth ColorChecker + UMIST colours)			
Calibration based on	All 69 colours	UMIST (45 colours)	Macbeth (24 colours)
Mean CMC (2:1)	1.98	2.79	2.01
SD CMC (2:1)	1.28	1.57	1.74
Compare	69 colours to 45 colours	45 colours to 24 colours	69 colours to 24 colours
z score	-0.320	-0.799	-1.110
Decision at 0.05 level	No difference	No difference	No difference
Decision at 0.01 level	No difference	No difference	No difference
Decision at 0.005 level	No difference	No difference	No difference
Decision at 0.002 level	No difference	No difference	No difference

Table 12 List of problematic colours

Colour number	CMC (2:1)			Category		
	Worst	Mean	Best	Worst	Mean	Best
7 (orange)	10.60	4.57	1.07	1.80	2.89	3.80
12 (orange yellow)	7.79	3.09	0.79	2.05	3.23	3.90
16 (yellow)	7.30	2.59	0.15	2.25	3.24	3.80
18 (cyan)	10.23	6.21	0.49	1.70	2.52	3.70
Other correct colours	5.85	2.07	0.05	2.25	3.35	4.00

duced as shown in Appendix 2, the category 3 (a near visually close match) rating, which is the mean of 24 ColorChecker colours, equates to a CIE Lab ΔE 3.0 or CMC (2:1) 1.5, in 4800 tests of paired results.

A colour difference lower than or equal to this value was proved to have high acceptability by all observers. If further visual experiments are carried out in the future, the following recommendations should be considered in order to reduce the standard deviation of the mean experiment percentage classification:

- to increase the number of observers;
- to identify the most reliable observers to repeat the experiments;
- to prevent an over-estimate or under-estimate of the mean percentage category by removing the highest and the lowest scores from observers.

From the example of the overall category distribution in Sections 3.1 and 3.2, the observers are encouraged to reject several sample matches corresponding to low colour differences. Others of a substantially higher colour difference were considered as

acceptable. Therefore, it is inadvisable to use CIE Lab ΔE colour difference values alone as a definitive guide to what is deemed as visually acceptable.

The first camera using the first calibration method achieved mean colour differences for CMC (2:1) of 1.71 and 1.70 based on ColorChecker 24 colours and UMIST 45 colours respectively. If an enlarged data set (ColorChecker 24 colours plus UMIST 45 colours) was used, the mean colour difference for CMC (2:1) increased to 1.98. This showed that using more colours for calibration does not necessarily give a better result.

ACKNOWLEDGEMENTS

The research described arises from the DTI Link Project AFM/65 'Colour Calibration for Food Appearance Measurement'. The work by Dr Pointer and Professor Attridge at the University of Westminster during the project provided some of the background data, and also the camera system design principles used in the modelling process. The authors are also indebted to all staff and students who contributed to the colour matching experiment.

REFERENCES

- Faugeras, O.** In Proceedings of Computer Vision—ECCV 90 First European Conference on *Computer Vision Antibes*, Antibes, France, April 1990.
- Stokes, M.** Colorimetric tolerances of digital images. MSc thesis, RIT, University of Rochester, 1991.
- Berns, R. S.** Methods for characterising CRT displays. *Displays*, 1996, **16**(4), 173–182.

4 McCamy, C. S., Marcus, H. and Davidson, J. G. A color-rendition chart. *J. Appl. Photographic Engng*, 1976, 2(3).

5 Hunt, R. W. G. *Measuring Colour*, 3rd edition, 1998 (Fountain Press, Kingston upon Thames).

6 Oulton, D. P. and Porat, I. The control of colour by using measurement and feedback. *J. Textile Inst.*, 1992, 83(3).

7 Oulton, D. P., Porat, I., Boston, C. and Walsby, R. Imagemaster: precision colour communication based on CIE calibrated monitor screens. In Proceedings of 5th International Conference on *High Technology*, Chiba, Japan, September 1996.

8 Oulton, D. P., Boston, C. J. and Walsby, R. Building a precision colour imaging system. In The Fourth Colour Imaging Conference on *Colour Science, Systems and Applications*, November 1996, pp. 14–19.

9 Ishihara, S. *Tests for Colour-Blindness*, 1954 (Kanehara Shuppan Company Limited).

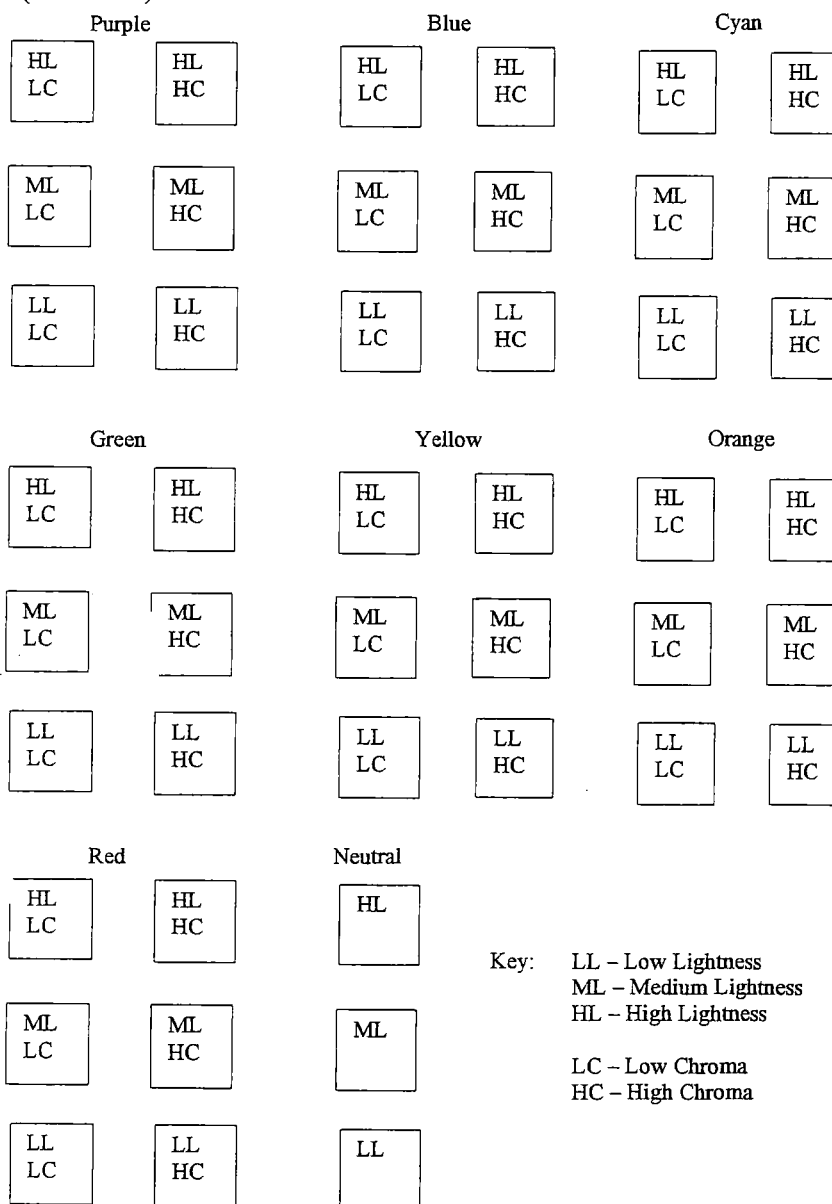
10 Spiegel, M. R. *Theory and Problems of Statistics*, 1980 (McGraw-Hill).

11 Tabachnick, B. G. and Fidell, L. S. *Using Multivariate Statistics*, 3rd edition, 1996 (Harper Collins College Publishers).

12 Keller, G. and Warrack, B. *Statistics for Management and Economics*, 4th edition, 1997 (Duxbury Press).

APPENDIX 1

UMIST colour cards (45 colours)



APPENDIX 2

Experiment brief: the accuracy of imaged colour

The experiment involves visually ranking coloured patches on a computer screen. Images will be shown of the Gretag Macbeth ColorChecker™. This is one of the most photographed colour charts in existence and is used in photography and reprographics as a standard set of colours for measurement and the calibration of reprographic devices.

Colour is modelled as a three-dimensional space, the three dimensions being lightness, chroma and hue. Numerical values can be applied to these values and mathematical transforms produce *X*, *Y* and *Z* values, which allow each of the 16 million or so discernible colours to be pinpointed at a specific point in this colour space. Even very small differences in these values can cause a significant shift in colour space (darker or lighter, more pure in colour or greyer, more blue or more red) which can be perceived by the human visual system.

In the following experiment the acceptability of colour patches is to be ranked. The colours in the top image are those which have been measured from the physical sample. Those in the image below have been produced by calibration of digital camera images. Each set of patches will be shown in turn and each ranking is to be placed in the corresponding chart.

The following system is to be used in the ranking:

- 4: A PERFECT MATCH
There is absolutely no colour difference,
e.g. no difference at all between two blue patches.
- 3: A NEAR MATCH
A discernible difference although the two are very similar,
e.g. a slight difference can be seen between two blue patches.
- 2: A POOR MATCH
A large difference between two patches,
e.g. the coloured patches may still be a similar blue but are very noticeably not a match.
- 1: A BAD MATCH
Great dissimilarity is apparent,
e.g. patches are both essentially blue but not the same blue.
- 0: NO OBVIOUS CORRESPONDENCE TO TARGET COLOUR
e.g. one patch appears blue, the other green, red or purple!

APPENDIX 3

Intra- and inter-observer consistency

Average observer category for each experiment

1. Ten experiments based on 24 colours:

	1	2	3	4	5	6	7	8	9	10	Observer mean percentage category
1	71.9	94.8	78.1	82.3	76	75	77.1	70.8	76	78.1	78.01
2	79.2	96.9	85.4	84.4	84.4	84.4	82.3	79.2	79.2	79.2	83.46
3	80.2	84.4	94.8	90.6	88.5	84.4	78.1	77.1	77.1	80.2	83.54
4	70.8	82.3	74	70.8	70.8	67.8	63.5	58.3	62.5	65.6	68.64
5	76	83.3	72.9	77.1	78.1	70.8	70.8	71.9	71.9	75	74.78
6	79.9	97.9	85.4	82.3	85.4	86.5	85.4	82.3	87.5	89.6	86.22
7	80.2	96.9	78.1	78.1	70.8	68.8	67.7	76	62.5	76	75.51

8	82.3	92.7	81.3	79.2	79.2	72.9	71.9	75	82.3	78.1	79.49
9	85.4	96.9	86.5	86.5	83.3	83.3	75	74	75	75	82.09
10	80.2	99	87.5	84.4	88.5	83.3	81.3	78.1	82.3	81.3	84.59
11	83.3	97.9	92.7	91.7	92.7	90.6	89.6	95.8	90.6	92.7	91.76
12	94.8	97.9	84.4	79.2	81.3	80.2	69.8	70.8	72.9	73.9	80.52
13	78.1	96.9	83.3	85.4	83.3	76	75	80.2	78.1	75	81.13
14	93.8	97.9	90.6	89.6	91.7	90.6	88.5	85.4	91.7	90.6	91.04
15	87.5	94.8	85.4	86.5	83.3	88.5	75	77.1	78.1	81.3	83.75
16	89.6	94.8	87.5	83.3	89.6	87.5	83.3	89.6	89.6	88.5	88.33
17	72.9	85.4	71.9	74	71.9	70.8	69.8	70.8	68.8	74	73.03
18	87.5	94.8	72.9	75	84.4	84.4	74	79.2	77.1	81.3	81.06
19	95.8	100	86.5	93.8	80.2	92.7	91.7	93.8	89.6	94.8	91.89
20	91.7	100	92.7	92.7	96.9	96.9	91.7	95.8	95.8	97.9	95.21
Mean experiment percentage category	83.06	94.28	83.60	83.35	83.02	81.77	78.08	79.06	79.43	81.41	82.70 = Average
Standard deviation	7.45	5.67	6.96	6.50	7.24	8.57	8.32	9.43	9.36	8.34	

2. Three experiments based on 45 colours:

				Observer mean percentage category
	1	2	3	
1	82.2	78.3	64.4	75
2	88.9	77.2	62.2	76.1
3	73.3	68.3	58.3	66.6
4	88.9	83.3	75.6	82.6
5	94.4	91.1	79.4	88.3
6	86.7	76.7	70.6	78
7	91.1	78.3	82.8	84.1
8	71.1	77.2	72.2	73.5
9	83.3	86.1	75.6	81.7
10	92.2	73.9	62.8	76.3
11	86.7	78.9	81.6	82.4
12	89.4	79.4	72.8	80.5
13	92.2	85	78.3	85.2
14	81.1	74.4	57.8	71.1
15	92.2	82.2	85	86.5
Mean experiment percentage category	86.25	79.4	72	79.19 = Average
Standard deviation	6.915	5.57	8.98	