UDC 655.3.024:658.5
ORIGINAL SCIENTIFIC PAPER
RECEIVED: 14-01-2011
ACCEPTED: 08-02-2011

ACTA GRAPHICA 184

Increasing the Accuracy of Colour Reproduction System Evaluation by Proper Sampling

Authors

Davor Donevski*, Diana Milčić, Dubravko Banić

Faculty of Graphic Arts
University of Zagreb, Croatia
**E-mail: davor.donevski@grf.hr*

Abstract:

Test charts are a tool commonly used for the evaluation of colour reproduction system accuracy. They may vary in size and choice of values tested. Within statistical quality control, the importance of random sampling is very often stipulated. However, in the context of the evaluation of colour reproduction system accuracy, samples are very small with respect to the number of possible device inputs. It is therefore very common that practically usable test charts of sizes of a few hundred patches underestimate the maximum error. Including a larger number of critical values such as those at the gamut boundary increases the probability of estimating the maximum error more accurately. However, the sample is no more random and the inclusion of a larger number of critical values inevitably increases the central tendency measures. The aim of this research is to improve the maximum error estimate, while retaining the central tendency estimates within the range of statistically insignificant differences. This is achieved by adding critical values to the chart of sufficient number of patches.

Keywords:

CRS Accuracy, Chart Size, Patch Selection

1. Introduction

Colour reproduction devices have a large number of possible input values. In the case of simple output devices with RGB drivers, i.e. 3 channels, each 8-bit, the number of permutations with repetition equals 256^3 = 16 777 216. In order to estimate the accuracy of colour reproduction system, only a much smaller sample of a few hundred values is practically usable. In statistical sampling it is common to use random samples. However, it is generally known that the sample size (in the context of this work, the number of test chart patches) is important in order to estimate the maximum and central tendency measures more accurately. The important consideration with respect to the central tendency measures is that using larger sample decreases the width of the confidence interval. In the case of the maximum (or minimum) error, it is known that increasing sample size increases the probability of including extreme values from the population. Tools used for the accuracy evaluation of colour reproduction devices are test charts and they can be considered as samples of a given size (number of patches). Various methods have been developed for optimal selection of training set values (cf. *Chou & Luo, 2009*), but little attention was given to test set value selection. In many references, such as *Johnson (2002)*, *Cheung and Westland (2004)*, charts of size 288 patches are considered as sufficiently accurate for the estimation of colour reproduction system accuracy. However, this pertains to the differences in estimated errors between random samples of different sizes. Their estimated errors do not differ significantly in samples of size 288 or larger. Despite this fact, these random samples are still too small to yield a probability high enough to include extremes from the population. The only way to estimate the maximum error more accurately is to include critical values such as those at the gamut boundary. *Green (2001)* showed that the inclusion of gamut boundary values in training targets increases the model accuracy. However, the effect of including such values in test targets was not presented. The inclusion of those values certainly affects central tendency measures. The question still open is to which extent they are affected with respect to the sample size and number of critical values included, and this is covered by this research.

2. Methodology

This research was conducted on a laser printer with a driver accepting RGB input values. The driver converts RGB inputs to device colorant space. As there is no direct access to the device colorant space, the device space consists of RGB values accepted by the device driver. The device was characterized using a 918 patches characterization chart and creating a standard ICC device profile. Although the inverse transformation ($L^*a^*b^*$ to RGB) accuracy is of more practical importance, the forward transformation (RGB to $L^*a^*b^*$) accuracy was tested for practical reasons, and with respect to the fact that the forward and inverse transformations are generally similarly accurate. In order to inspect the differences between different samples and sample sizes, three random samples for each of the three different sizes were created from a pool of RGB input values, totalling in 9 samples. Many authors, such as *Green (2002)*, stipulate the importance of using independent test set, and these were used in this research. The samples were printed on the laser printer and its responses were measured using a spectrophotometer. The conditions were illumination D50, 2° observer and 45°/0° measuring geometry. Measured $L^*a^*b^*$ values were compared to the values obtained by transforming RGB charts to $L^*a^*b^*$ space using the device ICC profile. Values were compared in terms of the colorimetric ΔE difference. After having determined the differences between different random samples of the same size and differences between different sample sizes, the effect of including critical values was tested. Samples of different sizes were extended by including additional 18 gamut boundary colours. The values predicted by the colour management system were compared to the measured device responses as previously described.

3. Results and discussion

Table 1 contains evaluation results for three different random samples of size 144. It shows the statistics of ΔE colorimetric differences between measured device responses and values predicted by the colour management system. It is obvious that the central tendency measures do not differ significantly for the three random samples, considering the widths of the 95% confidence intervals. However, the maximum error range equals 3,11 and this can be considered quite large.

Table 1, Estimation results on three different random samples of size 144

# patches	min	median	mean	max	95% C.I.
144	1,90	5,47	5,54	11,35	0,60
144	1,48	5,36	5,68	13,66	0,69
144	1,37	5,27	5,44	10,50	0,60

Table 2 contains evaluation results for three different random samples of size 288. The central tendency measures ranges between the three samples are remarkably larger than in the case of samples of size 144. The confidence intervals are narrower as the sample is larger. Nevertheless, the central tendency measures do not differ significantly for the three random samples. The maximum error range equals 4,21, which is even larger than in the case of samples of size 144.

Table 2, Estimation results on three different random samples of size 288

# patches	min	median	mean	max	95% C.I.
288	1,23	5,39	5,89	17,75	0,57
288	0,78	5,12	5,49	13,54	0,50
288	1,34	5,54	5,95	16,15	0,55

Table 3 contains evaluation results for three different random samples of size 576. In this case the central tendency measures ranges are somewhat smaller then in the case of samples of size 288. Each of the three samples falls within the limits of the 95% confidence interval of other samples and it can be stated that they do not differ significantly. The maximum error range equals 2,92 which is quite large.

Table 3, Estimation results on three different random samples of size 576

# patches	min	median	mean	max	95% C.I.
576	0,84	5,31	5,62	17,66	0,34
576	1,06	5,31	5,80	16,86	0,37
576	0,57	5,66	5,81	14,74	0,33

Table 4 contains evaluation results for 18 gamut boundary values. It is known that device characterization models are least accurate in those regions, and the results in Table 4 clearly show that both, central tendency and maximum errors are much larger than in the case of any of the random samples (*Table 1 to Table 3*).

Table 4, Estimation results on 18 gamut boundary values

# patches	min	median	mean	max	95% C.I.
18	4,68	13,57	13,98	25,07	5,89

Table 5 shows the evaluation results for random samples of different sizes with and without 18 critical values. For the sample size 144, including 18 gamut boundary values resulted in increasing the estimated mean value by almost ΔE 1. The maximum error was increased by ΔE 11,41. These differences are very large. In the case of the sample size 288, the inclusion of 18 addition values increased the mean error by considerably smaller amount then in the case of sample size 144. However, it is still a significant increase which shows that sample size 288+18 did not prove to be sufficiently large in this case. Adding 18 critical values to the sample of size 576 resulted in a considerably smaller increase of mean error than in the case of other two sample sizes. This increase should not be considered significant, which leaves a conclusion that the sample of size 576+18 proved to be sufficiently large to estimate the maximum error accurately, while not significantly affecting the mean error.

Table 5, Evaluation results of random samples with and without additional critical values

Sample size/error	144	144+18	288	288+18	576	576+18
mean	5,68	6,60	5,89	6,37	5,62	5,87
max	13,66	25,07	17,75	25,07	17,66	25,07

4. Conclusion

The results in this research showed that random sampling, often employed in statistical quality control, does not provide sufficiently accurate maximum error estimates in the context of the evaluation of colour reproduction system accuracy. The reason is that practically usable sample sizes are much smaller than the population. It was also shown that adding critical values to the random samples increases the accuracy of the maximum error estimates, but it inevitably affects the central tendency measures by increasing them. The amount of this increase depends on the difference between means of random sample and critical values sample and the sizes of both. These results show that in the particular case sample of 576 values proved to be sufficiently large to retain sufficiently accurate estimate of the mean error when 18 critical values are added to it. However, the number of critical values sufficient to accurately estimate the maximum error was not part of this research and should be investigated separately. It can be stated that the evaluation accuracy can be improved by adding critical values, provided that the sample is sufficiently large. The required sample size depends on the difference between device characterization model performance inside the gamut and on its boundary.

References

CHEUNG, V & WESTLAND, S 2004, 'A comparative study of the characterization of colour cameras by means of neural networks and polynomial transforms', *Journal of Coloration Technology*, vol. 120 no.1 pp. 19-25

CHOU, Y, LI, C & LUO, M 2009 'A New Colour Selection Method for Characterising Digital Cameras', in *proceedings of the Seventeenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Albuquerqe

GREEN, P 2002, 'Overview of characterization methods', in P Green & L MacDonald (ed.), *Colour engineering*, John Wiley & Sons, Chichester

GREEN, P 2001 'A test target for defining media gamut boundaries', in *Imaging: Device-Independent Color, color Hardcopy and Graphic Arts: proceedings of the SPIE conference*, San Jose

JOHNSON, T 2002, 'Methods for characterizing colour scanners and digital cameras', in P Green & L MacDonald (ed.), *Colour engineering*, John Wiley & Sons, Chichester