Affective Attributes in Image Quality of a Mobile LCD

***Youn Jin Kim*[1], *M. Ronnier Luo*[1], *Peter Rhodes*[1], *Wonhee Choe*[2], *Seongdeok Lee*[2] *and Changyeong Kim*[2]**
[1]Colour & Imaging Group, University of Leeds, Leeds, UK
[2]Samsung Advanced Institute of Technology, Yongin, South Korea

Abstract

A psychophysical experiment for describing image quality of a 2-inch QVGA mobile liquid crystal display (LCD) was carried out using a category judgment method. Five natural test images were rendered in terms of 8 physical parameters: peak-white luminance, resolution, bit depth, correlated colour temperature (CCT), lightness with linear and non-linear alterations, chroma and hue. Ten observers rated each of the rendered images using 9 categories (1 to 9), according to 7 perceptual attributes: naturalness, clearness, sharpness, contrast, colourfulness, quality and preference. In total, 18900 judgments were made. The whole set of psychophysical data were used to build an empirical image quality model and, through a stepwise regression method[1], it was found that naturalness and sharpness are the most important attributes quantifying image quality.

Introduction

Image quality has been recognised as one of the top considerations in the display manufacturing industry where they always face a trade-off between quality and cost.[19] The aim is to achieve the highest quality with reasonable cost. This situation requires a metric which can accurately represent the quality of an image accounting for human visual perception. However, defining or evaluating image quality is not simple, although it is innately understood.[2] Image quality can be evaluated physically (objective image quality) or psychophysically (subjective or perceptual image quality). Objective evaluation involves physical measurement of images and generally fails to consider human visual characteristics. Therefore, psychophysical experiment results are required for developing metrics that represent the response from a panel of observers. Subjective image quality modelling process was refined and generalised as image quality circle (IQC)[19] and it has been adopted in industry. The subjective image quality research can be divided into two major approaches: external and internal reference. The former assumes that the image quality of reproductions corresponds to perceptible visual difference from its origin.[3] Thus, the overall procedure of rating image quality is usually based on impairment, e.g. method of limit or pair comparison. A number of these metrics have been suggested and widely used. For example, colour difference between a pair of uniform colour patches can be calculated as the Euclidean distance between the corresponding coordinates in 1976 CIELAB colour space.[4] It is still widely used in industry due to its simplicity, despite significant non-uniformity in blue region.[4] In addition, it is the only colour space recommendation by CIE. There was a collaboration between international leading colour scientists from several countries to develop an accurate colour difference metric for small colour difference which resulted in CIEDE2000 and recommended by CIE.[5,6] Since those traditional metrics were derived for evaluating colour difference of simple uniform colour patches, S-CIELAB was published in 1996 as an image difference metric accounting for the spatial properties in image contexts.[7] This model was refined into a modular framework.[18]

Internal reference image quality can be similarly described as the image quality of a given image corresponding to perceptible visual difference from its memory prototype. The category judgment method is typically used for this approach, in which observers judge a single image by perceptual comparison with a memorised reference, when the original is not present.[7,8] There has been some effort to appraise an image without an original based on information theory[8] and the similarity to the memory colours of sky, grass, and Caucasian skin.[9] This idea was supported by the fact that the appearance of particular memory colours is different from that of original.[10] Fernandez et al.[2] found that observers from 5 different cultural backgrounds generally agreed with each other and even some biases between them were not as visually significant as image contexts or observer variability.

This study aims to investigate which attributes are affective in the image quality of a mobile LCD by means of internal reference evaluations through 18900 psychophysical assessments including repetition.

Image Quality Assumptions

Image quality (IQ) judgment is an event that approximates the overall perception of the image's excellence by human observers. For a given complex colour image, it is assumed that it involves two steps for human observers to appraise IQ. Observers firstly perceive an image that is determined by some physical traits (low-level attributes), such as image statistical measures, display specifications and viewing conditions, and then evaluate individual perceptual (or high-level) attributes such as naturalness, clearness, sharpness, contrastness, colourfulness, and so on. Secondly, the overall quality of the image can then be determined as a function of those attributes. For example, people assess an image under a certain viewing condition and perceive how natural, sharp, or colourful it is, and then form an overall impression as to the quality of the image.

Experimental

Setup

The subjective image quality experiment was performed using the colour display of a Samsung SCH-S250 mobile phone[15] in a dark room. It is 2-inch QVGA sized and its colour gamut is similar to sRGB in CIE 1931 xy chromaticity diagram.[21] A Minolta CS-1000 tele-spectroradiometer (TSR) was used for colour measurement.

The display was characterised using PLCC (Piecewise Linear Interpolation Assuming Constant Chromaticity) method.[11] A 9-equally stepped greyscale was measured and utilised for training a characterization model. With the combinations of 0, 64, 128, 192 and 255, 125 colours were selected to test the characterisation model. The median difference between the model prediction of the test colours and their corresponding measurement was 4.0 of CIELAB colour difference units.

Figure 1. *Test images*

Test Images

The number of test images in this study is five.[12] The contexts consisted of facial skin (Caucasian, Black, and Oriental), sky, grass, water, and fruit colours, as listed in Figure 1. It is assumed that mobile phone users usually view images of facial and natural (sky, grass) scenes under outdoor daylight viewing conditions. Therefore, these colour images were selected in addition to some memory colours such as fruit.

Procedure

Ten PhD students from University of Leeds, who passed the Ishihara test, participated. They were asked to rate each of the displayed images on the mobile LCD from the distance of 25 centimetres in a dark room, using a 9-point scale (1 to 9) for 7 perceptual attributes: naturalness, clearness, sharpness, contrast, colourfulness, quality and preference. All categories were described by a symmetrical design of quantitative adjectives originally suggested by Bartleson[13] and listed in Table 1. Equal-perception intervals were assumed between two consecutive categories. The collected category scale data were analysed in terms of z score following Torgerson's Law of Categorical Judgment.[14, 19]

Table 1. The definition of Bartleson categories[13]

Category	Definition
1	Least imaginable "ness"
2	Very little "ness"
3	Mildly "ness"
4	Moderately "ness"
5	"Ness"
6	Moderately highly "ness"
7	Mildly highly "ness"
8	Very highly "ness"
9	Highest imaginable "ness"

Table 2 lists the rendered levels of each physical parameter and converting functions. Forty-four rendered images, which consist of different levels of 8 physical parameters, were prepared. Fifty images of them were randomly selected and assessed twice. Combined with the 5 distinct test images and 7 perceptual attributes, this requires 1890 observations for each observer. In total, 10 observers made 18900 judgments. Each observer took part in 4 separate sessions.

Observers were trained prior to the main experiment. In each session, they were asked to fully adapt to the viewing conditions in a dark room which took approximately 3 minutes. Each image displayed on a mobile phone was assessed by one observer at one time and they were asked to judge a category in terms of each of the 7 attributes. The sequence of questions was randomised.

Table 2. Levels of physical parameters (O: Output & I: Input)

Physical Parameter	Level	Function
Peak-White Luminance	1.0, 0.8, 0.6, 0.4, 0.2	O = k X I
Resolution (ppi)	200, 180, 160, 140, 120, 100, 80	-
Bit Depth	8, 7, 6, 5, 4	-
CCT (K)	5400, 6500, 9300	-
Lightness linear	1.0, 0.9, 0.8, 0.7, 0.6	O = k X I
Lightness nonlinear	1.0, 1.2, 1.4, 1.6, 1.8	O = 100 X $(I/100)^{a}$
Chroma	1.0, 0.8, 0.6, 0.4, 0.2	O = k X I
Hue (°)	-60, -45, -30, -15, 0, 15, 30, 45, 60	O = Offset + I

Rendering Algorithms

Peak-White Luminance

Peak-white luminance levels of the mobile LCD were altered by means of multiplicative transformation. For each pixel in an image, a corresponding set of CIE XYZ values was obtained using the characterisation model of display. The tristimulus values were then multiplied by 0.8, 0.6, 0.4, and 0.2 to simulate lower levels of the display peak-white.

Resolution

Resolution can affect the ability to distinguish and recognise fine spatial detail. Pixels-per-inch (ppi) is often used to express the resolution of digital images. The resolution of the test images was changed using bicubic re-sampling method via *Adobe Photoshop 7.0* including two steps: adjusting resolution and re-sampling the image.

Bit depth

Bit-depth is the number of bits used to define each pixel. A digital colour image has a bit depth typically ranging from 8 to 24 bits. In most cases the bits are equally divided for each of the three channels (R, G, and B).

$$T = 2^{N} \tag{1}$$

$$I = 2^{MaxBit}/2^{N} - 1 \tag{2}$$

where *MaxBit* represents the number of bits of a display's single channel reproducing the simulated rendered colours or images, i.e. 8 bits in the case of this mobile LCD display. *N* is the target bit depth to be simulated.

Equations 1 and 2 illustrate a bit depth manipulation algorithm for a single channel from 1 to 8 bits (3 to 24 bits for three colour channels). The number of tone levels (*T*) is proportional to the bit-depth (*N*) as a power function and it is expressed in Equation 1. In addition, the interval (*I*) between consecutive levels can be formed following the relationship as Equation 2 to design evenly stepped tones.

Correlated Colour Temperature (CCT)

Since there is no available function to set the CCT in the mobile display, a simulation method was adopted. This involved the modelling of colour appearance change using a 15-inch LCD monitor for which the CCT could be controlled. For each of three CCTs of the LCD: 5400, 6500 and 9300K, a PLCC[11] characterisation model was implemented. Each model gave different XYZ values of images for each CCT setting. The XYZ values were then converted to RGB via the reverse PLCC characterisation model to reproduce images under the 3 CCTs.

Lightness, Chroma, and Hue

The test images were also rendered by adjusting each pixel's lightness, chroma, and hue angle separately. They can establish the impact of colour attributes in image quality. For each pixel, a scaling factor was multiplied for lightness with linear alteration and chroma changes. For hue rendering, scaling factors were added. Lightness with nonlinear alteration was another method used to render lightness that applies a power factor (Table 2).[23]

Results and Discussion

Image Quality vs. Physical Parameters

Most image attributes in this experiment showed that there is a difference at the 95% confidence interval between different levels of each physical parameter. When the confidence interval is considered, it can be said that the results are highly independent on image contexts.

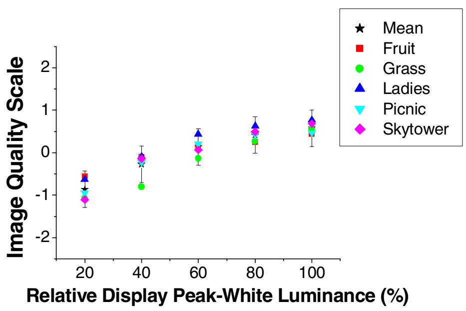

Figure 2. *Image quality assessment when varying the peak-white luminance of the mobile display*

Figure 2 plots image quality visual results (image quality scale values) for varied display peak-white luminance levels. The horizontal axis of the plot represents the percentage luminance level for the test images reproduced on the original status of the display, i.e. 100% is the original. As can be seen, when luminance is reduced by up to 20% of the original, the scale values are also linearly decreased.

The default resolution of the display was 200 pixels-per-inch (ppi). The test images were rendered and re-sampled to simulate display conditions with different resolutions from 200 to 80 ppi. Figure 3 shows that image quality scale values were linearly reduced for decreasing resolutions.

A nonlinear effect was found when increasing bit depths from 4 to 8 bits (Figure 4). In the 2-inch sized mobile LCD, the perceptible visual difference between 6-bit and 8-bit images was not large.

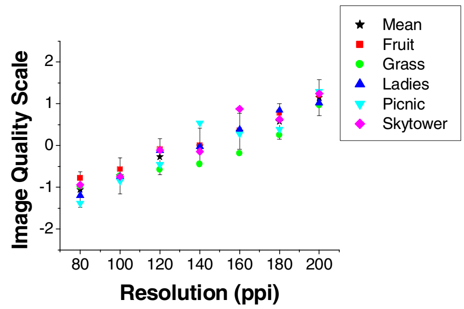

Figure 3. *Image quality assessment when varying the resolution of the mobile display*

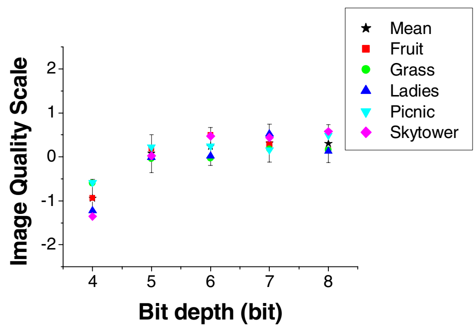

Figure 4. *Image quality assessment when varying the bit depth of the mobile display*

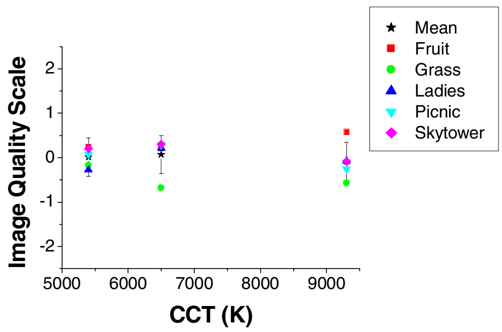

Figure 5. *Image quality assessment when varying the CCT of the mobile display (5400, 6500, and 9300 K)*

As shown in Figure 5, there was no statistical image quality difference between CCTs (5400, 6500, and 9300K). This implies that observers fully adapted to each white point. Vogel et al.[16] studied the optimal white-point of a display and found that there is optimal and acceptable CCT region in different test images. Their observers were allowed to view and compare all the renderings in the same screen and choose the most preferred one. However, in this study, only a single image was assessed and judged.

In Figure 6, image quality results are plotted against relative lightness ratio (%). It shows that scale values were changed for lightness in linear manipulation. The percentage of relative lightness level for an original image varies along the horizontal axis. Image quality was enhanced as lightness increases up to 90%. This effect in general agrees with the peak-white luminance as shown in Figure 2 as expected.

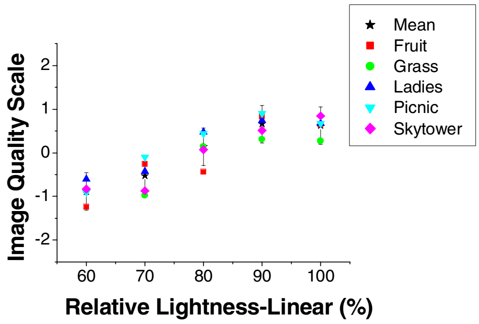

Figure 6. *Image quality assessment when varying the lightness of test images using linear manipulation*

For lightness with non-linear adjustment, image quality scale values are plotted in Figure 7. The horizontal axis represents the power factor of non-linear rendering as described. The level of 1.2 shows the best performance, but there is very little statistical difference from that of 1.0, (original). The nonlinear lightness function used in Table 1 has the property with a lightness reduction accompanied by a higher parameter.

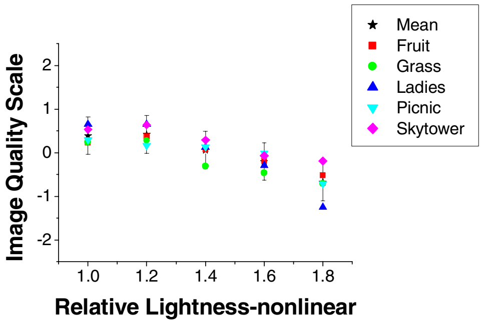

Figure 7. *Image quality assessment when varying the lightness of test images using nonlinear manipulation*

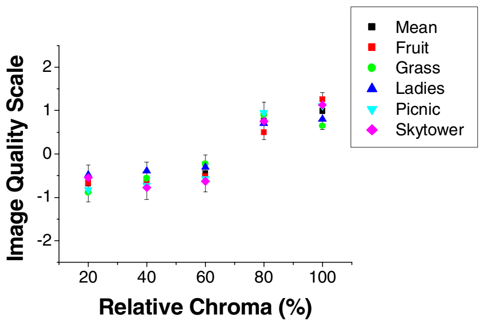

Figure 8. *Image quality assessment when varying the chroma of test images*

Figure 8 illustrates that higher chroma shows higher scale values. There was a drastic increase from level of 60 to 80 %. Observers would regard a level of 60 as being more greyish.

It is clearly seen, in the hue rendering experiment, that some data were image dependent. Some images showed that there is a difference out of the 95% confidence interval in hue angle rendering. As can be seen in Figure 9, sky colour dominated image (Skytower) showed the highest point for level –15 and fruit colour dominated image (Fruit) had a peak point for level 15. However, the other three images showed the highest scale values for level 0 (original). A related research also demonstrated that hue rendering generates ambiguous peaks.[2] Hue rendering also can cause some colours out of display colour gamut. Those colours may reduce the overall quality of images.

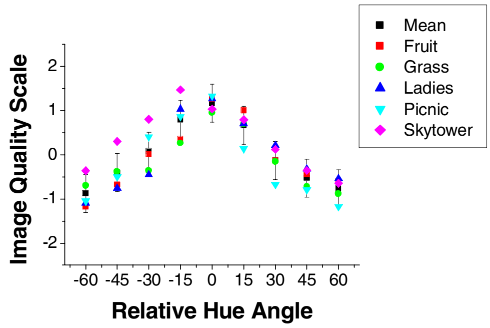

Figure 9. *Image quality assessment when varying the hue angle of test images*

Inter-Comparison

The image quality results were inter-compared between the perceptual attributes in terms of the Pearson correlation coefficient (r).[1] In Table 3, naturalness and quality showed a very high linear correlation (0.96). Clearness seemed to be highly associated with sharpness (0.97). Quality and preference were judged to be the same attribute (0.99).

Table 3. Inter-comparison between attributes (N: Naturalness, CL: Clearness, S: Sharpness, CT: Contrastness, CF: Colourfulness, IQ: Image Quality, P:Preference)

	N	CL	S	CT	CF	IQ	P
N	1.00	0.66	0.59	0.75	0.83	**0.96**	0.96
CL		1.00	**0.97**	0.89	0.58	0.82	0.80
S			1.00	0.84	0.53	0.77	0.74
CT				1.00	0.73	0.86	0.84
CF					1.00	0.82	0.83
IQ						1.00	**0.99**
P							1.00

Integrated Empirical Image Quality Modelling

All psychophysical data were used to develop an empirical image quality metric through a stepwise regression method[1]. Naturalness showed highest weights in image quality (or preference) and clearness followed it. It was found that clearness is highly correlated to sharpness, as shown in Table 3. This resultant fact, in which naturalness plays important roles in image quality, agrees with the findings of Yendrikhovskij et al.[9] and Boust et al.[17]. As can be seen in Figure 10, the correlation coefficient between the developed image quality model (IQm) using the two main attributes (Equation 3) and the corresponding MOS (Mean Observer Score) was 0.988.

MOS is a mean category of all responses from observers for each image. IQm output can be more obviously represented when MOS, i.e. from 1 to 9, is used, rather than scale values derived from the z scores. The correlation coefficient between MOS and scale values was 0.995.

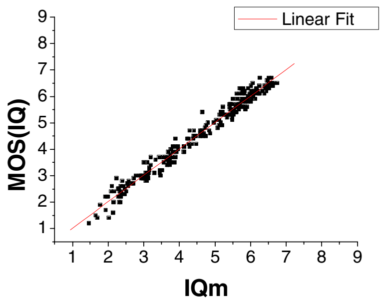

Figure 10. *MOS versus IQm (Image Quality Model): r=0.988*

$$IQm = 0.7626N + 0.4090S - 1.093 \tag{3}$$

where IQm: Image Quality Metric
N: Naturalness
S: Sharpness

Summary

Empirical image quality metrics were developed using the obtained psychophysical data. A stepwise regression method[1] was performed between the MOS of image quality and that of the other 5 attributes, i.e. naturalness, clearness, sharpness, contrast, and colourfulness. It was found that naturalness and sharpness are the principal affective attributes in the image quality of a mobile LCD.

References

[1] N. R. Draper and H. Smith, Applied regression analysis, John Wiley & Sons, Inc., New York (1966).
[2] S. R. Fernandez, M. D. Fairchild, and K. Braun, Analysis of observer and cultural variability while generating "preferred" color reproductions of pictorial images. Journal of Imaging Science and Technology, 49(1), 96 (2005).
[3] M. D. Fairchild, Measuring and modelling image quality, Chester F. Carlson Center for Imaging Science industrial associates program, Rochester (1999).
[4] R. W. G. Hunt, Measuring colour, 3rd ed. Fountain Press, London, England (1998).
[5] M. R. Luo, G. Cui, and B. Rigg, The development of the CIE 2000 colour-difference formula: CIEDE2000, Color Research and Application, 26, 340 (2001).
[6] S. Westland and C. Ripamonti, Computational colour science, John Wiley & Sons, Ltd, Chichester, England (2004).
[7] X. Zhang and B. A. Wandell, A spatial extension of CIELAB for digital color image reproduction, SID 96 Digest (1996).
[8] T. J. W. M. Janssen, Computational image quality, PhD thesis, Eindhoven, IPO (1999).
[9] S. N. Yendrikhovskij, F. J. J. Blommaert, and H. de Ridder, Colour reproduction and the naturalness constraint. Color Research and Application, 24, 53 (1999).
[10] R. W. G. Hunt, The reproduction of colour, 6th ed., John Wiley & Sons Ltd, Chichester, England (2004).
[11] D. L. Post and C. S. Calhoun, An evaluation of methods for producing desired colors on CRT monitors, Color Research Application. 14, pg. 172 (1989).
[12] J. Morovic and Y. Wang, Influence of test image choice on experimental results, Proc. IS&T/SID Eleventh Color Imaging Conference, pg. 143 (2003).
[13] C. J. Bartleson, Visual measurements. In Bartleson, C. J. & Grum, F. (Ed), Optical Radiation Measurements, 5, Orlando, FL: Academic Press (1984).
[14] W. S. Torgerson, Theory and methods of scaling. Malabar, FL: Robert E. Krieger Publishing Co. (1985).
[15] http://www.samsung.com/PressCenter/PressRelease
[16] I. M. L. C. Vogels and I. E. J. Heynderickx, Optimal and acceptable white-point settings of a display, Proc. IS&T/SID Twelfth Color Imaging Conference, pg. 233 (2004).
[17] C. Boust, F. Cittadini, M. B. Chouikha, H. Brettel, F. Vienot, S. Berche, and G. Alquie, Does an expert use memory colors to adjust images? Proc. IS&T/ SID Twelfth Color Imaging Conference, pg. 347 (2004).
[18] M. D. Fairchild, Color appearance models 2nd ed., John Wiley & Sons Ltd, Chichester, England (2005).
[19] P. G. Engeldrum, Psychometric scaling: a toolkit for imaging systems development, Imcotek Press, Winchester (2000).
[20] B. W. Keelan, Handbook of image quality: characterization and prediction, Marcel Dekker, New York (1994).
[21] S. Susstrunk, R. Buckley, and S. Swen, Standard RGB color spaces, Proc. IS&T/SID Seventh Color Imaging Conference, pg. 127 (1999).
[22] A. Sharma, Understanding color management, Thomson Delmar Learning, New York (2004).
[23] C. Sano, T. Song, and M. R. Luo, Colour differences for complex images, Proc. IS&T/SID Eleventh Color Imaging Conference, pg. 121 (2003).

Author Biography

Youn Jin Kim received his B.S. degree in Physics from Daejin University (Korea) in 2003 and MSc in Imaging Science from University of Derby (UK) in 2004. He has worked as a PhD student at the Colour & Imaging Group at the University of Leeds (UK) since 2004. His research interests are in colour image quality and enhancement. This work is supported by the Samsung Advanced Institute of Technology.

