Journal of Imaging Science and Technology® 52(4): 040904–040904-11, 2008.
© Society for Imaging Science and Technology 2008

Investigation of Large Display Color Image Appearance I: Important Factors Affecting Perceived Quality

Seo Young Choi▲, M. Ronnier Luo▲, Michael R. Pointer and Peter A. Rhodes
Department of Color Science, University of Leeds, Leeds, LS2 9JT United Kingdom
E-mail: seoyoung228@googlemail.com

Abstract. *A large-scale psychophysical experiment was performed to establish the important image appearance attributes controlling the perceived quality of images presented on a large display under dark surround conditions. Six image appearance attributes were chosen: Colorfulness, contrast, naturalness, visual information, sharpness, and image quality. A nine-point qualitative category scale was used to rate these six attributes for eight test images, each of which had 22 derivative images which varied in lightness, chroma, and sharpness. The influences of the three image manipulations on the six attributes, and the psychophysical relations between image quality and its constituent attributes were investigated. Multiple regression and factor analysis were conducted to derive an empirical image quality model. It was found that there were high correlations among the five attributes forming image quality: Between sharpness and contrast, and between naturalness and visual information. Furthermore, colorfulness, contrast, and naturalness were key attributes to have a significant impact on image quality.*
© 2008 Society for Imaging Science and Technology.
[DOI: 10.2352/J.ImagingSci.Technol.(2008)52:4(040904)]

INTRODUCTION

Understanding human perception of the quality of images is necessary for the development of displays, including making technical improvements such as gamut extension, increasing luminance, and reducing power consumption and display thickness. This has been one of main research themes in imaging industries and many studies have been conducted to achieve this aim. Most of them, however, focused on individual image appearance attributes which comprise image quality. Fedorovskaya and de Ridder[1,2] studied the naturalness-quality relationship of complex images by manipulating "colorfulness" through chroma and lightness adjustments to the images. They reached two conclusions: That high quality images were more colorful than those images considered as being the most natural, and image quality was not a monotonic function of colorfulness. Calabria[3] performed a large-scale investigation into image contrast and its effect on image preference. It was shown that images having medium contrast were preferred. All of these studies were based on an empirical approach in which subjective impressions of image appearance attributes could be understood by the results of observers' assessments on manipulated images.

A different approach was taken by Janssen[4] who described image quality not by the evaluation of diverse images but by understanding the visual-information process in the context of the human visuo-cognitive system. Two attributes were found to be the main perceptual constraints determining image quality. First, observers acquire visual information from the outside world via the reproduced scenes on imaging devices from which they construct an internal representation. Second, this is interpreted by means of a "confrontation" with memory. The first step requires that the reproduced images have a "usefulness" attribute (i.e., the image should be visually representative of the outside world) and the second requires a "naturalness" attribute.

The ultimate goal of all this research was to create empirical models that can mathematically predict the human perceptual image appearance. One approach towards this goal is measuring the perceived image appearance difference between an original image and its reproduction, leading to an image-difference metric such as iCAM.[5] Another approach involves image quality, which is the combination of a number of perceptual attributes, often called the "nesses."[6] This concept considers only individual images and not pairs. The attributes forming image quality are first modeled and then combined to determine image quality. Combination requires information about the psychophysical relationships between image quality and its constituent attributes. Colorfulness and contrast have been shown not to have a linear relationship with image quality[1–3] and so Engeldrum[7] has proposed empirical image quality models using logistic and Gaussian-like functions rather than a simple linear function.

As introduced thus far, there have been many research activities to develop models that are capable of predicting the perceived quality of images, as well as investigations into diverse image appearance attributes and their relationships. Most of the studies have, however, concentrated on characteristics of individual attributes. Image quality models have generally been developed by combining some of the attributes contributing towards image quality, but without first establishing which image appearance attributes were the key factors affecting image quality. This study was therefore carried out to identify perceptual attributes that have a significant impact on image quality using a large display that has recently become dominant in our lives. Five attributes which had previously individually been investigated were chosen and assessed under a dark surround: Colorfulness, contrast,

▲IS&T Member.

Received Jan. 22, 2008; accepted for publication Mar. 19, 2008; published online Jul. 17, 2008.
1062-3701/2008/52(4)/040904/11/$20.00.

Figure 1. The test images: Adults, Kids, Fruits, Harbor, Park, Sheep, Seashore, and Pier (clockwise). (Available in color as Supplemental Material on the IS&T website, www.imaging.org).

naturalness, sharpness, and visual information. In the second part of this article, the resulting key image appearance attributes will again be evaluated for the final goal of this work: The quantification of the surround effect on the color image appearance of a large display.

EXPERIMENTAL SETUP

Psychophysical Experimental Setting

The psychophysical experiment was carried out in a darkened room. Observers consisting of graduate students in Color and Imaging Group at the Department of Color Science in the University of Leeds participated in each assessment for the six image appearance attributes (colorfulness, contrast, naturalness, sharpness, visual information, and quality of color images). All observers undertook the Ishihara Test for color blindness to ensure their normal color vision. Age range of the observers was 24–37. Most observers had experiences with psychophysical experiments evaluating diverse image appearance attributes. Two minutes of adaptation time were given to each observer. The viewing distance was 2 m at which the size of images seen on the display subtended a visual field of $26.3°(\mathrm{H}) \times 15.2°(\mathrm{V})$. Eight test images were carefully chosen in order to cover a wide range of image contents. These are introduced in Figure 1: two portraits (*Kids* and *Adults*), one fruit (*Fruits*), and five natural scenes (*Seashore, Sheep, Pier, Harbor*, and *Park*). To scale observers' perceptions of the six attributes, a categorical judgment technique was adopted. Observers were asked to rate each image using a nine-point verbally labeled category scale. For example: "9" corresponded to the "highest quality imaginable," "5" to the "average quality," and "1" to the "lowest quality imaginable." Each image was assessed by 10–14 observers. The experiment was divided into six observing sessions. Each session contained 184 assessments (8 images $\times$ 23 manipulations $\times$ 1 image appearance attribute) and lasted for approximately 15–20 min.

A 42 in. plasma display panel (Samsung, PPM42H3) with 1024×768 pixel resolution was used to present the images. The reference white of this display was 174 cd m^{-2} with a correlated color temperature of 8940 K. The colorimetric characterization model was built using a three-dimensional (3D) LUT ($13 \times 13 \times 13$ RGB values) and tetrahedral interpolation.[8] The spectral power distributions of 2197 colors were measured using a Minolta CS1000 telespectroradiometer providing XYZ tristimulus values. To test the accuracy of the characterization model, CIELAB color differences (ΔE^*_{ab}) between the measured and predicted tristimulus values of the 64 test colors were calculated, giving average errors of 1.16 and 1.46 for the forward and reverse models, respectively.

Tools Providing Variations to Test Images

The eight test images were rendered using 22 methods to provide a wide, but realistic, range of variations to observers. RGB values of each pixel in a test image were transformed into XYZ tristimulus values using the display characterization model developed earlier. The CIECAM02 color appearance model[9] was then used to manipulate each original image in terms of lightness (J) and chroma (C). Manipulating J in the frequency domain created a sharpness change. Each method was labeled in three parts: First, the attribute rendered; second, the type of manipulating function; and third, the amount of variation given to the original image. For example, LSL is a lightness manipulation (L) by sigmoid function (S) with a large variation (L). The 22 methods are summarized in Table I using their names and according to the type of manipulation function. Each of the original eight test images could therefore produce 22 derivative manipulated images.

Lightness (J) *manipulations*

Eleven manipulations in the J channel were performed: Four linear, three sigmoid, three inverse sigmoid functions, and the local color correction method. For overall darkened images, four linear functions with four different slopes were used. These are expressed in Eq. (1).

$$J_{\mathrm{output}} = S \times J_{\mathrm{input}}, \tag{1}$$

where S values are 0.95, 0.9, 0.85, and 0.8 for LL095, LL09, LL085, and LL08 respectively.

The three sigmoid functions and three inverse sigmoid functions are expressed in Eqs. (2) and (3).

$$J_{\mathrm{output}} = \frac{100}{[1(1+M^E)] \times \{1 + [M/(0.01 \times J_{\mathrm{input}})]^E\}}, \tag{2}$$

where $M = 1.23$ and $E = 1.45$ for LSS, $M = 0.75$ and $E = 1.9$ for LSM, and $M = 0.63$ and $E = 2.35$ for LSL. The "S" in LSS dictates a small variation, "M" in LSM a medium variation, and "L" in LSL a large variation.

$$J_{\mathrm{output}} = 100 \times M \times \left[\frac{1 - 0.01 \times [1/(L + M^E)] \times J_{\mathrm{input}}}{0.01 \times [1/(1+M^E)] \times J_{\mathrm{input}}}\right]^{-1/E}, \tag{3}$$

where $M = 1.23$ and $E = 1.45$ for LISS, $M = 0.75$ and $E = 1.9$ for LISM, and $M = 0.63$ and $E = 2.35$ for LISL.

The sigmoid function reduced the lightness in the dark areas of the original image but increased the lightness in light areas. This function was controlled by M and E in Eq. (2). As the value of M was smaller and E was larger, there was more lightness reduction for dark regions and more lightness increase for light regions. The inverse sigmoid function provided an opposite effect on the original image.

Table I. Summary of the 22 image-manipulation methods.

Functions	Linear function	Sigmoid function	Inverse-Sigmoid function	Using HFE (High-Frequency Emphasis) Filter		
Meaning of the *third part* in the name of image manipulation	*Slope* in linear function	*Amount of variation* *L* (large) *M* (medium) *S* (small)		*Cutoff frequency* (1/11—most sharpened image)	Using LCC (Local Color Correction) Method	Using Barten's CSF (Contrast Sensitivity Function)
Lightness	LL*095*	LS*L*	LIS*L*		LLCC	
	LL*09*	LS*M*	LIS*M*			
	LL*085*	LS*S*	LIS*S*			
	LL*08*					
Chroma	CL*09*	CS	CIS			
	CL*08*					
	CL*07*					
	CL*06*					
Sharpness				SHFE*1/3*		SCSF
				SHFE*1/5*		
				SHFE*1/7*		
				SHFE*1/11*		

For the local color correction method,[10] the background luminance factor (Y_b) in CIECAM02 was computed from the absolute luminance value of each pixel; Y_b in each pixel$=100\times$luminance in each pixel/luminance of reference white. This individual Y_b value was used to compute a new lightness value for each pixel. The original lightness values of the image could thus be controlled on pixel-by-pixel basis, so that the original dark pixels were brightened but the original light pixels were darkened.

Chroma (*C*) *Manipulations*

The chromatic transfer functions used in this study included four linear, one sigmoid, and one inverse sigmoid functions. The four linear functions with four different slopes are expressed in Eq. (4).

$$C_{\text{output}} = S \times C_{\text{input}}, \tag{4}$$

where S values are 0.9, 0.8, 0.7, and 0.6 for CL09, CL08, CL07, and CL06 respectively.

For the sigmoid and inverse sigmoid functions, the same Eqs. (2) and (3) as were applied to the lightness manipulation were used; however "$C_{\text{input}}/C_{\text{max}}$" (maximum chroma value in an image) was an input instead of "$0.01\times J_{\text{input}}$" and only $M=0.63$ and $E=2.35$ were used. The sigmoid function reduced the chroma of less colorful areas of the original image, but increased the chroma of more colorful areas, leading to an increase of chromatic contrast. The inverse sigmoid function provided the inverse effect on the original image.

Figure 2 illustrates the functions that were used for manipulating images in the lightness and chroma domains: The two linear functions with two slopes (0.95 and 0.6), three sigmoid, and three inverse sigmoid functions. A 45° line is also shown to represent no change between $0.01\times J_{\text{input}}(C_{\text{input}}/C_{\text{max}})$ and $J_{\text{output}}(C_{\text{output}})$.

Sharpness Manipulations

Two methods were applied to increase sharpness. A constant, 1.1, was multiplied to the amplitude components of two frequency ranges: 0.6–9.3 cpd covered the top 50% sensitivity according to Barten's contrast sensitivity function[11] and 23.97–31.96 cpd corresponded to the edge areas. For other frequency ranges, a constant, 0.9, was multiplied to their amplitude components. Second, a high frequency emphasis filter was used and its equation is given in Eq. (5). Since the cutoff frequency parameter (d) was smaller, the image became sharper due to a reduction in low frequency information; function SHFE1/11 produced the sharpest image.

$$Filter = 1 + 1.5 \times \left\{ 1 - \exp\left[\frac{-x^2}{(2 \times d^2)} \right] \right\}, \tag{5}$$

where x is frequency, $d=1024\times P$, cutoff frequency parameter. P equals 1/3, 1/5, 1/7, or 1/11 for SHFE1/3, SHFE1/5, SHFE1/7, and SHFE1/11 respectively. The constant, 1024, is the horizontal resolution of the display in pixels used in this study.

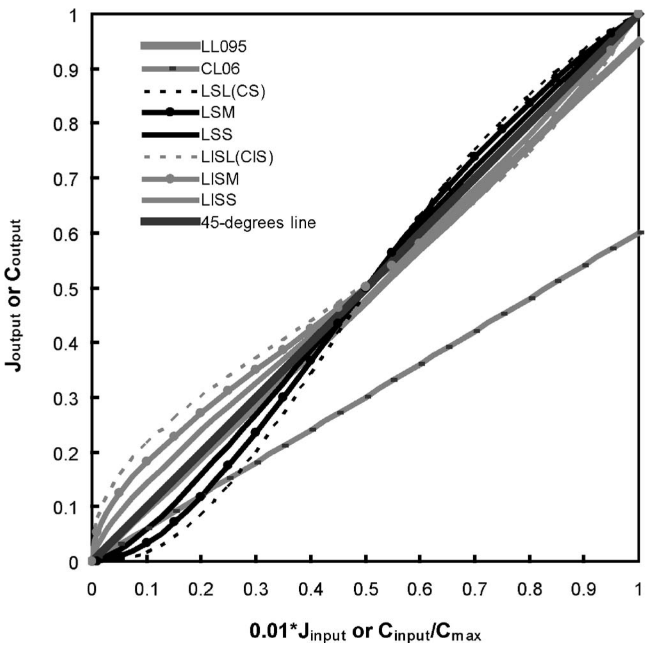

Figure 2. Illustration of the two linear, three sigmoid, and three inverse-sigmoid functions. Also shown is a line at 45°.

RESULTS AND DISCUSSIONS

Observer Variations

Observer variation was computed from the coefficient of variation (CV) defined in Eq. (6), which is a statistical measure to represent the agreement between two sets of data. Two test images (*Fruits* and *Pier*) and their manipulated images were assessed twice by eight observers in terms of colorfulness, contrast, and image quality. The CV values were computed between the two repeated judgments to represent intraobserver agreement. To determine inter-observer agreement, the CV values between the individual observer data and mean data of all observers were calculated using all visual results (i.e., eight test images and their manipulated images assessed by 10–14 observers in terms of the six attributes).

$$CV = \frac{100}{\bar{y}}\left[\sum (x_i - y_i)^2/n\right]^{1/2}, \tag{6}$$

where n is the number of assessed images. For intra-observer agreement, x_i and y_i are the first- and second-judgment data, respectively. When assessing inter-observer agreement, x_i and y_i are individual observer data and average data of all observers. The mean value of the y_i data set is $\bar{y}$.

The mean CV values for intra-observer agreement were 19, 18, and 20 for colorfulness, contrast, and image quality respectively. Table II summarizes the resulting mean CV values (covering all observers) for inter-observer agreement. The mean CV values for inter-observer agreement are 17, and 18, and 22 for colorfulness, contrast, and image quality, respectively, which indicate that observers performed similarly in terms of within an individual observer and between observers for these three attributes. The largest mean CV value is seen for naturalness in Table II. The three attributes

Table II. Inter-observer agreement in terms of mean CV values for each of the six attributes.

Inter-Observer agreement	Mean CV
Colorfulness	17
Contrast	18
Sharpness	17
Naturalness	25
Visual information	20
Image quality	22

having the smallest mean CV value are colorfulness, contrast, and sharpness. It can therefore be assumed that naturalness is the most difficult for observers to scale, whereas colorfulness, contrast, and sharpness are relatively easier to scale. In other words, observers might use different criteria in the evaluation of naturalness, whereas similar criteria might be used in the assessment of colorfulness, contrast, and sharpness.

Influences of Image Manipulation on the Six Image Appearance Attributes

The raw experimental data were the category numbers assigned by each observer. To convert the category-scaling data into equal-interval scale values, Case V of Thurstone's law of comparative judgments was adopted.[12] These converted data were interval scale values of image appearance attributes for color images viewed on the display.

The mean scale value of the eight test images is plotted with the scale value of each of the eight test images simultaneously against each of the 23 image manipulations in Figures 3–8 for each of the six image appearance attributes. The scale values of the eight test images are expressed using different symbols. The mean scale values of all test images are symbolized using a horizontal bar (-) and their 95% confidence intervals using error bars. These symbols will be common to the six graphs for the six image appearance attributes. There is an arrow pointing to the original image in each figure.

Image Dependency

To examine whether the visual results for the six attributes are image independent, 95% confidence intervals, Eq. (7), of the mean scale values were computed and are shown in Figures 3–8 using error bars.

$$CI = \frac{1.96}{\sqrt{2 \times N}}, \tag{7}$$

where CI is 95% confidence interval, and N represents the number of observers who assessed each of the six image appearance attributes.

Most of scale values for the eight test images fall within the 95% confidence interval from the mean of the eight test images for most image manipulations and for each of the six attributes in Figures 3–8. This indicates that the changes in

Figure 3. Perceived image-colorfulness scale against the 23 image manipulations in order from lowest colorfulness to highest colorfulness. Also shown are the 95% confidence intervals of the mean scale values of the eight test images.

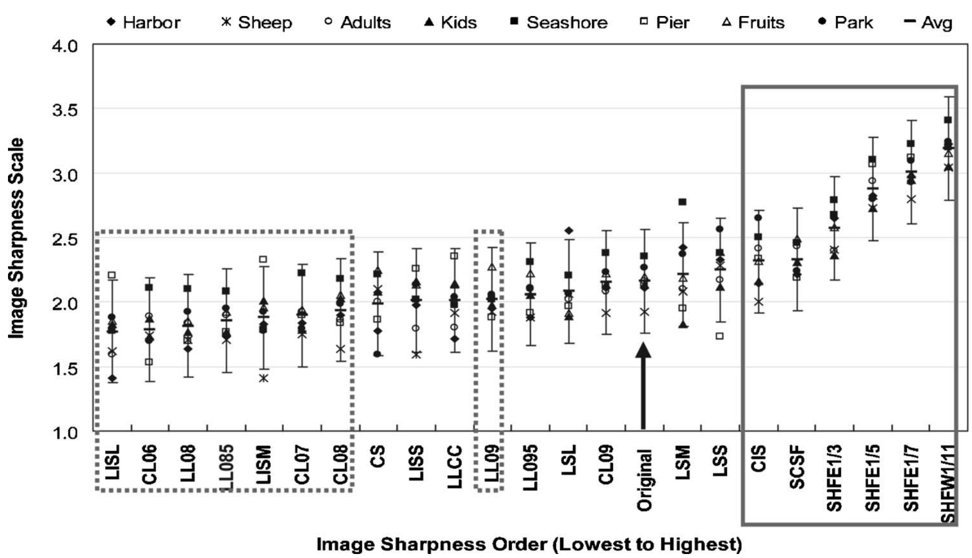

Figure 4. Perceived image-sharpness scale against the 23 image manipulations in order from lowest sharpness to highest sharpness.

each of the six attributes affected by the 23 image manipulations are similar regardless of image content. Hence, the complete interval scales for each of the six attributes were made by averaging scale values across the eight test images. For the following result parts, comparisons will be made using these mean scale values.

Independent samples t-test was conducted to establish whether each of the six attributes was perceived significantly differently between the particular manipulated image and the original. This test was constructed on the observed differences between two mean scale values of the original image and each of the 22 manipulated images. The main visual phenomena in the results are discussed in brief below.

Image Colorfulness

The mean (across the eight test images) and individual image colorfulness scale values are plotted against the 23 image manipulations in Figure 3. The manipulated images inside the boxes indicated by broken and solid lines were found to be statistically less colorful and more colorful than the original image, respectively.

- The CIS image (in which low chromatic areas appear more chromatic, but high chromatic areas appear less chromatic than in the original image) was perceived to be the most colorful. On the other hand, the CS image (the opposite effect compared to the CIS image) looked

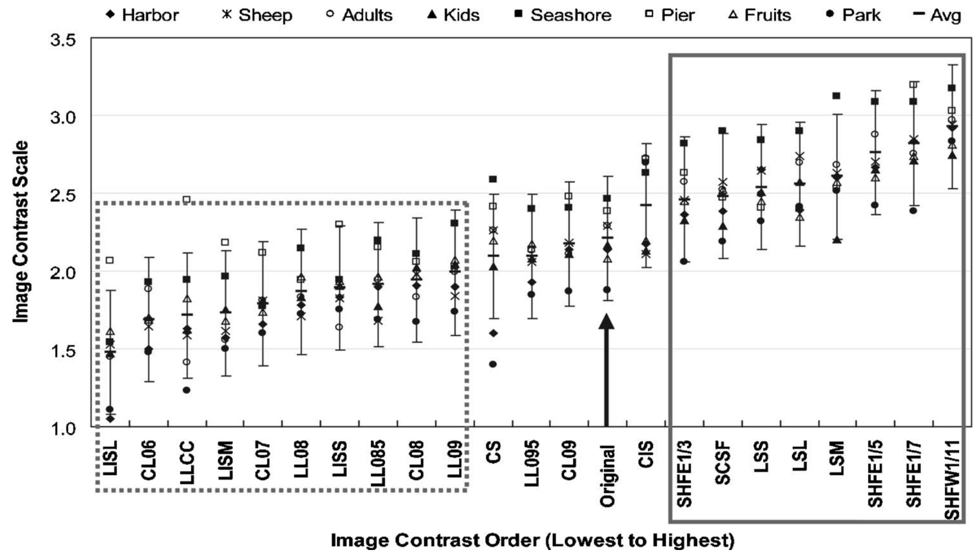

Figure 5. Perceived image-contrast scale against the 23 image manipulations in order from lowest contrast to highest contrast.

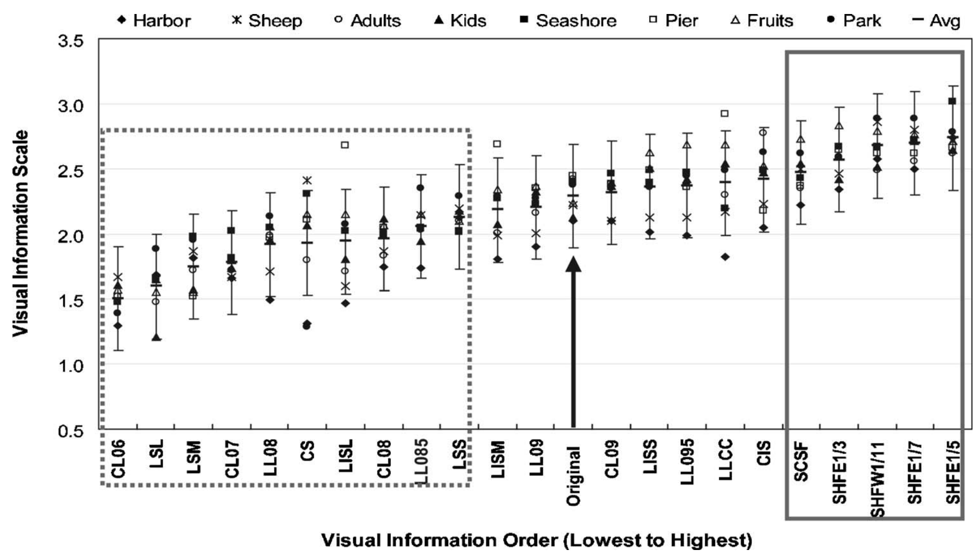

Figure 6. Perceived visual-information scale against the 23 image manipulations ranked from lowest visual information to highest visual information.

to be significantly less colorful than the original image. This indicates that increasing chroma for low chromatic areas in the image is more effective in order to improve perceived colorfulness compared to increasing chroma for high chromatic areas in the image.

- Among the five least colorful images, the LISL image (where dark areas are lighter and light areas are darker) had very much less lightness contrast while the other four images had decreased chroma. The LISL image might appear washed out due to the marked decrease in lightness-contrast, leading to noticeably less colorfulness.
- There is a tendency for the sharpened images (SCSF, SHFE1/3, SHFE1/5, SHFE1/7, and SHFE1/11) to appear slightly more (but not significantly more) colorful than the original.

Image Sharpness

The mean (across the eight test images) and individual image sharpness scale values are plotted against the 23 image manipulations in Figure 4. The manipulated images inside the boxes indicated by broken and solid lines were respectively found to be significantly less sharp and sharper than the original image.

- Apart from the five sharpened images, the CIS image that appeared the most colorful was also assessed as being markedly sharper than the original image. This indicates that an enhanced colorfulness can also cause images to appear sharper.
- The LISL image that looked to be significantly less colorful than the original in Figure 3 was judged the least

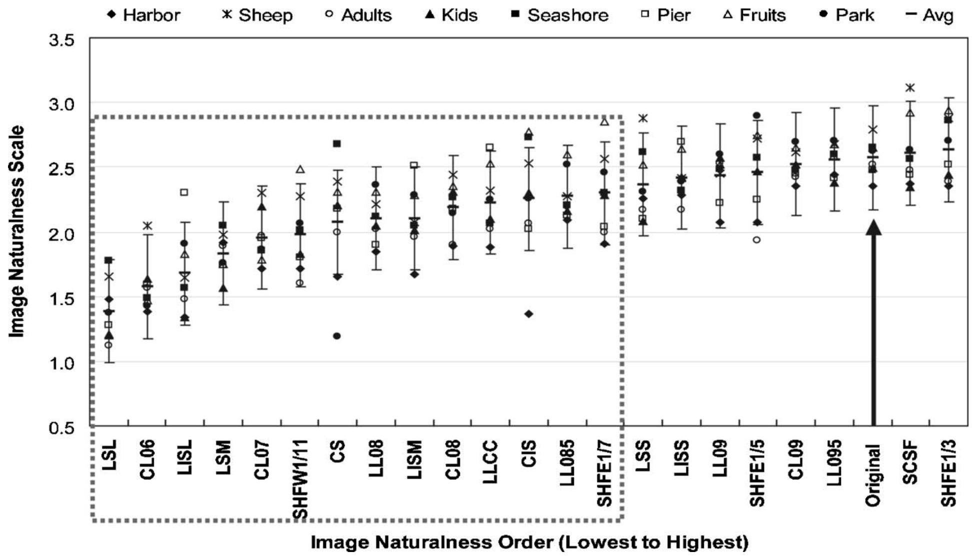

Figure 7. Perceived image-naturalness scale against the 23 image manipulations in order from lowest naturalness to highest naturalness.

sharp. It is thought that the much decreased lightness-contrast in the LISL image may cause a lowering in sharpness due to decreased colorfulness.
- Among the five most sharpened images, it can be seen that enhancing edge information by cutting the low frequency component (in SHFE images) results in a more noticeable increase in perceived sharpness. Enhancing the frequency range covering the top 50% human contrast sensitivity (in SCSF image) is less effective.

Image Contrast
The mean (across the eight test images) and individual image contrast scale values are plotted against the 23 image manipulations in Figure 5. The manipulated images inside the boxes indicated by broken and solid lines were respectively judged to have significantly lower contrast and higher contrast than the original image.

- The three sharpened images were perceived to have the highest contrast, followed by the three lightness-based sigmoid manipulated images (dark areas are darker and light areas are lighter).
- From the manipulated images in the box represented by the broken lines, it is seen that perceived contrast diminishes as chroma, lightness, and lightness-contrast in images decrease.
- These two results suggest that the mean image contrast can be affected not only by lightness but also by chroma and sharpness changes.

Visual Information
The attribute "visual information" was adopted from the image quality semantics study by Janssen in which the term "usefulness" was used instead of visual information.[4] In the current experiment, observers were instructed to judge visual information as "Does an image provide similar visual information as would be expected from the real scene?" The mean and individual visual information scale values are plotted against the 23 image manipulations in Figure 6. The manipulated images inside the boxes indicated by broken and solid lines were estimated to have significantly lower visual information and higher visual information than the original image, respectively.

- The highest ranking images are all sharpened images. The sharpest image produced by SHFE1/11 was not, however, perceived to have the best visual information. This implies that too much image sharpening can inhibit observers from getting enough visual information.
- Images which are lower ranking than the original have a large reduction in chroma, lightness, or lightness-contrast. Additionally, three lightness-based sigmoid functions provided lower scale values than the original image, due to lack of distinguishability in dark areas of images.

Image Naturalness
The mean and individual image naturalness scale values are plotted against the 23 image manipulations in Figure 7. The manipulated images inside the box indicated by broken lines were judged to appear significantly less natural than the original image.

- None of the image manipulations can noticeably improve image naturalness compared to the original image.
- The images that appeared significantly less natural than the original image can be categorized according to four characteristics: Loss in colorfulness, loss of shadow detail, washed-out appearance due to considerably decreased lightness-contrast, or too much sharpening. These are considered to be important factors contributing towards the perception of naturalness.

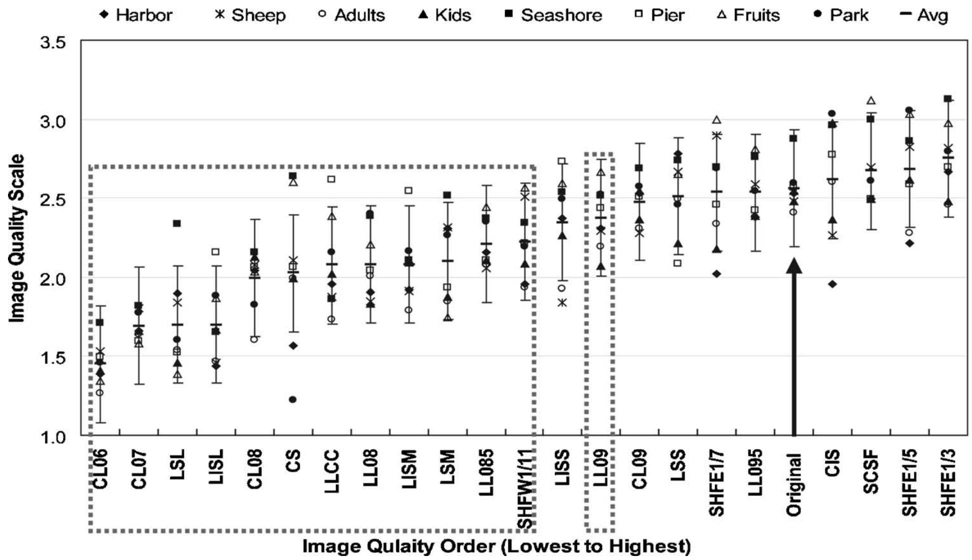

Figure 8. Perceived image-quality scale against the 23 image manipulations in order from lowest quality to highest quality.

Image Quality

The mean and individual image quality scale values are plotted against the 23 image manipulations in Figure 8. The manipulated images inside the box represented by the broken lines were judged to have significantly lower image quality than the original image.

- As with the results for image naturalness (Figure 7), there is no manipulated image having a higher quality than the original image.
- The manipulations which decrease chroma (CL06 and CL07), and the two extreme lightness-based sigmoid (LSL) and inverse sigmoid (LISL) manipulations showed the poorest image quality, which also looked most unnatural.
- The lightness-based sigmoid functions produced images having lower image quality and naturalness than any others, whereas these functions also resulted in higher image contrast. Hence, when evaluating image quality, the two attributes of naturalness and contrast may compromise each other.

Relationships between Image Quality and Other Image Appearance Attributes

The psychophysical relationships between image quality and each of the five image appearance attributes were investigated. Figures 9(a)–9(g) plot the mean scale values of image quality against those of each of colorfulness, sharpness, contrast, visual information, and naturalness, respectively. The data points are plotted with four different symbols (■, ○, + and ×) corresponding, respectively, to lightness, chroma, and sharpness manipulations, together with the original image. A best-fit curve was also given so as to indicate the trend in the relationship between image quality and each of the five attributes.

Figure 9(a) shows that image quality increases with an increase in colorfulness, and stabilizes when colorfulness reaches a certain level. The CIS image (where low chromatic areas appear more chromatic but high chromatic areas appear less chromatic than in the original image) is judged to be more colorful than the original image but with no improvement in image quality. If images look much more colorful than CIS images, image quality may begin to decrease after its maximum. This was proven by other studies.[1,2] The same relationship between image quality and image sharpness can be seen in Figure 9(b), and between image quality and image contrast in Figure 9(c), i.e., an inverted-U shape. Image quality increases to a certain sharpness level (SHFE1/3 and SHEF1/5) and contrast level (CIS, SHFE1/3, 1/5, 1/7 and SCSF) before falling. The highly sharpened image (SHFE1/11) and the image having a loss in shadow detail (LSL) both have higher image contrast than others, however their image quality is lower. Figure 9(d) shows image quality scale plotted against visual information scale. It is seen that image quality rises and then is slightly saturated with an increase in visual information. Overall, sharper images provided greater texture detail and thus provided more visual information. Figure 9(e) illustrates a clear positive linear relationship between image quality and naturalness.

Important Image Appearance Attributes Affecting Image Quality

Multiple regression and factor analysis using the principal component method were conducted to determine the important image appearance attributes affecting image quality. The five attributes (colorfulness, sharpness, contrast, visual information, and naturalness) were independent variables used to predict image quality in these analyses. Multicollinearity can cause problems in multiple regression analysis where there are high intercorrelations between independent variables.[13] Hence, factor analysis was performed to remove highly intercorrelated variables.

The results of factor analysis are described in Tables III and IV. The "Cumulative %" column explains the percentage of variance accounted for by the first *n* rotated compo-

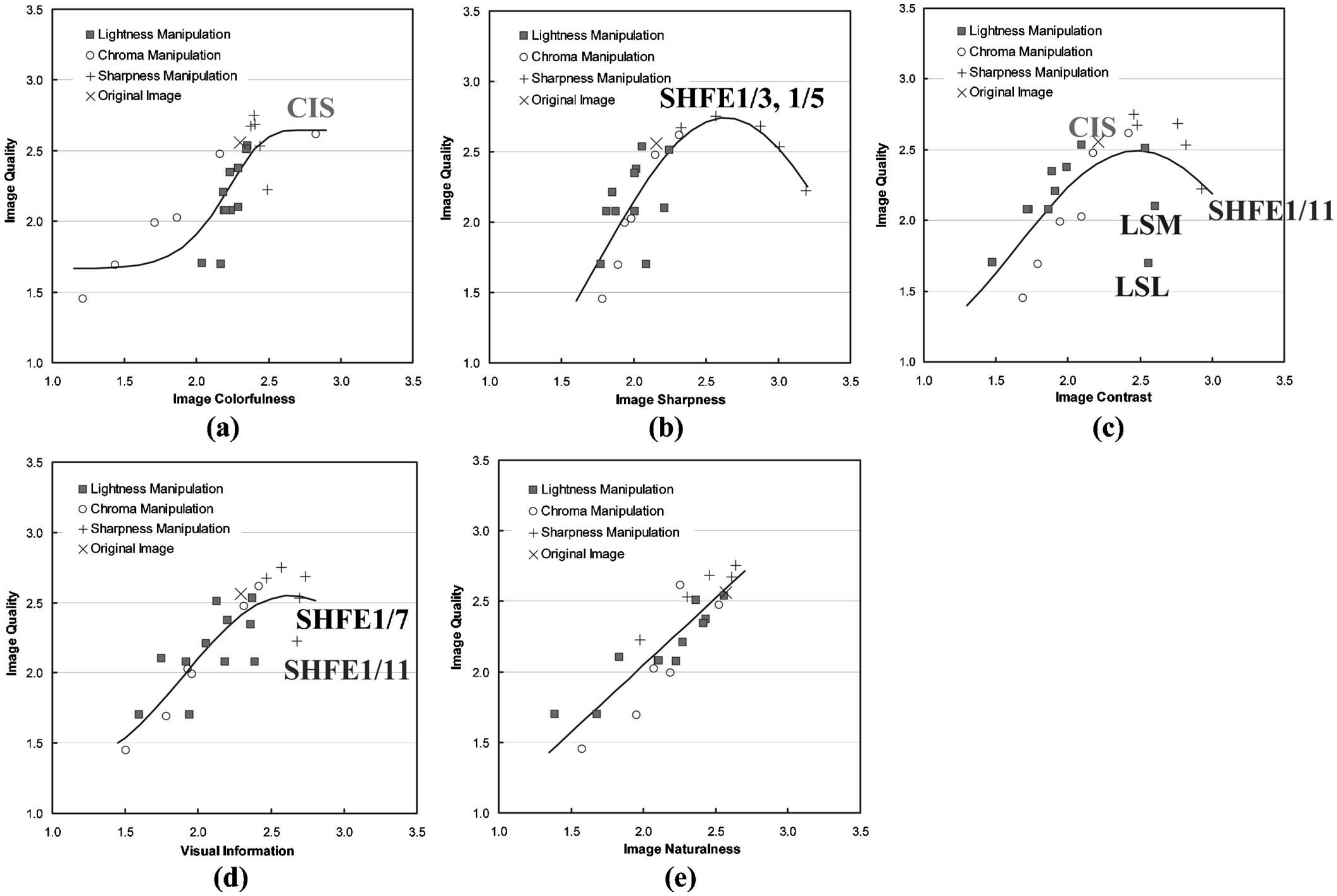

Figure 9. Image quality scale versus (a) image colorfulness, (b) image sharpness, (c) image contrast, (d) visual information, and (e) image naturalness. Also shown are (a) Weibull fitting curve, (b)–(d) Sinusoidal fitting curves, and (e) fitting line.

Table III. Total variance explained by the first component, and the first and second components.

Component	Cumulative %
1	40.44
2	80.84

nents in Table III. For example, the first two components account for nearly 81% of the variability in the five independent variables used for determining image quality. In order to know what the first and second components actually represent, correlation coefficients between the components and five independent variables were computed; these are given in Table IV. The first component is the most highly correlated with contrast and sharpness. The contrast is, however, better representative, since it is less correlated with the second component. The second component is most highly correlated with naturalness. These two findings suggest that we should focus on contrast and naturalness for controlling image quality.

The results of factor analysis indicate that image colorfulness does not arise from any single component. Therefore, to further investigate colorfulness, the mean score values of contrast, sharpness, naturalness, visual information, and quality are plotted against those of colorfulness in Figure 10. Image sharpness does not seem to be affected by variation in image colorfulness. Image contrast and visual information are not changed beyond a certain level of image colorfulness. Image quality has the largest difference between the least and most colorful images. Although image quality has a clear linear relationship with image naturalness [Figure 9(e)], their relationships with change in image colorfulness are not identical. That is, as image colorfulness increases, image quality tends to flatten out while image naturalness

Table IV. The correlation coefficients between each of five attributes and each of two components in the rotated component matrix.

Five perceptual attributes	Component 1	Component 2
Contrast	**0.94**	0.14
Sharpness	**0.89**	0.31
Naturalness	0.01	**0.92**
Visual information	0.45	**0.79**
Colorfulness	0.37	0.66

Table V. The empirical image quality models with different independent variables.

Number of independent variables	Contrast	Sharpness	Naturalness	Visual Information	Colorfulness	Constant	R
5	**0.20**	0.11	**0.68**	0.02	**0.21**	**−0.46**	0.927
4	**0.19**	**0.13**	**0.69**		**0.21**	**−0.46**	0.927
3	**0.28**		**0.71**		**0.23**	**−0.42**	0.925
2	**0.35**		**0.79**			**−0.26**	0.906
1	**0.51**					**1.13**	0.532
		0.59				**0.95**	0.570
			0.87			**0.32**	0.831
				0.84		**0.40**	0.765
					0.65	**0.83**	0.628

Table VI. The CV values computed between the experimental and predicted image quality.

Independent variables used to derive image quality models	CV
All five attributes	**7**
Contrast, naturalness, colorfulness, and sharpness	**7**
Contrast, naturalness, and colorfulness	**7**
Contrast and naturalness	**8**
Contrast	**15**
Naturalness	**11**
Colorfulness	**15**
Sharpness	**16**
Visual information	**13**

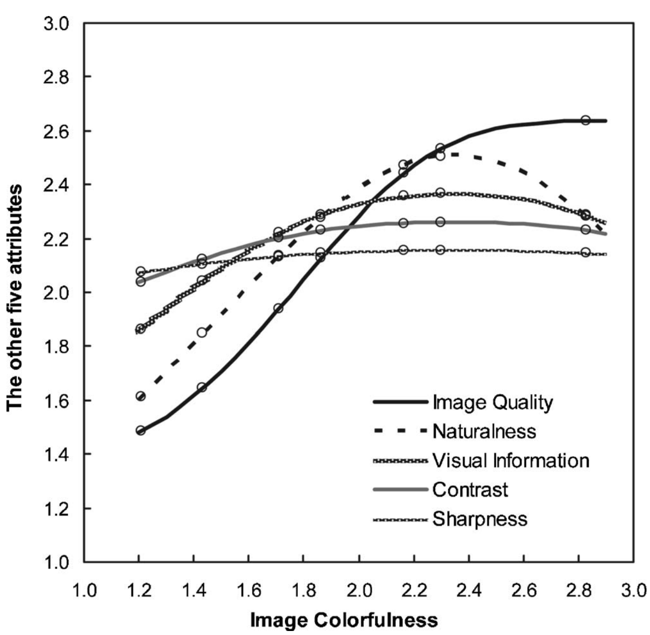

Figure 10. Mean scale values of contrast, sharpness, visual information, naturalness, and quality versus those of colorfulness.

falls after reaching a maximum. This implies that image colorfulness also needs to be considered as an important attribute influencing image quality in addition to the attributes of image contrast and image naturalness which were found in the earlier factor analysis.

Empirical image quality models were derived using different numbers of independent variables by multiple regression analysis. Table V summarizes coefficients for each independent variable and R value (multiple correlation coefficient) for each model. The coefficient for each independent variable shown in bold is statistically significant. Although the model with four independent variables gives the highest R value, sharpness and contrast were highly correlated (Pearson correlation coefficient=0.8). The model with three independent variables having an R of 0.925 may give a satisfactory prediction for the quality of images viewed on large displays. Among the five models using one independent variable, the model derived using naturalness performs the best, i.e., it has the largest R value of the five models.

In addition, the coefficient of variation (CV) was calculated in order to evaluate the agreement between observer-judged and predicted data using the empirical image quality models derived using different attributes (independent variables). Table VI shows the CV values for the nine models described in Table V. There is no CV difference for the top three models in Table VI, suggesting three attributes (contrast, naturalness, and colorfulness) are sufficient to explain image quality variations arising from images that differ in the lightness, contrast, and sharpness domains. Since the highest R value was seen for the image quality model derived using naturalness, among the single variable models in Table V, the corresponding CV value is the smallest in Table VI.

In summary, colorfulness, contrast, and naturalness were chosen as the key image appearance attributes affecting image quality. Naturalness is believed to be the most influential factor on image quality among them.

CONCLUSION

The psychophysical experiment described here was designed to investigate which attributes influence the image quality of a large display under a dark surround. The six image appearance attributes were evaluated using a categorical judgment method applied to eight test images: Quality, contrast, sharpness, visual information, naturalness, and colorfulness. The influence of changes in image lightness, chroma, and sharpness on these six attributes was examined. Subsequently, the psychophysical relationship between image quality and each of the other five image appearance attributes was investigated. The results revealed a strong positive linear relationship between image quality and naturalness. For colorfulness and visual information, image quality first increased and then stabilized at the high end. For the two spatial attributes (contrast and sharpness), image quality first rose and then fell from certain sharpness and contrast levels.

Finally, multiple regression and factor analyses were conducted to reveal which were the significant image appearance attributes affecting image quality among the five attributes studied. From these results, five independent variables were classified into two components. Contrast and naturalness were chosen as being representative for the two components in the factor analysis. Sharpness, which was highly correlated with contrast, and visual information, which was highly correlated with naturalness, were both removed. Colorfulness, however, did not result from any individual component. Also, variations in colorfulness affected image quality and naturalness differently. The results from multiple regression demonstrated that the empirical image quality model having three independent variables (naturalness, contrast, and colorfulness) had a similar performance to a model having all five independent variables. In conclusion, colorfulness, contrast, and naturalness were found to be important perceived attributes influencing image quality.

The eventual goal of this work is to quantify the effect of surround on color image appearance for large displays. The critical attributes chosen will be evaluated under different surround conditions together with image quality in the following paper.

REFERENCES

[1] H. de Ridder, “Progress naturalness and image quality: saturation and lightness variation in colour images of natural scenes”, J. Imaging Sci. Technol. **40**, 487 (1996).
[2] E. A. Fedorovskaya, “Chroma variations and perceived quality of color images of natural scenes”, Color Res. Appl. **22**, 97 (1997).
[3] A. J. Calabria and M. D. Fairchild, “Perceived image contrast and observer preference I. The effects of lightness, chroma, and sharpness manipulations on contrast perception”, J. Imaging Sci. Technol. **47**, 479 (2003).
[4] T. J. W. M. Janssen and F. J. J. Blommaert, “Image quality semantics”, J. Imaging Sci. Technol. **41**, 555 (1997).
[5] M. D. Fairchild and G. M. Johnson, “The iCAM framework for image appearance, image difference, and image quality”, J. Electron. Imaging **13**, 126 (2004).
[6] P. G. Engeldrum, “A framework for image quality models”, J. Imaging Sci. Technol. **39**, 312 (1995).
[7] P. G. Engeldrum, “Extending image quality models”, *Proc. IS&T’s PICS Conference* (IS&T, Springfield, VA, 2002) pp. 65–69.
[8] S. Y. Choi, M. R. Luo, P. A. Rhodes, E. G. Heo and I. S. Choi, “Colorimetric characterization model for plasma display panel”, J. Imaging Sci. Technol. **51**, 337 (2007).
[9] CIE Publication 159:2004, “A color appearance model for color management systems: CIECAM02”, (Central Bureau of the CIE, Vienna, 2004).
[10] N. Moroney, “Local color correction using non-linear masking”, *Proc. IS&T/SID’s Eighth Color Imaging Conference* (IS&T, Springfield, VA, 2000) pp. 108–111.
[11] P. Barten, “Evaluation of subjective image quality with the square-root integral method”, J. Opt. Soc. Am. **7**, 2024 (1990).
[12] P. G. Engeldrum, *Psychometric Scaling* (Imcotek, Winchester, MA, 2000) pp. 94–98.
[13] D. H. Sanders and R. K. Smidt, *Statistics: A First Course* (McGraw–Hill, New York, 2000), p. 566.