



Contrast evaluation methods for natural color images in display systems: within- and cross-content evaluations^{*}

Qiao-song CHEN, Choon-woo KIM^{†‡}

(School of Information and Communication Engineering, Inha University, Incheon 402-751, Korea)

[†]E-mail: cwkim@inha.ac.kr

Received Jan. 2, 2011; Revision accepted June 9, 2011; Crosschecked Sept. 28, 2011

Abstract: Contrast evaluation can be used as a criterion to evaluate performance of contrast enhancement algorithms and to compare contrast capability of display systems. This paper deals with contrast evaluation models for natural color images. Two separate models are defined for within- and cross-content evaluations. The former is to differentiate the perceived contrast of the images with the same content. The latter is to discriminate the differences in contrast among the images with different contents. Perception mechanisms are quite different for within- and cross-content evaluations. Local contrast plays an important role in within-content evaluation. In contrast, global contrast dominates the contrast perception for cross-content evaluation. Results of human visual experiments show that the proposed evaluation models outperform previous methods for both within- and cross-content evaluations.

Key words: Contrast evaluation, Global contrast, Local contrast, Within content, Cross content, Statistical significance

doi:10.1631/jzus.C1100004

Document code: A

CLC number: TP317.4

1 Introduction

As various display systems such as liquid crystal display (LCD), plasma display panel (PDP), and organic light-emitting diode (OLED) are introduced into markets, image quality becomes one of the most important performance indicators. There are many factors affecting the quality of a displayed image, including contrast, sharpness, colorfulness, and naturalness. This paper is focused on the quantitative evaluation of contrast on natural color images. The proposed measure of contrast can be used as a criterion to evaluate performance of contrast enhancement algorithms and to compare the contrast capability of various display systems.

Contrast has been referred to as a quality factor that discriminates the content or objects of the image. Generally speaking, image contrast can be divided into two categories: global and local contrasts.

Global contrast indicates a perceived difference in luminance and chroma. Local contrast describes the distinguishable degree of image details. Even though contrast is an important factor in describing image quality, there is no universally accepted definition. Michelson (1927) and Weber (1846) proposed measures of contrast for simple patterns. However, these measures are not suitable for describing the contrast of natural images because they depend only on the maximal and minimal luminance values. In Agaian *et al.* (2007) and Panetta *et al.* (2008), Michelson's or Weber's contrast was calculated for non-overlapping sub-images and averaged. However, utilization of the non-overlapping sub-images cannot completely describe the level of local contrast on natural color images.

Unlike Michelson's or Weber's contrast focusing on simple patterns such as sinusoidal grating, Peli (1990) proposed a local contrast measure for natural images. The ratio of energy in the band-passed image to that in the low-passed image was defined as a local contrast measure to account for human contrast

[‡] Corresponding author

^{*} Project supported by the Inha University Research Grant, Korea
 © Zhejiang University and Springer-Verlag Berlin Heidelberg 2011

sensitivity. Chen *et al.* (2006) used Tenengrad's criterion (Krotkov, 1989), which is a local contrast measure based on the gradient information, to measure performance of contrast enhancement techniques. Matkovic *et al.* (2005) calculated averages of luminance differences between neighboring pixels at various spatial resolutions. The contrast measure was described as a weighted sum of the averaged differences. However, the measures in Krotkov (1989), Peli (1990), and Matkovic *et al.* (2005) are insufficient because a complete description of contrast should include global contrast as well.

Contrast measures that use color information as well as luminance can be found in Calabria and Fairchild (2003b) and Pedersen *et al.* (2008). Pedersen *et al.* (2008) calculated the variances of L^* , a^* , and b^* in the CIE $L^*a^*b^*$ color space. Geometric mean of the calculated variances was proposed as a measure of contrast. Variances of L^* , a^* , and b^* can be regarded as measures of global contrast. Calabria and Fairchild (2003b) selected three attributes to describe both the global and local contrasts. Standard deviations of lightness and chroma were calculated from the original image to describe global contrast. Standard deviation of lightness was obtained from the high-passed image to account for local contrast. The contrast measure was defined as a linear combination of three attributes.

The performance of the contrast measure can be examined by comparing the calculated measure with the results from human visual experiments. Calabria and Fairchild (2003b) and Pedersen *et al.* (2008) used Pearson's correlation, Spearman rank correlation, and the root mean square error (RMSE) to evaluate the performance of contrast measure. A high value of correlation or a low value of RMSE implies that the contrast measure faithfully matches the visual perception. There are two types of evaluation methods: within-content and cross-content evaluations. The former is to differentiate the images with the same content exhibiting different levels of contrasts. The latter is to discriminate the differences in contrast among the images with different contents. In most previous works, within-content evaluation was employed for performance evaluation. Works on cross-content evaluation are scarce. Pedersen *et al.* (2008) used 15 images having different contents in human visual experiments. For each image, observ-

ers were asked to give a score representing the perceived degree of contrast without a reference image. Cross-content evaluation plays an important role in many applications. For example, content-dependent contrast enhancement algorithms have been proposed to improve the image quality of digital TVs (Kim *et al.*, 1999). In this application, it is desirable to maintain the same level of the enhanced contrast for consecutive frames having different contents.

The objective of this paper is to propose efficient contrast evaluation models that faithfully match the visual perception. Attributes describing global and local contrasts are selected based on contrast perception. Since the brightest and darkest areas on the image are often regions of interest in contrast perception, a dynamic range of luminance is chosen to describe global contrast. Also, the standard deviation of luminance is chosen as a complementary attribute to overcome the limitation of the dynamic range. These two attributes are combined to represent the global contrast in luminance. It is known that the level of perceived contrast is enhanced as the average chroma value of image increases (Calabria and Fairchild, 2003a). Thus, the average dynamic range of chroma is selected as the attribute to describe global contrast in chroma. Finally, gradient information is used to measure the level of local contrast because it describes the change of luminance and chroma in image details.

Visual perception mechanisms for within- and cross-content evaluations are quite different. Generally speaking, within-content evaluation can be described as a 'full-reference comparison'. In other words, contrast perception is achieved by comparing the same content or objects. Cross-content evaluation is more complicated. It can be regarded as a 'no-reference comparison'. Observers have no reference to compare when evaluating the contrast of the images with different contents. In this case, the brightest and darkest areas on the image are the most likely regions of interest.

In this paper we propose two separate contrast evaluation models. Both of them are defined as a linear combination of the aforementioned attributes. Coefficients of the linear combination are determined by applying the linear regression to the subjective scores of human visual experiments and objective values of the attributes. Training images for

model generation are selected such that the minimal difference in perceived contrast between images should be around 1.5 JND (just noticeable difference) scales (Keelan, 2002; ISO 20462-1:2005). Paired comparison is used for human visual experiments for model generation because one-to-one comparison makes it easier for observers to determine their preference (Gescheider, 1984; Engeldrum, 2000).

To verify the performance of the proposed evaluation models, separate human visual experiments are performed using test images. In the visual experiments for model evaluation, both paired comparison and category judgment (Bartleson, 1984; Engeldrum, 2000) are employed.

2 Attributes for contrast evaluation

In this paper, attributes describing global and local contrasts are selected based on contrast perception. Their linear combinations serve as the proposed contrast evaluation models. Each of the selected attributes is formulated in the CIE $L^*a^*b^*$ color space because it is a perceptually uniform system (Fairchild, 2005). ‘Perceptually uniform’ means that changes of the same scale in visual perception would yield the same changes in the color space. To make the selected attributes reflect human visual perception, the CIE $L^*a^*b^*$ color space is used.

2.1 Attributes for global contrast in luminance

In this subsection, attributes specifying the perceived difference in luminance are described. The brightest and darkest areas on the image play an important role in the perception of global contrast in luminance. Thus, the dynamic range of L^* (referred to as DRL) is chosen as an attribute.

DRL should be calculated carefully to match visual perception. First, details of the image should be excluded because the global contrast depends on the overall impression of the image. In this study, a simple average filter of size 2×2 is applied to remove the high frequency contents. Second, having very few pixels with extreme gray levels would not affect the perceived global contrast in luminance. Existence of these pixels will make a low contrast image have a large value of DRL. They should be excluded in the calculation of DRL. Fig. 1a illustrates a night sky

image with shining stars and lamplights. Fig. 1c shows the histogram of L^* of the image in Fig. 1a. Due to the existence of bright pixels of the stars and lamplights, DRL would have a large value even though the image in Fig. 1a is perceived as an image of low contrast. In this study, pixels with the low probability of occurrences in the histogram of L^* are excluded in the calculation of DRL. The number of pixels to be removed is set to 0.2% of the total pixels. Fig. 1b illustrates the image after removing pixels with the low probabilities. In Fig. 1b, it can be observed that the pixels of shining stars and lamplights disappear. Also, a considerable reduction in the value of DRL can be observed in Fig. 1d. In summary, the average filter is applied to the original image and then the pixels with the low probability of occurrences are removed. DRL is then determined as the difference between the maximum and minimum values of L^* .

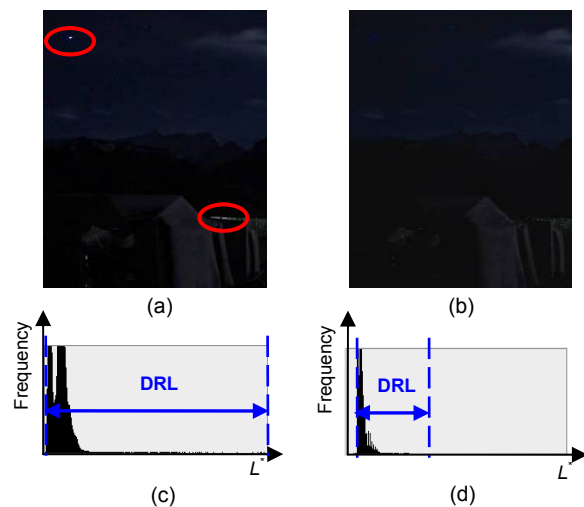


Fig. 1 Influence of pixels with extreme gray levels
(a) Original image; (b) Image after removing pixels with the low probabilities; (c) Histogram of (a); (d) Histogram of (b). DRL: dynamic range of L^*

However, DRL alone is insufficient for fully specifying the global contrast in luminance. For example, even though the images in Figs. 2a and 2b have the same value of DRL, the image in Fig. 2b is perceived as a higher contrast image due to its heavier accumulations in black and white areas. In this study, the standard deviation of L^* (referred to as SDL) is selected as the complement to DRL. For two images with the same value of DRL, the image with

heavier accumulations in black and white areas will have the larger value of SDL. In the example given by Fig. 2, the value of SDL for the image in Fig. 2a is 27.0, and the SDL of the image in Fig. 2b is 31.6.

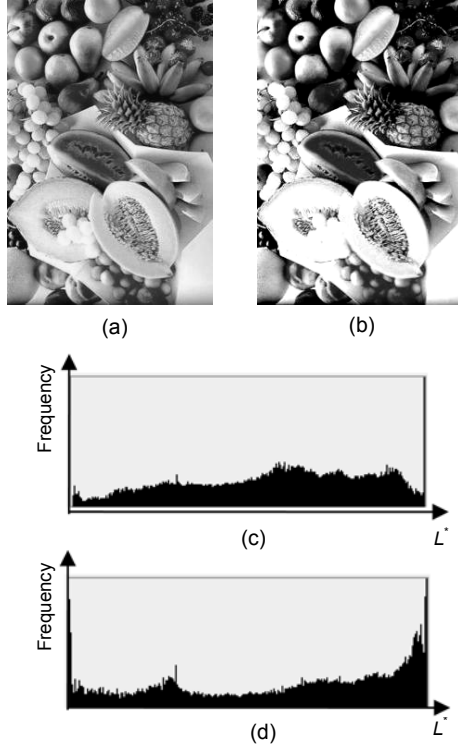


Fig. 2 Example of two images with the same value of dynamic range of L^* (DRL)

(a) Low contrast image; (b) High contrast image; (c) Histogram of (a); (d) Histogram of (b). The values of the standard deviation of L^* (SDL) are 27.0 and 31.6 for the images in (a) and (b), respectively

Global contrast in luminance is modeled by both DRL and SDL as in Eq. (1):

$$V_{GCL} = \omega_{DRL} \cdot V_{DRL} + \omega_{SDL} \cdot V_{SDL}, \quad (1)$$

where V_{GCL} represents the value of the proposed measure of global contrast in luminance, and ω_{DRL} and ω_{SDL} represent the coefficients for values of DRL and SDL (V_{DRL} and V_{SDL}), respectively.

2.2 Attributes for global contrast in chroma

Experiments performed in Calabria and Fairchild (2003a) indicate that the perceived contrast is enhanced as the average chroma value of the image increases. Thus, the averaged dynamic range of

chroma is selected as an attribute to describe global contrast in chroma. Value of chroma is calculated as the square root of the sum of a^{*2} and b^{*2} in the CIE $L^*a^*b^*$ color space. Unlike the calculation of DRL in the 1D space, calculation of the dynamic range of the chroma should be performed in the 2D space, namely, the a^*-b^* plane. The maximum value of chroma for a given hue can be regarded as the dynamic range of chroma, as illustrated in Fig. 3. The averaged dynamic range of chroma is specified by Eq. (2):

$$V_{GCC} = \frac{1}{N_s} \sum_{i=1}^{N_s} C_{\max_i}, \quad (2)$$

where V_{GCC} represents the value of the proposed measure of global contrast in chroma, N_s the number of hue subsections, and C_{\max_i} the maximum value of chroma for the i th subsection. The effect of various values of N_s will be explained in Section 4.3.

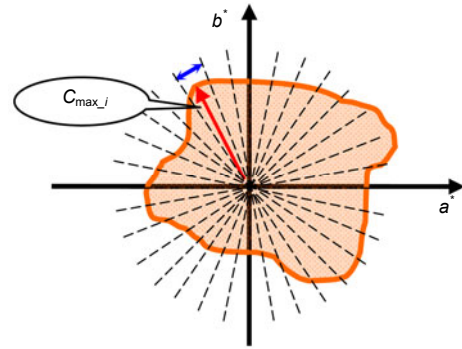


Fig. 3 Calculation of global contrast in chroma

2.3 Attributes for local contrast

Local contrast describes the distinguishable degree of image details. The sum of averaged gradients is selected to describe the local contrast. First, the input image is divided into 3×3 non-overlapping sub-images. Four Sobel masks are applied to each sub-image to calculate the gradients in four directions: vertical, horizontal, 45° , and 135° directions. Fig. 4 shows one of the four Sobel masks to measure the gradients in the vertical direction. For example, the gradient along the left column in Fig. 4 is calculated using Eq. (3):

$$g(i-1) = \{[L^*(i-1, j+1) - L^*(i-1, j-1)]^2 + [a^*(i-1, j+1) - a^*(i-1, j-1)]^2 + [b^*(i-1, j+1) - b^*(i-1, j-1)]^2\}^{1/2}, \quad (3)$$

where $L^*(i, j)$, $a^*(i, j)$, and $b^*(i, j)$ represent CIE $L^*a^*b^*$ values of the pixel at location (i, j) . Similarly, the gradients along the middle and right columns are calculated. These calculations are repeated for every sub-image and then averaged. The same procedure is applied to the remaining three directions. In this paper, V_{LC} represents the value of the proposed measure of local contrast, and is equal to the sum of averaged gradients in the four directions.

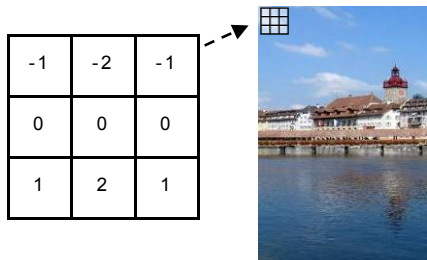


Fig. 4 One of the four Sobel masks used in the calculation of gradients in the vertical direction

3 Proposed contrast evaluation model

Fig. 5 illustrates the flowchart to determine the proposed contrast evaluation models. To design contrast evaluation models that faithfully match the human visual results, attributes are selected based on characteristics of contrast perception. Values of the selected attributes are calculated from each of the training images. Human visual experiments are performed to obtain subjective scores. The proposed contrast evaluation models are generated by applying the linear regression technique to the subjective scores of human visual experiments and objective values of the attributes. The proposed contrast evaluation models are designed as linear combinations of four attributes and a constant as in Eq. (4):

$$PC = (\omega_{DRL} \cdot V_{DRL} + \omega_{SDL} \cdot V_{SDL}) + \omega_{GCC} \cdot V_{GCC} + \omega_{LC} \cdot V_{LC} + C, \quad (4)$$

where PC represents the perceived measure of contrast. The weighted sum of V_{DRL} and V_{SDL} in the parentheses specifies the value of global contrast in luminance. V_{GCC} and V_{LC} describe the global contrast in chroma and local contrast, respectively. ω_{DRL} , ω_{SDL} , ω_{GCC} , and ω_{LC} are the coefficients. C is a constant to ensure that the range of PC is the same as that of the subjective scores.

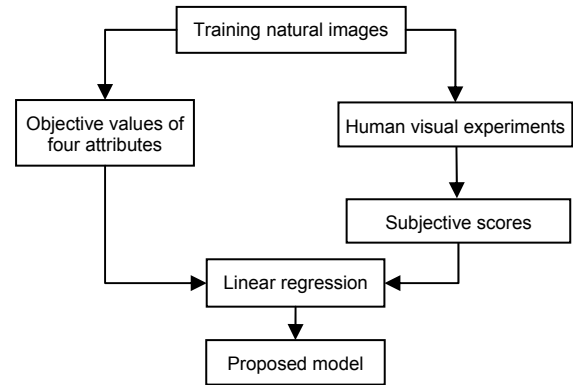


Fig. 5 Flowchart to determine the proposed contrast evaluation models

Two separate contrast evaluation models are constructed: within- and cross-content evaluations of contrast. Thus, two different human visual experiments are performed. Within-content evaluation is used to differentiate the degrees of contrast of the images with the same content (Fig. 6a). Cross-content evaluation is used to discriminate the differences in contrast among the images with different contents (Fig. 6b).

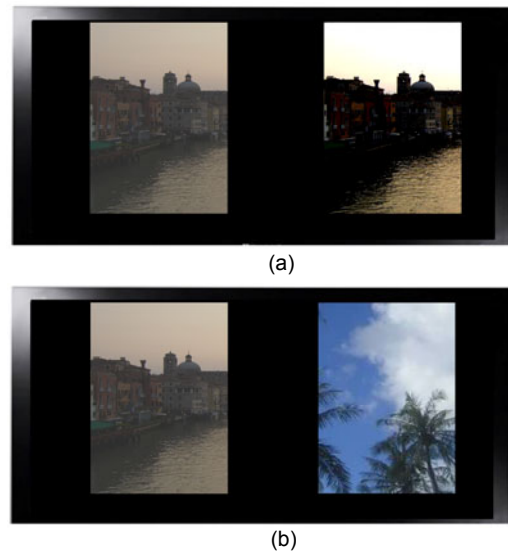


Fig. 6 Within-content (a) and cross-content (b) evaluations

Conditions common to both the within- and cross-content evaluations are described next. In a dark room, human visual experiments are carried out on a standard sRGB LCD monitor with LED backlight. Paired comparison is employed because one-to-one comparison makes it easier to determine

the preference (Gescheider, 1984; Engeldrum, 2000). When the number of training images is n , each observer will make C_n^2 visual comparisons. Twelve observers participated and were asked to choose the image giving the impression of higher contrast. If unable to distinguish the difference, they were allowed to say that two images had the same perceived contrast. The preferred image receives a score of 1 and the other receives a score of 0. When judged as the same level of contrast, both images receive a score of 0.5. Subjective scores of human visual experiments are calculated by the Case V of Thurstone's law of comparative judgments (Thurstone, 1927). Eq. (5) is used for conversion from the outcomes of the paired comparisons to JND values (Keelan, 2002):

$$JND = \frac{12}{\pi} \arcsin \sqrt{p} - 3, \quad (5)$$

where p is the probability of preference in the paired comparisons. Suppose that image A is compared with image B by 10 observers. When seven observers give a score of 1 to image A, one observer gives 0.5, and two observers give 0, p for image A becomes $7.5/10=0.75$.

3.1 Proposed model for within-content evaluation

Fig. 7 illustrates the training images selected for within-content evaluation. To perform human visual experiments for within-content contrast, a set of images representing different levels of contrast needs to be generated for each of the training images. It was reported in Calabria and Fairchild (2003a) that image lightness, chroma, and sharpness affect the perceived contrast. In this study, histogram stretch and compression are applied to generate images with different levels of lightness. Variations in chroma are achieved by multiplying a constant with the chroma of the original image. Unsharp masking is used to generate images with different levels of sharpness.

The selection of training images for within-content evaluation is explained by the example in Fig. 8. Suppose that $C(1)$ to $C(8)$ in Fig. 8a represent variations of the 'Blue' image, arranged in ascending order of the perceived contrast. Fig. 8b lists the JND matrix obtained from paired comparison. When JND between $C(i)$ and $C(i+1)$ is too small, observers would not be able to discriminate the perceived dif-

ference. It was reported in ISO 20462-1:2005 that the difference in JND between $C(i)$ and $C(i+1)$ should be around 1.5 JND scales. In Fig. 8b, the number in the

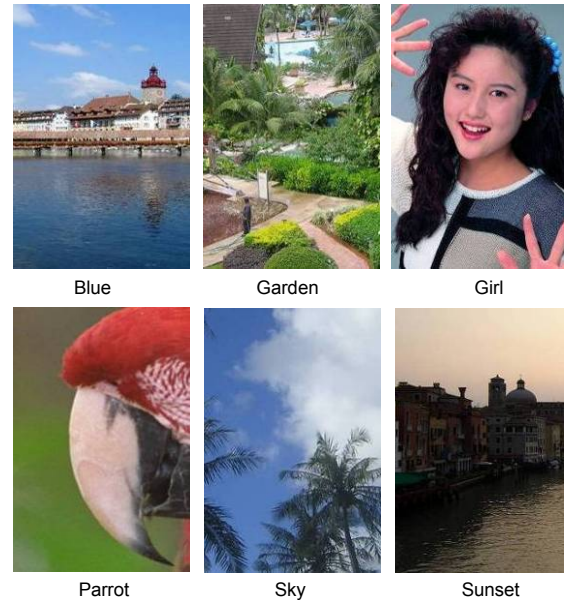
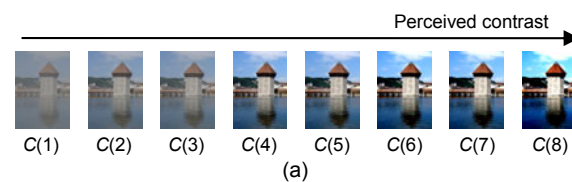


Fig. 7 Six training images for within-content evaluation 'Blue' contains both low and high frequency content and represents an image of high colorfulness. 'Garden' represents high spatial details with high colorfulness. 'Girl' is a portrait image that contains both dark and bright content against bright background. 'Parrot' and 'Sky' consist of relatively few details with a medium level of colorfulness. 'Sunset' represents both low and high spatial frequency content in a dark scene



	C(1)	C(2)	C(3)	C(4)	C(5)	C(6)	C(7)	C(8)
C(1)	0	1.23	2.32	3	3	3	3	3
C(2)	-1.23	0	0.42	1.39	3	3	3	3
C(3)	-2.32	-0.42	0	1.39	2.32	3	3	3
C(4)	-3	-1.39	-1.39	0	1.87	1.56	2.32	3
C(5)	-3	-3	-2.32	-1.87	0	1.39	2.32	3
C(6)	-3	-3	-3	-1.56	-1.39	0	0.68	1.56
C(7)	-3	-3	-3	-2.32	-2.32	-0.68	0	1.39
C(8)	-3	-3	-3	-3	-3	-1.56	-1.39	0

Fig. 8 Example for the selection of training images in within-content evaluation

(a) Eight generated images of 'Blue'; (b) Example of human visual experiment results. The selected images are shaded in the 1st row in (b). Also, the JND values examined for training image selection are shaded

$(i+1)$ th row and $(i+2)$ th column of the matrix represents the JND difference between $C(i)$ and $C(i+1)$. For example, 1.23 in the 2nd row and 3rd column represents the JND difference between $C(1)$ and $C(2)$. $C(1)$ is selected because it is the first image. $C(2)$ is also selected because the JND difference between $C(1)$ and $C(2)$ is 1.23. However, $C(3)$ is not chosen because the JND difference between $C(2)$ and $C(3)$ is 0.42. Because $C(3)$ is not chosen, the JND difference between $C(2)$ and $C(4)$ is examined next to decide whether to select $C(4)$. In this example, $C(4)$ is selected because the JND difference between $C(2)$ and $C(4)$ is 1.39. This procedure is repeated to select appropriate training images for visual experiments.

In this study, 16 images are generated to represent different levels of contrast for each of the six training images in Fig. 7. Paired comparison is performed using 16 images, among which 12 images satisfying the aforementioned condition for the JND value are selected. Results from paired comparison using 12 within-content images are employed for model generation.

The proposed contrast evaluation model for within-content is generated by applying the linear regression technique to the subjective scores of human visual experiments and objective values of four attributes. Contrast evaluation models specified in Eq. (4) are generated for each of the six images in Fig. 7. Coefficients of the six generated models are averaged to yield Eq. (6):

$$PC_{\text{within}} = (0.0330V_{\text{DRL}} + 0.1490V_{\text{SDL}}) + 0.1563V_{\text{GCC}} + 0.6426V_{\text{LC}} - 13.2181, \quad (6)$$

where PC_{within} represents the proposed model for within-content evaluation. In Eq. (6), the attribute representing the local contrast V_{LC} has the largest weight. This may be due to the fact that image details are the most likely regions of interest for within-content evaluation. The value of constant in Eq. (6) is selected to maintain the ranges of PC_{within} and JND would be the same.

Probability of statistical significance (Fisher, 1925), the P -value, is calculated to examine the level of significance for the attributes used in the linear regression. Desirable attributes should have a P -value less than 0.05 (Fisher, 1925). In addition to the proposed model, single image perceived contrast (SIP)

(Calabria and Fairchild, 2003b) defined as a linear combination of three attributes is considered in the calculation of P -value. Note that the SIP model is generated using the same training images selected in this study for fair comparison. Table 1 lists the calculated P -values. For the proposed model, all of P -values are less than 0.05 except for the ‘Sunset’ image, where P -values for V_{GCC} and V_{LC} are greater than 0.05. This may be due to the fact that this image is darker than other training images. However, P -values for three attributes used for SIP, standard deviations of lightness, chroma, and high-passed lightness, σ_{L} , σ_{C} , and σ_{Lh} , are greater than 0.05 in many cases.

Table 1 P -values for the six training images of within-content evaluation

Image	Proposed measure			SIP		
	V_{GCL}	V_{GCC}	V_{LC}	σ_{L}	σ_{C}	σ_{Lh}
Blue	0.0009	0.0071	0.0002	0.0047	0.5736	0.0381
Garden	0.0079	0.0003	0.0213	0.0077	0.0109	0.2553
Girl	0.0001	0.0270	0.0001	0.0001	0.1079	0.0017
Parrot	0.0001	0.0153	0.0067	0.0001	0.4870	0.1800
Sky	0.0001	0.0058	0.0052	0.0010	0.2680	0.0405
Sunset	0.0001	0.1154	0.3313	0.0001	0.1763	0.2785

V_{GCL} represents the value of global contrast in luminance in Eq. (1); V_{GCC} and V_{LC} describe the global contrast in chroma and local contrast, respectively. σ_{L} , σ_{C} , and σ_{Lh} represent the standard deviations of lightness, chroma, and high-passed lightness, respectively. Bold fonts indicate the P -values that are smaller than 0.05

3.2 Proposed model for cross-content evaluation

Fig. 9 illustrates 16 training images selected for cross-content evaluation. For each of the 16 images in Fig. 9, 16 images representing different levels of contrast are generated by varying the lightness, chroma, and sharpness. Thus, 256 images are available for human visual experiments. To impose the same condition for the JND as in the within-content evaluation, the number of paired comparisons becomes $C_{256}^2=32640$. An alternative procedure is adopted in this study to indirectly satisfy the condition for the JND to extract reliable results from human visual experiments. Suppose that $C(i, j)$ ($i=1, 2, \dots, 16; j=1, 2, \dots, 16$) denote 256 generated images. Index i represents 16 images in Fig. 9 and index j represents 16 variations of the i th image in Fig. 9. The value of PC_{within} defined in Eq. (6) is calculated for all 256 images. $C(i, j)$ are arranged such that the

values of PC_{within} for $C(i, 1)$ ($i=1, 2, \dots, 16$) are in ascending order. Also, for a given index i , $C(i, j)$ ($j=1, 2, \dots, 16$) are arranged such that the values of PC_{within} are in ascending order.

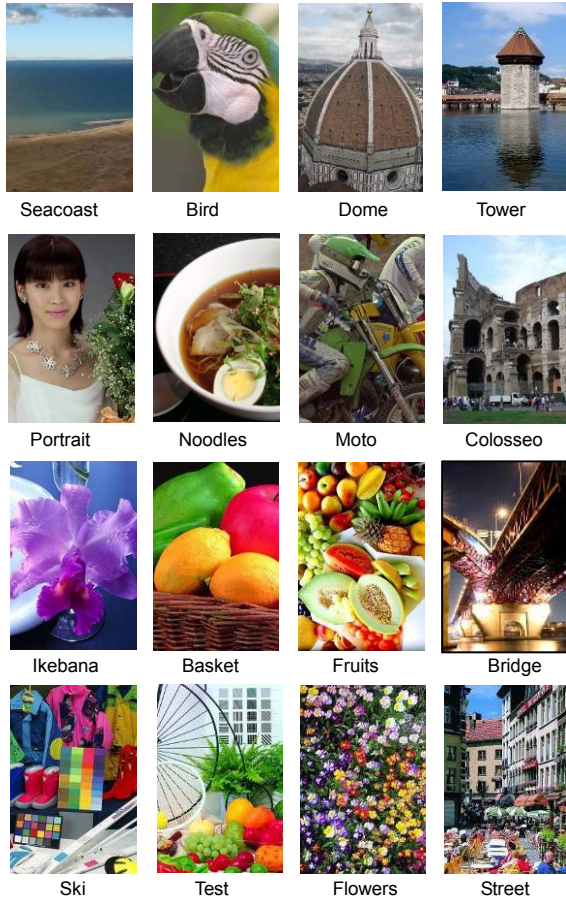


Fig. 9 Sixteen training images for cross-content evaluation

The procedure to select 16 images from $C(i, j)$ ($i=1, 2, \dots, 16; j=1, 2, \dots, 16$) is explained using an example given in Fig. 10. For simplicity, the ranges of i and j are assumed to be 1 to 8. Fig. 10a illustrates the values of PC_{within} for $C(1, j)$ and $C(2, j)$ ($j=1, 2, \dots, 8$). $C(1, 1)$ represents the image with the smallest value of PC_{within} . It is selected as the first image to be used for human visual experiments. Differences between $C(1, 1)$ and $C(2, j)$ ($j=1, 2, \dots, 8$) are calculated. When the JND difference is greater than 1 and smaller than 3, the corresponding $C(2, j)$ is selected as the next training image for human visual experiments. When none of $C(2, j)$ ($j=1, 2, \dots, 8$) satisfy this condition, image generation by varying the lightness, chroma, and sharpness is repeated. Similarly, the remaining sample images are selected for human visual experiments.

In this study, 16 images are generated to represent different levels of contrast for each of the images in Fig. 9. Among the 256 images of $C(i, j)$ ($i=1, 2, \dots, 16; j=1, 2, \dots, 16$), 16 images are selected based on the values of PC_{within} using the aforementioned procedure. They serve as training images for model generation.

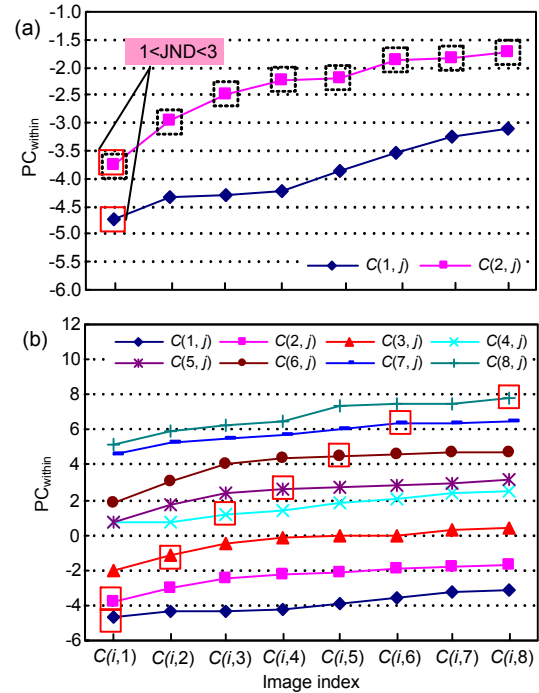


Fig. 10 Example for the selection of training images in cross-content evaluation

(a) Selection of training images for $C(1, j)$ and $C(2, j)$; (b) Selection of eight training images. Solid squares represent the selected images satisfying the JND condition in this example; dashed squares represent the candidates in $C(2, j)$ before image selection

The following model for cross-content evaluation is generated by applying the linear regression technique to the subjective scores of human visual experiments and objective values of four attributes:

$$PC_{cross} = (0.0523V_{DRL} + 0.0284V_{SDL}) + 0.0292V_{GCC} + 0.0430V_{LC} - 7.1876, \quad (7)$$

where PC_{cross} represents the proposed measure for cross-content evaluation. The weight of V_{LC} in Eq. (7) is much smaller than that in Eq. (6). This may be due to the fact that image details are not the major regions of interest for cross-content evaluation because there is no reference. However, the brightest and darkest

areas on images are the most likely regions of interest for cross-content evaluation. This may be the reason why the weight of V_{DRL} in Eq. (7) is much greater than that in Eq. (6).

Table 2 lists the calculated P -values for the attributes of the proposed evaluation model and SIP. Again, note that the SIP model is generated using the same training images selected in this study for fair comparison. For the proposed model, the P -values for V_{GCL} and V_{GCC} are less than 0.05, but the P -value for V_{LC} is greater than 0.05. However, all P -values for three attributes in SIP are greater than 0.05.

Table 2 P -values for 16 training images of cross-content evaluation

Proposed measure	P	SIP	P
V_{GCL}	0.0013	σ_L	0.0736
V_{GCC}	0.0349	σ_C	0.0915
V_{LC}	0.0912	σ_{Lh}	0.1859

V_{GCL} represents the value of global contrast in luminance in Eq. (1); V_{GCC} and V_{LC} describe the global contrast in chroma and local contrast, respectively. σ_L , σ_C , and σ_{Lh} represent the standard deviations of lightness, chroma, and high-passed lightness, respectively. Bold fonts indicate the P -values that are smaller than 0.05

4 Performance evaluation

Separate visual experiments were performed independently using test images. In addition to paired comparison, the category judgment method (Bartleson, 1984; Engeldrum, 2000) was employed for the visual experiments. For the category judgment method, observers were asked to classify the image into one of the seven ordered categories based on their impression of contrast. The seven categories were labeled as highest, very high, high, moderate, mild, very little, and least, representing different levels of contrast perception, and assigned scores of 7 to 1, from highest to least. The averaged score serves as the result of category judgment for the test image.

To compare the performance of the proposed models, five different contrast assessment models were used, including the Tenengrad criterion (TEN) (Krotkov, 1989), measure of enhancement (EME) (Agaian et al., 2007), the global contrast factor (GCF) (Matkovic et al., 2005), SIP (Calabria and Fairchild, 2003b), and $L^*a^*b^*$ -variance (Pedersen et al., 2008). TEN is a local contrast measure based on gradient

information (Krotkov, 1989). EME is defined as the average of Weber's contrast calculated for non-overlapping sub-images (Agaian et al., 2007). GCF is defined as the weighted sum of averages of luminance differences between neighboring pixels calculated at various image resolutions (Matkovic et al., 2005). SIP is defined as a linear combination of standard deviations of lightness, chroma, and high-passed lightness (Calabria and Fairchild, 2003b). $L^*a^*b^*$ -variation is specified by the geometric mean of three variances of L^* , a^* , and b^* in the CIE $L^*a^*b^*$ color space (Pedersen et al., 2008).

The performance of contrast assessment models was evaluated according to the correlation coefficient between the subjective scores from the human visual experiments using test images and the calculated values of the models.

4.1 Performance of within-content evaluation

The 16 test images illustrated in Fig. 11 were used in visual experiments for model evaluation.

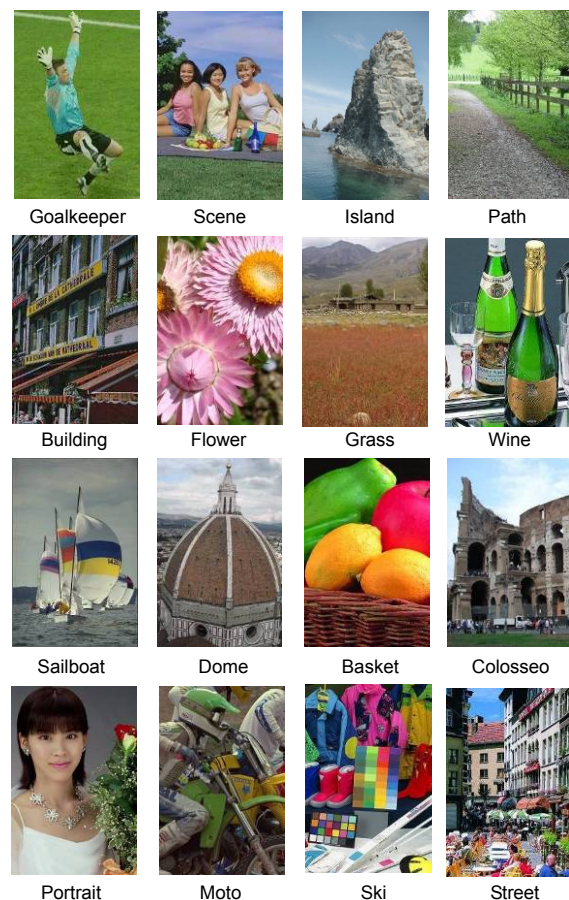


Fig. 11 Sixteen test images for within-content evaluation

4.1.1 Performance evaluation based on paired comparison

Subjective scores from human visual experiments based on paired comparison were calculated by the Case V of Thurstone’s law of comparative judgments (Thurstone, 1927). The correlation coefficients calculated for test images are listed in Table 3. The proposed measure PC_{within} achieves a higher value of correlation coefficient R_{within} compared with previous methods employed in the experiments. Moreover, there are no significant variations in the values of R_{within} by the proposed measure. However, previous works do not provide consistent values of R_{within} for all 16 test images.

Table 3 Model evaluation for within-content based on paired comparison

Test image	TEN	EME	GCF	SIP	$L^*a^*b^*$ -variance	PC_{within}
Goalkeeper	0.86	0.88	0.88	0.91	0.82	0.98
Scene	0.82	0.81	0.83	0.93	0.72	0.97
Island	0.86	0.84	0.87	0.97	0.95	0.98
Path	0.84	0.81	0.84	0.87	0.89	0.92
Building	0.88	0.81	0.89	0.91	0.92	0.92
Flower	0.68	0.73	0.71	0.76	0.90	0.97
Grass	0.81	0.83	0.83	0.90	0.80	0.93
Wine	0.74	0.80	0.75	0.88	0.97	0.98
Sailboat	0.82	0.83	0.84	0.92	0.96	0.98
Dome	0.82	0.82	0.86	0.90	0.91	0.94
Basket	0.66	0.78	0.67	0.82	0.80	0.92
Colosseo	0.80	0.83	0.84	0.88	0.89	0.94
Portrait	0.83	0.91	0.85	0.81	0.88	0.91
Moto	0.81	0.81	0.83	0.90	0.96	0.97
Ski	0.74	0.81	0.76	0.95	0.85	0.96
Street	0.84	0.89	0.85	0.95	0.96	0.97
Average	0.80	0.82	0.82	0.89	0.89	0.95

4.1.2 Performance evaluation based on category judgment

In addition to paired comparison, the results from category judgment were used to verify the performance of the proposed model for within-content evaluation. The correlation coefficients calculated for test images are listed in Table 4. The proposed measure PC_{within} achieves a higher value of R_{within} compared with previous methods used in the experiments.

4.2 Performance of cross-content evaluation

The 16 test images illustrated in Fig. 12 were used in visual experiments for model evaluation.

Table 4 Model evaluation for within-content based on category judgment

Test image	TEN	EME	GCF	SIP	$L^*a^*b^*$ -variance	PC_{within}
Goalkeeper	0.89	0.92	0.91	0.91	0.73	0.94
Scene	0.82	0.82	0.84	0.91	0.67	0.94
Island	0.78	0.77	0.80	0.96	0.95	0.95
Path	0.84	0.81	0.83	0.89	0.94	0.93
Building	0.88	0.81	0.88	0.95	0.98	0.96
Flower	0.74	0.78	0.77	0.77	0.85	0.95
Grass	0.74	0.76	0.77	0.92	0.83	0.93
Wine	0.73	0.79	0.74	0.90	0.94	0.96
Sailboat	0.81	0.83	0.83	0.88	0.93	0.95
Dome	0.79	0.79	0.82	0.85	0.90	0.92
Basket	0.64	0.74	0.64	0.85	0.80	0.90
Colosseo	0.80	0.83	0.84	0.84	0.84	0.91
Portrait	0.74	0.81	0.76	0.92	0.91	0.92
Moto	0.81	0.82	0.81	0.87	0.95	0.94
Ski	0.75	0.81	0.77	0.95	0.85	0.95
Street	0.83	0.86	0.84	0.93	0.96	0.96
Average	0.79	0.81	0.80	0.89	0.88	0.94

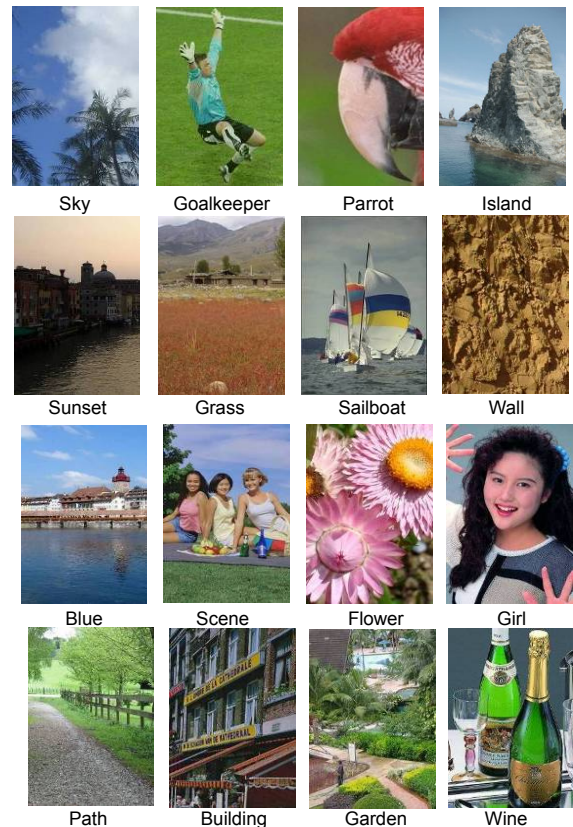


Fig. 12 Sixteen test images for cross-content evaluation

4.2.1 Performance evaluation based on paired comparison

Fig. 13 illustrates the relationships between the objective measures and subjective scores from human visual experiments using test images. Unlike previous measures in Figs. 13a–13e, the proposed measure in Eq. (7) provides an approximately linear relationship with subjective scores, as illustrated in Fig. 13f.

The performance of six different contrast assessment methods can also be verified by calculating the correlation coefficients R_{cross} between the subjective scores from the human visual experiments and the values of objective measures. Table 5 shows the results.

Compared with the within-content evaluation in Table 3, the values of correlation coefficients based on the previous methods are considerably reduced for cross-content evaluation. Drop in the values of the correlation coefficients for previous methods may be due to the fact that observers mostly rely on global contrast rather than local contrast when evaluating images with different contents. Among the five previous works employed in the experiments, SIP yields the highest correlation coefficients of 0.74 because three attributes of SIP can describe both the global and local contrasts. In contrast, the proposed model

for cross-content evaluation yields a correlation coefficient of 0.90. It outperforms all previous works employed in the experiments.

Table 5 Model evaluation for cross-content based on paired comparison

Image	JND	TEN	EME	GCF	SIP	$L^*a^*b^*$ -variance	PC_{cross}
Sky	-2.54	747620	4.83	0.49	1.99	20.89	-3.32
Goalkeeper	-1.90	475382	5.69	0.65	1.82	16.65	-2.88
Parrot	-1.56	115543	3.53	0.39	1.98	29.39	-1.94
Island	-0.12	1950505	9.81	1.01	2.85	23.45	-2.16
Sunset	-1.05	1067899	14.21	0.58	3.24	30.90	-1.23
Grass	-1.34	798451	7.82	0.71	1.74	20.95	-1.41
Sailboat	-0.74	1029266	6.70	0.69	2.37	19.49	-1.22
Wall	0.28	3143653	19.10	1.54	2.43	21.17	-1.79
Blue	1.29	1099854	8.54	0.68	2.42	25.44	-0.32
Scene	-0.01	1367712	9.26	0.88	2.69	29.95	-1.29
Flower	1.66	1449847	8.65	1.05	2.76	34.33	-0.16
Girl	1.63	2123913	29.63	1.26	4.41	37.33	0.20
Path	0.93	4307378	20.21	2.02	3.50	31.18	-0.12
Building	0.18	3485316	20.85	1.57	2.98	29.06	-1.33
Garden	1.50	6139658	30.54	2.39	3.73	32.78	0.45
Wine	1.81	5102878	40.84	2.40	5.26	41.07	0.77
R_{cross}		0.67	0.70	0.71	0.74	0.72	0.90

Smaller JND values indicate lower levels of the perceived contrast. The third to seventh columns contain the calculated measures defined in previous methods. The last column shows the values of PC_{cross} defined in Eq. (7). R_{cross} is the correlation coefficient for cross-content evaluation

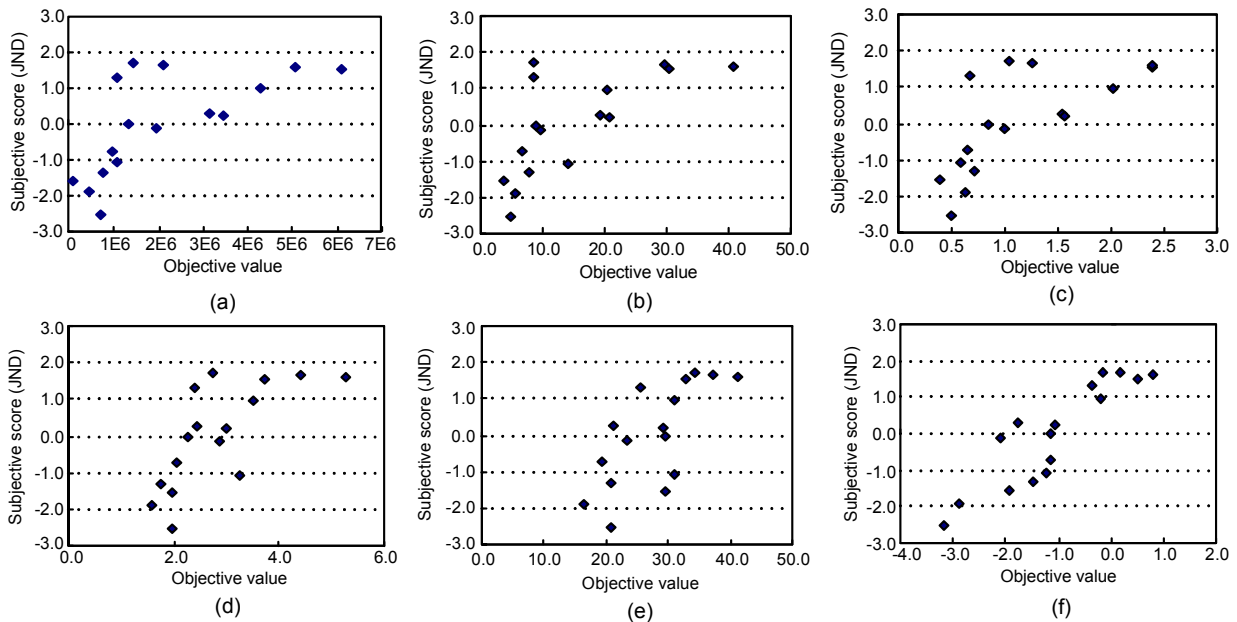


Fig. 13 Objective measures vs. subjective scores for 16 test images of cross-content evaluation
 (a) TEN; (b) EME; (c) GCF; (d) SIP; (e) $L^*a^*b^*$ -variance; (f) Proposed measure

4.2.2 Performance evaluation based on category judgment

The results from category judgment are also used to verify the performance of the proposed model for cross-content evaluation. The correlation coefficients calculated for test images are listed in Table 6. The proposed measure PC_{cross} outperforms those of previous methods used in the experiments with a much higher value of R_{cross} .

Table 6 Model evaluation for cross-content based on category judgment

Image	Degree	TEN	EME	GCF	SIP	$L^*a^*b^*$ -variance	PC_{cross}
Sky	2.42	747 620	4.83	0.49	1.99	20.89	-3.32
Goalkeeper	2.83	475 382	5.69	0.65	1.82	16.65	-2.88
Parrot	3.50	115 543	3.53	0.39	1.98	29.39	-1.94
Island	4.33	1950 505	9.81	1.01	2.85	23.45	-2.16
Sunset	3.75	1067 899	14.21	0.58	3.24	30.90	-1.23
Grass	3.58	798 451	7.82	0.71	1.74	20.95	-1.41
Sailboat	4.33	1029 266	6.70	0.69	2.37	19.49	-1.22
Wall	5.00	3143 653	19.10	1.54	2.43	21.17	-1.79
Blue	5.83	1099 854	8.54	0.68	2.42	25.44	-0.32
Scene	5.08	1367 712	9.26	0.88	2.69	29.95	-1.29
Flower	5.92	1449 847	8.65	1.05	2.76	34.33	-0.16
Girl	6.00	2123 913	29.63	1.26	4.41	37.33	0.20
Path	5.67	4307 378	20.21	2.02	3.50	31.18	-0.12
Building	4.92	3485 316	20.85	1.57	2.98	29.06	-1.33
Garden	5.83	6139 658	30.54	2.39	3.73	32.78	0.45
Wine	6.42	5102 878	40.84	2.40	5.26	41.07	0.77
R_{cross}		0.66	0.70	0.71	0.73	0.73	0.91

The second column represents the results of category judgment. Smaller values indicate lower levels of the perceived contrast

To summarize the results of experiments, Table 7 lists the correlation coefficients, R_{within} and R_{cross} , for within- and cross-content evaluations. Paired comparison and category judgment result in similar values of correlation coefficients.

Table 7 Comparison of R_{within} and R_{cross}

	R_{within}	R_{cross}
Paired comparison	0.95	0.90
Category judgment	0.94	0.91

4.3 Effect of hue subsections

To investigate the effect of hue subsection to calculate V_{GCC} in Eq. (2), the correlation coefficients R_{GCC} between V_{GCC} and the results of human visual

experiments using test images were calculated for various numbers of hue subsections (N_s). Fig. 14 illustrates the results of R_{GCC} for five different values of N_s . There are no significant differences in the values of R_{GCC} when N_s is greater than 8. N_s is set to 90 in this study.

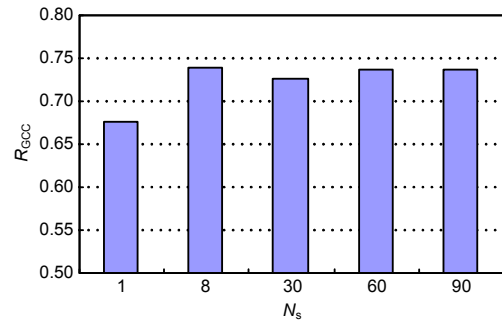


Fig. 14 Relationship between the number of hue subsections (N_s) and the correlation coefficient between V_{GCC} and the results of human visual experiments (R_{GCC})

5 Conclusions

This paper deals with new contrast evaluation models that faithfully match visual perception. Attributes describing global and local contrasts are selected based on the characteristics of contrast perception. Global contrast indicates a perceived difference in luminance and chroma. Local contrast describes the distinguishable degree of image details. Two separate contrast evaluation models are proposed: within- and cross-content evaluations. The proposed models are defined as the linear combinations of the selected attributes. Local contrast plays an important role in within-content evaluation. However, global contrast dominates in cross-content evaluation where there is no reference image. Results of the experiments show that the proposed models outperform previous works, especially for cross-content evaluation.

References

Agaian, S., Silver, B., Panetta, K., 2007. Transform coefficient histogram based image enhancement algorithms using contrast entropy. *IEEE Trans. Image Process.*, **16**(3):741-758. [doi:10.1109/TIP.2006.888338]

Bartleson, C.J., 1984. Optical Radiation Measurements. Academic Press, New York.

Calabria, A.J., Fairchild, M.D., 2003a. Perceived image contrast and observer preference I. The effects of lightness, chroma, and sharpness manipulations on contrast

- perception. *J. Imag. Sci. Technol.*, **47**(6):479-493.
- Calabria, A.J., Fairchild, M.D., 2003b. Perceived image contrast and observer preference II. Empirical modeling of perceived image contrast and observer preference data. *J. Imag. Sci. Technol.*, **47**(6):494-508.
- Chen, Z.Y., Abidi, B.R., Page, D.L., Abidi, M.A., 2006. Gray-Level Grouping (GLG): an automatic method for optimized image contrast enhancement - part I: the basic method. *IEEE Trans. Image Process.*, **15**(8):2290-2302. [doi:10.1109/TIP.2006.875204]
- Engeldrum, P.G., 2000. Psychometric Scaling: a Toolkit for Image Systems Development. Imcotek Press, Winchester.
- Fairchild, M.D., 2005. Color Appearance Models. John Wiley and Sons, Ltd., West Sussex.
- Fisher, R.A., 1925. Statistical Methods for Research Workers. Oliver and Boyd Ltd., London.
- Gescheider, G.A., 1984. Psychophysics: Method, Theory, and Application. Lawrence Erlbaum, Hillsdale.
- ISO 20462-1:2005. Photography - Psychophysical Experimental Methods for Estimating Image Quality - Part 1: Overview of Psychophysical Elements. International Organization for Standardization, Geneva.
- Keelan, B.W., 2002. Handbook of Image Quality: Characterization and Prediction. CRC Press, New York. [doi:10.1201/9780203910825]
- Kim, S.Y., Han, D., Choi, S.J., Park, J.S., 1999. Image contrast enhancement based on the piecewise-linear approximation of CDF. *IEEE Trans. Consum. Electron.*, **45**(3):828-834. [doi:10.1109/30.793618]
- Krotkov, E.P., 1989. Active Computer Vision by Cooperative Focus and Stereo. Springer-Verlag, New York.
- Matkovic, K., Neumann, L., Neumann, A., Psik, T., Purgathofer, W., 2005. Global Contrast Factor—a New Approach to Image Contrast. Computational Aesthetics in Graphics, Visualization and Imaging, p.159-168.
- Michelson, A.A., 1927. Studies in Optics. University of Chicago Press, Chicago.
- Panetta, K., Wharton, E., Agaian, S., 2008. Human visual system-based image enhancement and logarithmic contrast measure. *IEEE Trans. Syst. Man Cybern. Part B*, **38**(1):174-188. [doi:10.1109/TSMCB.2007.909440]
- Pedersen, M., Rizzi, A., Hardeberg, J.Y., Simone, G., 2008. Evaluation of Contrast Measures in Relation to Observers Perceived Contrast. 4th European Conf. on Color in Graphics, Imaging, and Vision and 10th Int. Symp. on Multispectral Color Science, p.253-258.
- Peli, E., 1990. Contrast in complex image. *J. Opt. Soc. Am. A*, **7**(10):2032-2040. [doi:10.1364/JOSAA.7.002032]
- Thurstone, L.L., 1927. A law of comparative judgment. *Psychol. Rev.*, **34**(4):273-286. [doi:10.1037/h0070288]
- Weber, E.H., 1846. Tastsinn und das Gemeingefühl. Wagner's Handwörterbuch der Physiologie, Braunschweig (in German).



www.zju.edu.cn/jzus; www.springerlink.com

Editor-in-Chief: Yun-he PAN

ISSN 1869-1951 (Print), ISSN 1869-196X (Online), monthly

Journal of Zhejiang University

SCIENCE C (Computers & Electronics)

JZUS-C has been covered by SCI-E since 2010

Online submission: <http://www.editorialmanager.com/zusc/>

Welcome Your Contributions to JZUS-C

Journal of Zhejiang University-SCIENCE C (Computers & Electronics), split from *Journal of Zhejiang University-SCIENCE A*, covers research in Computer Science, Electrical and Electronic Engineering, Information Sciences, Automation, Control, Telecommunications, as well as Applied Mathematics related to Computer Science. JZUS-C has been accepted by Science Citation Index-Expanded (SCI-E), Ei Compendex, INSPEC, DBLP, Scopus, IC, JST, CSA, etc. Warmly and sincerely welcome scientists all over the world to contribute Reviews, Articles, Science Letters, Reports, Technical notes, Communications, and Commentaries.