https://doi.org/10.2352/issn.2169-2629.2021.29.170

Models to predict naturalness and image quality for images containing three memory colors: sky, grass, and skin

Jason Ji[1], Dalin Tian[1], Ming Ronnier Luo[*1]
[1]State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou, China
****m.r.luo@zju.edu.cn***

Abstract

When evaluating the image quality, people mostly would like to concentrate on the color appearance of memory objects, representing the naturalness and reality of the image scene. Generally, an image with objects which have perfect memory colors reproduction will give natural and harmonious feelings. Many previous studies had proved the critical role of naturalness in image quality assessment, but it was still tough to scale the image naturalness precisely. In this study, natural images with blue sky, green grass, and skin colors were selected and partially rendered to develop the model of preference and naturalness of typical memory colors. A psychophysical experiment was conducted to collect the visual data of these images. Afterward, the psychophysical data were used to build the preference models and naturalness models, respectively. The models were then compared with previous studies. Results showed that the new models could accurately predict the preference and naturalness of target memory colors.

Introduction

The rapid evolution of digital imaging media provides people tremendous joy in observing the world, from CRT display to LCD to nowadays OLED display, QD-LCD, etc. The image quality is always highly desired by customers, no matter at which stage. Therefore, investigating the factors that affect image quality has been the research focus for a long time. In our previous study, seven visual image attributes together with overall image quality were evaluated on an OLED display [1]. An image quality model for this OLED display was then proposed, and the importance of those image attributes was compared based on principle components analysis. Results showed that in color-related attributes, colorfulness and reality were more effective to image quality. Gong et al. had tested the performance of various image attributes on the smartphone [2]. An image quality model including naturalness, colorfulness, brightness, and clearness was proposed. Results showed that naturalness and clearness were the most correlated attributes for image quality in natural scenes. Choi et al. developed a reference image quality model based on the experiment that dealt with a test display and a reference display [3]. Naturalness, colorfulness, and contrast were the major components in her model.

It is understandable from the above research that the naturalness effect on overall image quality was unparalleled, but in their work, there are no perfect naturalness models that could work well in any database. Gong's naturalness model was based on Yendrikhovskij's theory [4]. They clarified that memory colors are pretty consistent among different observers, and people might prefer to see an image in which the color appearance of a familiar object agrees with its memory color. Gong first utilized three color centers (sky, grass, and skin) based on their theory to detect the memory color areas. Afterward, she averaged the color differences with reference color centers as the degree of color shift, and an exponential formula was designed to linearize the model. This model could predict the naturalness in her database. However, in Choi's research, she firstly analyzed this naturalness model with her dataset [5]. She found that due to the overall tone shift of different images, sometimes the model predictions delivered an opposite trend compared with visual data. Therefore, she proposed a new naturalness model based on three components, i.e., reproduction of shadow detail, colorfulness, and sharpness. Zeng et al. collected a large set of memory colors from several natural scenes [6]. He used ellipsoids in CIELAB space to cover most of the memory colors. The conclusion was that those colors outside the ellipsoids could not be regarded as memory colors. Nevertheless, those ellipsoids could not predict naturalness, while the naturalness did not distribute as the probabilities of memory colors.

Above all, there has not been a satisfactory model to predict image naturalness. Therefore, the goal of the present study is to develop generic naturalness models based on a memory color image database.

Development of database

To investigate the trend of the naturalness distribution, a database including a set of images with three memory colors, i.e., sky, grass, and skin, was firstly generated. For each memory color, ten images were selected from the MIT-Adobe FiveK image database [7]. In total, 30 images were chosen. Note that only original images in this database were selected because those retouched by experts may lose their naturalness, so selected images were all-natural scenes without any preprocessing. Fig.1 shows part of the selected original images.

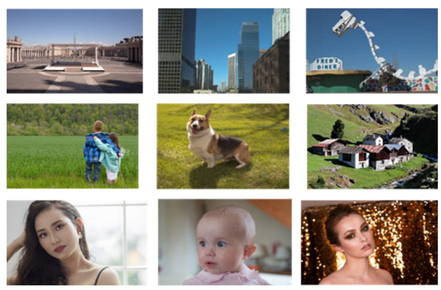

Figure 1. *Part of selected original images for sky, grass, and skin.*

Subsequently, original images were partially rendered to form the image database. Then, to precisely focus on each unique memory color's effect, the memory object areas in each image were cropped, and the rest was not processed. Fig.2 shows an example of a cropped image.

Figure 2. A sky original image and its cropped image.

The centers of these memory color pixels in CIELAB space were calculated. Around the center of each image, three circles in L*-a*, L*-b*, and a*-b* plane were generated. On the circumference, eight points were uniformly selected as the rendering targets. Considering if the rendering intensity was too small to distinguish their difference, the naturalness variation could not cover the best and worst circumstances, the radii of these circles were set to ensure exactly three of eight target points on each circumference were outside Zeng's memory color ellipsoids. Once these radii were decided, another set of circles with the same center and half radii were also generated, and eight target points on this circumference were selected. Afterward, the memory colors in each original image were translated to move the center of the memory color. This manipulation would push many pixels to the color gamut for some saturated images, leading to the color deviation. In this case, the target points of those images were specially adjusted to make sure less than 5% pixels exceed the color gamut. Fig.3 shows the example of one original image and its related renderings in a*-b* plane.

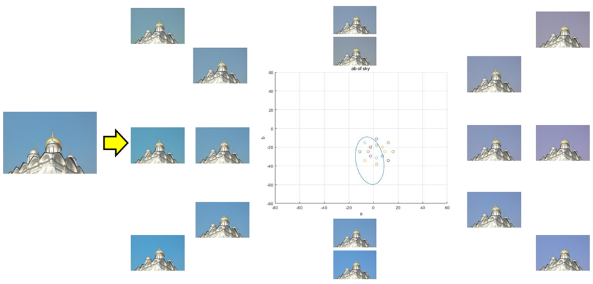

Figure 3. the original image and its 16 renderings in a-b* plane. The ellipse in right-center is Zeng's model and the points are the rendering targets.*

Consequently, for each image, (8 + 8) * 3 = 48 targets were selected around the original center. Thus, in total, there were 1470 images ((48 target points + 1 origin point) * 30 original images) generated and included in this dataset.

Psychophysical experiments

A psychophysical experiment was carried out on mobile phones because the mobile images on mobile phones were one of the most frequently observed in daily life. Four smartphones of the same model with OLED screens were used in this experiment. Their physical size was 6.1-inch, and their spatial resolution was 2240 pixels by 1080 pixels. It could achieve a peak luminance of around 400 cd/m^2. The automatic adjustment of brightness in these phones was switched off. According to the measurement, the average color difference among these phones was 1.8 Δ^*E_{ab}, which indicated that these four cellphones had almost the same color appearance and could be used simultaneously for this experiment. Four GOG models were developed for each phone to characterize these phones, and the one with the best prediction for all cellphones was finally selected. The prediction error of this model to all cellphones was less than 4 Δ^*E_{ab}, which was acceptable in this study. All measurements were taken by a tele-spectroradiometer (TSR), Konica Minolta CS-2000A Spectroradiometer.

Twenty observers aged from 20 to 28 participated in this experiment. They were trained about the whole procedure before the experiments. All observers undertook the Ishihara Test for the color defect to ensure normal color vision before carrying out this experiment. The smartphones were laid 25 centimeters in front of the observers while the light source was in the top, which agreed with CIE 0°/45°observing geometry. The environment light was accurately set with a 5000K correlated color temperature (CCT) and a 500lux luminance. Fig.4 shows the experimental environment.

Figure 4. The experiment situation

A forced-choice method was adopted in this experiment. Each observer was asked to evaluate naturalness and image quality in two separate sessions. The naturalness was defined as how close this image was to the real world, and the image quality was defined as how much this image was preferred. A more preferred image represents a higher image quality, so the below image quality also refers to preference. For each phase, observers were asked to view and estimate a displayed image in naturalness or image quality. Then they should choose natural/unnatural or preferred/not preferred. Each original image, together with its 48 renderings, was arranged in a random order for each observer. The name of memory objects in the current scene, the number of unevaluated images, and the current psychophysical scale were hinted at in the top left of the interface. Fig.5 shows the user interface.

Considering that the grass colors in the real world are so variant in different seasons, it is necessary to repeat the grass colors twice for modeling the naturalness and image quality of grass individually for spring and fall. Thus, four memory color groups were evaluated: sky, skin, spring grass, and autumn grass. Besides, 40 images were selected as repeats from all rendered images to test the observers' stability. Overall, (49 renderings * 10 original images * 4 memory colour groups + 40 repeats) * 2 psychophysical scales

Figure 5. The user interface on the smartphone

= 4000 judgements were made by each observer, and in total 4000 * 20 = 80000 sets of data were collected.

The visual results were transformed to percentage N% for each image, representing the probability that this image was assessed to be natural/preferred. For instance, if 15 out of 20 observers thought an image looked natural and the other 5 observers thought it looked unnatural, the final naturalness score N% for this image was 75%. Image quality scores were calculated with the same manipulation. These scores will be used in the following data analysis.

Data Analysis

Observer variation

Intra-observer variation was evaluated in terms of repeatability to represent the agreement of observers' twice assessment. For a perfect agreement, the repeatability should be 100%. A repeatability of 70% means a 30% variation between the two sets of data. The observers' data that had a repeatability lower than 50% would be removed to promise stability. The best and worst observers gave 91% and 62% repeatability, respectively, with a mean of 83%. For the inter-observer variation, the mean color difference of the mean (MCDM) was used here to reflect the agreement among different observers. Observers' selected memory colors were averaged, and the color difference between them was calculated. The mean value of four memory color groups' MCDM was 10.2 $\Delta^* E_{ab}$. These values indicate that the experimental results were quite reasonable compared with the previous study.

Correlation between naturalness and image quality

The correlation between naturalness and image quality was firstly investigated. Fig.6 shows the scatter plots and the correlation coefficients of four memory color groups. Results show that for sky and skin groups, the naturalness and image quality strongly correlate with each other, which agrees with previous studies. Furthermore, in the spring grass group, the correlation is stronger than the autumn grass group. This result indicates that for some grass colors, people recognize them as autumn grass while they prefer spring grass rather than them.

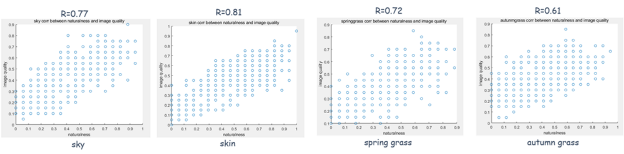

***Figure 6**. The correlation between naturalness and image quality for each memory colour group.*

Naturalness and image quality models

Individual models for each original image were developed to clarify the unity among original images in the same memory color group. The percentage *N%* was related to the modified color difference (ΔE'), calculated using the ellipsoid equation's L*, a*, b* of each image. Equation 1 shows the calculation procedure.

$$N\% = \frac{1}{1 + e^{(\Delta E' - \alpha)}} \quad (1)$$

and

$$\Delta E' = k_1 * \sqrt{(l - l_0)^2 + k_2 * (a - a_0)^2 + k_3 * (b - b_0)^2 + k_4 * (a - a_0)(b - b_0)}$$

Where L_0, a_0, b_0 represent the most natural or preferred color center in CIELAB color space. This modified color difference (*ΔE'*) could be regarded as the color shift from the most natural or preferred color center, representing the level of unnatural or un-prefer. The eight coefficients in Equation 1, including color center [L_0, a_0, b_0] and coefficients k_i (i = 1, 2, 3, 4) together with α were optimized by maximizing the correlation coefficient between the *ΔE'* values in Equation 1 and *N%* from the visual results.

Ten ellipsoids related to 10 original images were fitted for each memory color group using the above equation for naturalness and image quality, respectively. Fig.7 shows ten naturalness ellipses and ten image quality ellipses which are the projection of ellipsoid models in a*-b* plane for each memory color group. The ellipsoids parameters are set as N% is precisely 50%. In this condition, the ellipsoid represents the boundary of natural/unnatural or prefer/not prefer. Results indicate that the models are pretty converged for sky, skin, and spring grass, and a generic model for the same memory color images can be well fitted. However, for autumn grass, the models of some images do not follow the same trend. The reason is that for those images, the selected target points could not cover the most natural and preferred area of autumn grass color, so this model could not find the actual color center. Table.1 shows the mean color difference of the mean (MCDM) of each memory group. A larger MCDM represents a more different trend with average. For those images in which the fitted color center would not converge, the MCDM values were replaced by /.

Moreover, for those images which have large MCDM, their values are marked in red. The marked images were removed in the generic model fitting in the next step to fit a good model and get rid of the negative influence. For each memory color group, no more than four images were removed.

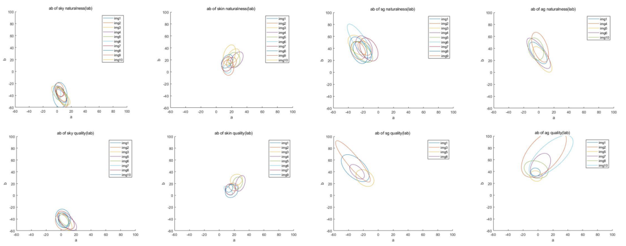

***Figure 7**. The individual model of each memory colour group. Top: naturalness models. Bottom: image quality models. From left to right: sky, skin, spring grass, and autumn grass. For some images, the ellipsoid model retrograded to hyperbola and are not plotted.*

Table 1. The MCDM of the colour center of each individual model.

	Sky (Naturalness)	Sky (Image quality)	Skin (Naturalness)	Skin (Image quality)	Spring grass (Naturalness)	Spring grass (Image quality)	Autumn grass (Naturalness)	Autumn grass (Image quality)
img1	2.9	7.8	7.7	6.7	5.8	7.0	2.9	9.3
img2	7.6	3.9	8.9	8.9	10.7	/	/	28.2
img3	23.6	3.5	12.9	14.4	11.0	18.8	/	22.6
img4	5.5	/	9.8	10.8	3.8	45.3	8.8	44.0
img5	14.8	11.7	8.4	9.8	6.5	19.2	13.0	15.8
img6	8.0	4.4	13.8	11.3	13.8	13.0	3.4	/
img7	7.5	2.5	6.7	8.5	9.8	/	/	7.9
img8	3.0	2.6	5.4	/	6.1	7.8	16.9	5.3
img9	5.7	/	10.5	7.9	10.2	12.9	/	/
img10	6.0	3.6	12.0	/	/	/	11.1	29.0
Mean	8.5	5.0	9.6	9.8	8.6	17.7	9.4	20.3
Standard deviation	6.3	3.2	2.7	2.4	3.2	13.0	5.7	13.2

Generic models for each memory color

According to the previous analysis, generic naturalness and image quality models could be fitted for each memory color group with the above ellipsoid equation. Fig.8 shows the fitted generic model for each memory color group in L*-a*, L*-b*, and a*-b* plane. Zeng's related ellipsoids are also plotted here as a comparison. Table.2 shows the details of each ellipsoid model, i.e., the lightness (L^*), chroma (C_{ab}^*) and hue (h_{ab}^*) of the fitted color center, the length of semi-long axes (A), the ratio between the semi-long and semi-short axes (A/B), the rotation angle of the ellipse in a*-b* plane (θ), the correlation coefficients between prediction and visual data (R), the average MCDM value of the color centers used in obtaining the generic model, and the color difference between the color centers of naturalness and image quality (*ΔE*).

The correlation coefficients R show that the skin colors models are not as accurate as other models (R = 0.68 for image quality and R = 0.76 for naturalness) because, in our previous study [8], the naturalness model of skin colors can be further sorted by the skin types including Chinese, Caucasian, and African, etc. The ten original images selected in this study consist of many different skin colors, so it is reasonable that this model could not accurately predict every image. Results also show a slight difference between the naturalness model and the image quality model for each group. Another novel discovery is that for the sky, spring grass, and autumn grass, the ellipsoid models are more or less out of Zeng's ellipsoid boundary. This result indicates that some colors out of Zeng's sky and grass ellipsoid could be regarded as memory colors and could even be considered natural or preferred.

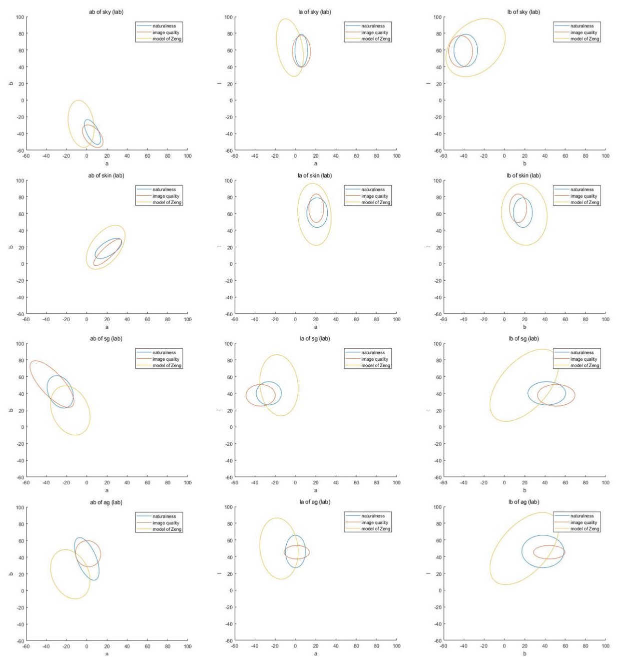

Figure 8. *The generic model for each memory colour group. From top to bottom: sky, skin, spring grass and autumn grass. From left to right: L*-a*, L*-b*, and a*-b* plane.*

Table 2. The details of generic models for each memory colour group.

	Naturalness model								Image quality model								ΔE
	L^*	C_{ab}^*	h_{ab}^*	A	A/B	θ	R	MCDM	L^*	C_{ab}^*	h_{ab}^*	A	A/B	θ	R	MCDM	
sky	59.4	38.7	278.5	10.74	2.70	111.70°	0.85	8.5	58.3	43.7	277.8	15.02	1.85	120.00°	0.82	5.0	5.17
skin	62.4	22.8	47.8	21.12	3.45	84.22°	0.76	9.6	61.3	25.6	40.3	19.41	2.78	71.57°	0.68	9.8	8.08
spring grass	40.1	49.4	122.2	18.98	1.64	106.78°	0.79	8.6	37.7	61.7	124.0	32.71	2.75	125.77°	0.74	17.7	12.66
autumn grass	46.3	37.9	90.2	26.19	2.80	107.90°	0.85	9.4	45.3	44.2	88.3	16.08	1.25	96.77°	0.73	20.3	6.66

Conclusions

In this study, a natural image database was first developed. Then, a psychophysical experiment was conducted to collect the visual assessments of these natural images. Afterward, naturalness and image quality models were developed for four memory colors: sky, skin, spring grass, and autumn grass. Results show that all models can give reasonably accurate predictions.

Acknowledgments

Most of the photos in this paper are courtesy of the MIT-Adobe FiveK Dataset [2011], which was created, described, and shared by Bychkovsky et al. [2011], for which we are grateful.

References

[1] D. Tian & L. Xu & M. R. Luo. (2019). The Image Quality Evaluation of HDR OLED Display. Electronic Imaging. 309-1. 10.2352/ISSN.2470-1173.2019.10.IQSP-309.

[2] R. Gong, H. Xu, M. R. Luo. (2015) Comprehensive model for predicting perceptual image quality of smart mobile devices[J]. Applied optics, 2015, 54(1): 85-95.

[3] S. Y. Choi, M. R. Luo, M. R. Pointer, and P. A. Rhodes. (2008) "Investigation of large display color image appearance I: important factors affecting perceived quality," J. Imaging Sci. Technol. 52, 040904.

[4] YENDRIKHOVSKIJ S.N., RIDDER H.DE, BLOMMAERT F.J.J. Color Reproduction and the Naturalness Constraint[J]. Color Research and Application,1999, 24(1):52-67.

[5] S. Y. Choi, M. R. Luo, M. R. Pointer and P. A. Rhodes, Investigation of large display colour image appearance. III: Modelling Image Naturalness, J. Imaging Sci. Tech., 53 (2009) 031104-1-12.

[6] H. Z. Zeng and M. R. Luo, Skin Colour Modelling of Digital Photographic Images, Journal of Imaging Science and Technology, (2011) 55(3): 030201 – 1-12.

[7] BYCHKOVSKY, V., PARIS, S., CHAN, E., AND DURAND, F. 2011. Learning photographic global tonal adjustment with a database of input / output image pairs. In The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition.

[8] Tian D., Peng R., Chang C., Luo M.R. (2021) Models to Predict Naturalness of Images Including Skin Colours. In: Zhao P., Ye Z., Xu M., Yang L., Zhang L., Zhu R. (eds) Advances in Graphic Communication, Printing and Packaging Technology and Materials. Lecture Notes in Electrical Engineering, vol 754. Springer, Singapore. https://doi.org/10.1007/978-981-16-0503-1_5