Colour and Tolerance of Preferred Skin Colours

Huanzhao Zeng and Ronnier Luo**, *Hewlett-Packard Company, Vancouver, WA, USA **University of Leeds, Leeds, UK*

Abstract

Skin tone has been an active research subject in preferred colour reproduction of photographic images. There is a consistent conclusion that preferred skin colours are different from actual skin colours. However, results about preferred skin colours from different reports are not consistent. To enhance skin colours of photographic images reliably, psychophysical experiments were conducted to determine preferred skin colours for preferred colour reproduction. Inter-observer variation and tolerance of preferred skin colours were analyzed as well.

1. Introduction

People, particularly facial patterns, are often the primary visual subject in a colour image. Reproducing skin colours pleasingly is a critical task in photographic colour reproduction. Since people typically rely on their memory of object colours to judge the quality of memory colour reproduction [1, 2], it is important to know the preferred colour region of skin colours instead of actual skin colours for preference colour reproduction.

Bartleson investigated the preferred colour reproduction of skin colours and found that the actual colour and the memory colour were significantly different in chroma, and the preferred flesh tone appeared to be yellower and higher chroma than the colours of actual flesh tone. Bartleson and Bray [3] further investigated the preferred colour reproduction of flesh tone, blue-sky, and green-grass. They found that the preferred colour for reproduced Caucasian complexions is to be the same chromaticness as the mean memory colour of the flesh tone. Sanders [4] studied the colour preference of natural objects, and found that the preferred Caucasian facial colour was more saturated than actual facial colours. Hunt et al. [5] studied the preferred colour reproduction in colour photography, and concluded that Caucasian skin had about the same purities as real skin but was a little yellower for reflection prints. Sanger et al. [6] studied offset printing samples of Mongoloid, Caucasoid, and African skin colours, and found that the African skin colour distribution was wider than the distribution of other races; chroma of skin colours increased steadily in the order of Caucasoid, Mongoloid, and African while the lightness increased in the reverse order; the hues were similar and the dominant wavelength was about 590nm in all distributions. Yano and Hashimoto [7] studied the preference of Japanese complexion, and found that the preferred complexion of Japanese women was shifted to a slightly higher saturation and was more reddish in hue than the actual complexion of Japanese women; the direction of hue shift is different from that of the preferred Caucasian women; and the saturation shift of Caucasian women is 3.5 times as large as that of the Japanese woman. Park et al [8] studied preferred skin colour reproduction on display. The skin colours were determined using a bivariate Gaussian function. Skin colours were moved toward a small colour region around preferred skin colour centres for preference enhancement. Kuang et al. [9] conducted psychophysical experiments to study the influence of different factors on skin colour preference of photographic colour reproduction. They concluded that background lightness has little influence on skin colour preference; capturing illuminant has significant influence on skin colour preference; the preference variances on Oriental and Caucasian skin colours are smaller than those on Indian and African American; and no significant culture difference among different ethnic observers. Bodrogi [10] used photo-realistic images containing sky, plant, and Caucasian skin colours, as well as standalone colour patches taken from the corresponding photo-realistic images to study memory colour shift on a calibrated CRT monitor. Memory colour shift in photo-realistic images was found larger than that in colour patches, and later (long-term) memory colours had higher chroma than the instant memory colour in both photo images and colour patches. For skin, observers' long-term memory colours tended to be yellower than original colours.

There is a consistent conclusion from different reports that preferred skin colours are different from actual skin colours. However, results about preferred skin colour regions from different studies are not consistent. As part of the research for skin colour enhancement of photographic images, psychophysical experiments were conducted to study preferred skin colours, inter-observer variation, and tolerance for preferred colour reproduction.

2. Phase I of the Experiment

The shape of skin colours can be approximated with an elliptical model [11]. To simplify the experiment, the lightness-independent skin ellipse model was applied to detect skin colours. An image database composed of 2500 digital images was used to train the skin model. The images were captured using various digital cameras under different conditions. The subjects cover different skin tones, including Caucasian, Oriental, Hispanics, African, and Indian. The skin model coefficients can be found in reference 11. The skin colour ellipse trained using all skin types is plotted in Fig. 1, where the ellipse is the skin colour boundary and the big orange dot in the centre is the statistical skin colour centre in CIE a*b* diagram.

Psychophysical experiments were conducted to determine preferred skin colour regions for preferred skin colour enhancement. The procedure are: 1) the skin colour region is uniformly sampled to obtain an N-point skin colour set; 2) skin colours of each test image is morphed toward each of these N skin colour points; and 3) observers evaluate the preference of the N versions of each image. A paired comparison method was used for the experiment. The number of judgements on each image are $m = N(N-1)/2$. If skin sample points are large (a large N), m may be too large for practical evaluation. As a compromise, 9 sample points were chosen. The total number of judgements on each image by an observer is 36. To ensure the quality of observations, each observer was to judge images for no more than 30 minutes in a single session. With this constraint, an observer was able to judge 4 images (36x4=144 pairs of comparisons) at a time. 12 images were judged by each observer. The experiment was divided in 3 sessions. Each observer completed a session each day and therefore finished all observations in three days.

Five workstations were used to conduct the experiment simultaneously, each with an HP L2335 Active Matrix TFT 23-inch LCD displays to display images. Each display's white point was adjusted to D65. The room lighting was adjusted to simulate a typical dim surround condition. 19 observers participated this phase of the experiment. All of them have normal colour vision. They were experienced at evaluating image quality because of involvement in colour and image quality evaluations that were conducted regularly. They were diverse in gender and race (Caucasian dominated), and ages were between 25 and 50.

A set of 9 pre-determined colour centres uniformly sampled within the skin colour ellipse in CIE a*b* diagram was used to morph skin colours of test images (see Fig. 1). The centre point #4 is at the skin elliptical centre. Because of a defect in the software to display images for the experiment, each display ICC profile was not applied for displaying images during the experiment. Therefore the nine colour centres are actually shifted to different positions (see Fig. 5).

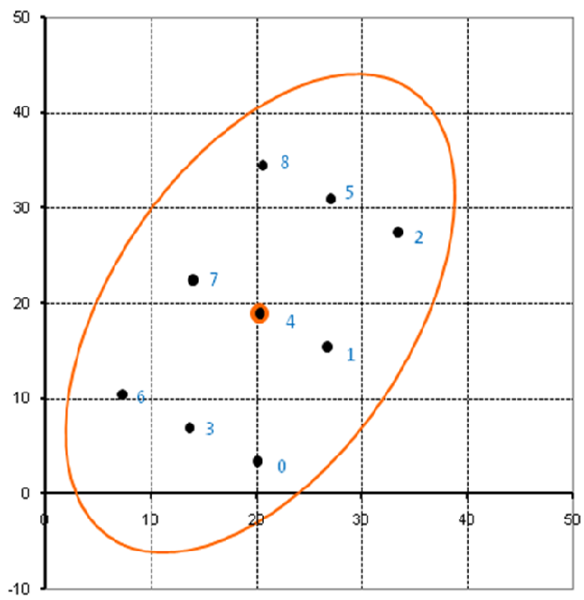


***Fig. 1**. Nine predetermined skin colour centres for the experiment*

Twelve sRGB images for the experiment were carefully selected to cover different skin types. The skin colours of each image were processed in a way that skin colours were morphed toward a skin colour centre. The adjustment gradually faded off for colours from the statistical skin colour centre to the skin colour boundary. Each image was first transformed from its source RGB colour space to L*a*b* using the source RGB ICC profile. The linear Bradford transformation was applied to adapt colours to D50 in ICC profiles [12]. The ellipse skin colour model was applied to compute Mahalanobis distance, $\Phi(X)$. $\Phi(X) = \rho$ defines the skin colour boundary. If $\Phi(X)$ was greater than ρ, a colour X was considered a non-skin colour, and no colour adjustment was applied. Otherwise, a weight was computed from the Mahalanobis distance to morph the skin colour toward each of the nine skin centres. A colour at the ellipse centre was adjusted the most, and colours on or outside the skin colour gamut were not adjusted. With nine predetermined skin colour centres, nine versions of images were created from each image.

2.1 Colour Distribution of Test Images

The 12 test images were carefully selected and visually judged to avoid biases in skin colour distributions. Quantitative analysis were performed to analyze skin colour distributions of the set of images and to compare them with the statistical distribution of skin colours. The purpose is to confirm that the skin colour distribution of the set of image is unbiased.

The skin colours of each image were labelled manually. Then the histogram of skin colours of the 12 images was generated. Two steps were performed to reduce noise in histogram computation: 1) images were de-noised prior to skin colour labelling; and 2) 5% of least occurred skin colours in each labelled skin image were removed. The skin colour histogram was applied to train an ellipsoid that covers 90% of labelled skin colours. Fig. 2 shows the skin colour ellipsoids in CIELAB colour space. The skin colour ellipse of the 12 images is illustrated in orange, and the ellipsoid of skin colours of the image database is drawn in black. Visualization of the 3-D plot and visualization of 2-D slices demonstrate that the orange ellipsoid is in the centre of the skin colour database. It confirms that the skin colour distribution of the 12 images is well balanced. These images are not biased toward a direction that may force observers to select a skin colour centre in an opposite direction to cancel off the colour shift. This avoids biased judgment from imbalanced skin colour distributions of images.

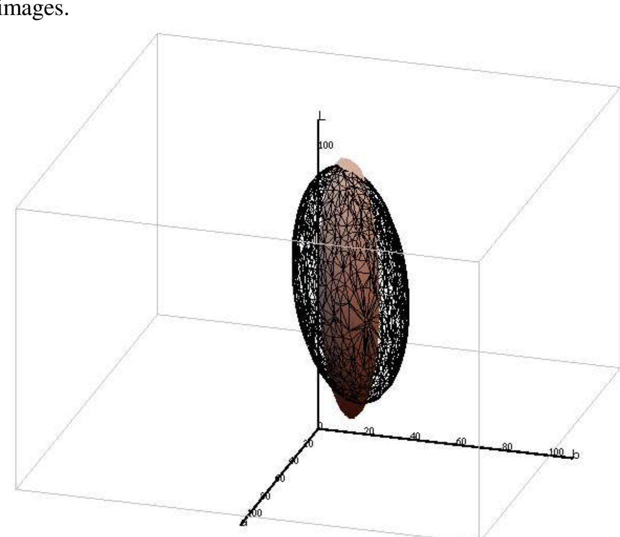


***Fig. 2**. The skin colour ellipsoid of the 12 images (orange) and the skin colour ellipsoid of the image database (black) in CIELAB colour space*

2.2 Preferred Skin Colour Centre

The Thurston's Law of Comparison Case 5 was applied to analyse the result [13]. The z-scores of each image at skin centres 0 to 8 obtained from all judgements are plotted in Fig. 3. A higher z-score means stronger 'prefer' (or less 'dislike'). Error bars represent the 95% confidence interval. The figure shows that skin centres #0 and #2 are least preferred to most images, and #4 and #7 are most preferred to most images.

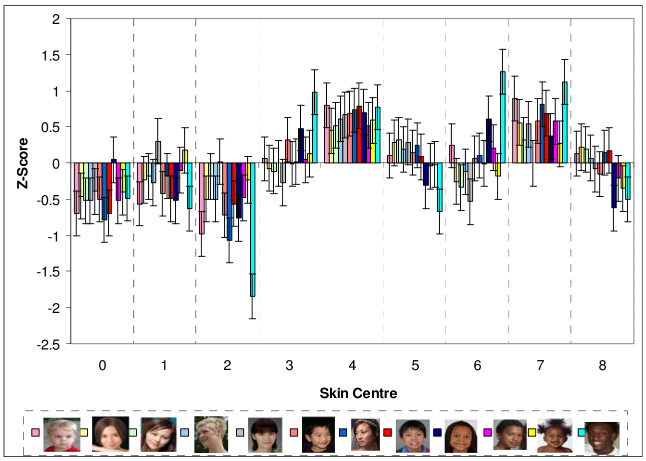


Fig. 3. z-score of each individual image at each skin colour centre

Although Fig. 3 shows the preference of each individual image, the plot is very busy and is difficult to see a general trend. To examine overall preference from all images, the z-scores of all images are averaged and plotted in Fig. 4. The skin colour locations #4 and #7 are most preferred; the locations #0 and #2 are least preferred; the next unfavoured location is #1; and other locations are close to neutral.

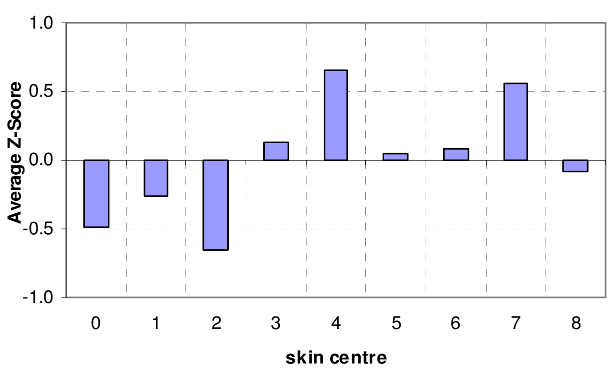


Fig. 4. z-scores averaged at each skin colour centre from all images

Characterising skin colour centres using a representative display ICC profile (an ICC profile that approximately represents the average of all five displays) and using z-scores of skin centres #4 and #7 as weights, a preferred skin colour centre was obtained: a* b* = (20.4, 23.4).

To understand the underlying trends of different skin colours, the skin colours were categorized into light, medium, and dark skin tones. The first five images were categorized as light skin-tone images (Caucasian dominated), the next three images were categorized as medium skin-tone images (Oriental), and the last four images were categorized as dark skin-tone images (African dominated). The average z-scores for these three categories are plotted in Fig. 5. The three groups have following similar preferences:

1) Skin centres #0 and #2 are strongly unfavoured;
2) Skin centre #1 is weakly unfavoured; and
3) Skin centres #4 and #7 are strongly preferred.

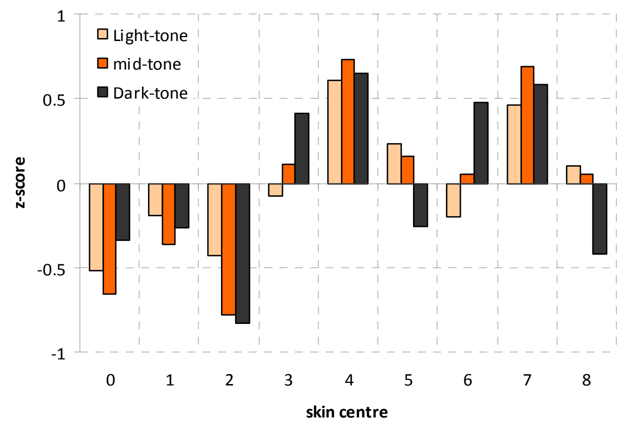


Fig. 5. Average z-scores of three skin categories

The performances of skin centres #3, #5, #6, and #8 among three groups are different. The light and medium skin tones are more similar, while the dark skin tone is more different from the other two. Skin centres #5 and #8 are unfavoured for the dark skin tone, but are weakly preferred or neutral for the light and medium skin tones. Centres #3 and #6 are strongly preferred for the dark skin tone, while they are close to neutral for the light and medium skin tones. Since #5 and #8 are more chromatic than the centre #4, and centres #3 and #6 are less chromatic than centre #4, the result seems to reveal that relative to light and medium skin tones, observers prefer less chromatic dark skin colours.

Fig. 6 shows preferred skin colour ellipse of light, medium, and dark skin colours in CIE a*b* diagram. The orange ellipse is the skin colour boundary, and the large orange dot is the centre of the ellipse. Dots marked with 0 to 8 are the nine skin colour centres used in the experiment. Display colour characterisation using a representative display ICC profile has been applied to characterise skin colour centres. The ellipses of three colour categories were generated using their positive z-scores as weights. Negative z-scores were considered low favoured and their corresponding positions were removed from analysis. Each ellipse was computed independently using a corresponding data set. The overall preferred skin colours are in the area around skin centres #4 and #7. The preferred dark skin colours are in a slightly less chromatic region. The hue preferences among different groups are very close with each other (the preferred hue centre of darker skin colours is very slightly lower).

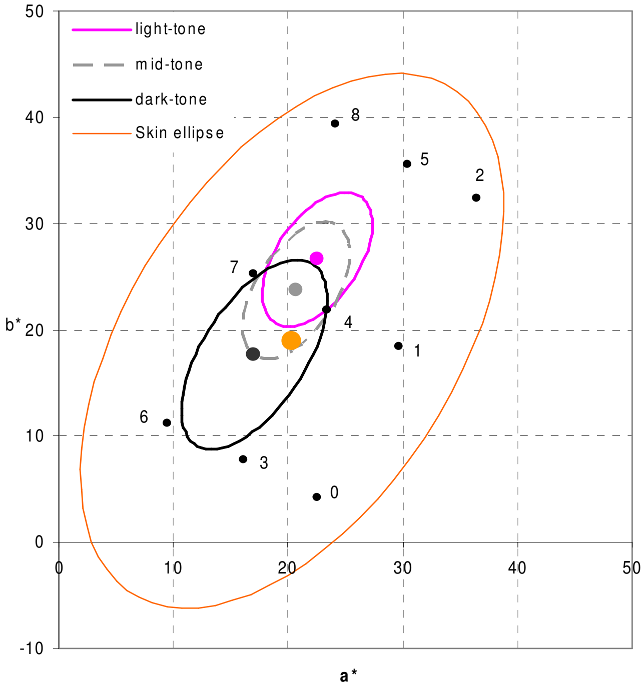


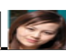
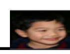
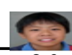
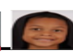
Fig. 6. Preferred skin colour regions of three skin colour groups

The scores of 9 skin centres judged by each observer were taken as weights to compute a set of a*b* that represents a preferred skin colour centre judged by an observer on an image. For each image, colour difference (ΔE^*_{ab}) between each individual preferred skin colour centre and the mean skin colour centre averaged from all 19 observers was computed to analyse inter-observer variation. ΔE^*_{ab} of all 19 observers by 12 images are listed in Table 1. Except for the last column and the last row, each grid represents ΔE^*_{ab} of an observer's judgements on an image. Different observers judged on the same image differently, and each observer judged different images differently as well. The last column is each observer's average ΔE^*_{ab} from all 12 images. The last row is the average ΔE^*_{ab} from all observers on each image. The average ΔE^*_{ab} from all observer judgements over all images is 4.3, ranged from 2.5 to 5.8.

There are 19 observer preferred skin colour centres on each image. Each observer's preferred skin colour centre against the mean skin colour centre of the image, (Δa^*, Δb^*), were computed. In Fig. 7, each point is a preferred skin colour centre of an observer judgement on an image. The spread of 19 points on each image represents inter-observer variances on skin colour preference for this image. An ellipse to cover 85% of 19 points of (Δa^*, Δb^*) from all observer judgements on each image was generated. All ellipses show very long major-axis and very short minor-axis, and have about the same orientation. Translating each ellipse centre to its original a*b, one will see that 1) the orientation of each ellipse lies within a small hue range, implying a small tolerance in preferred hue angles; and 2) their chroma variations are more tolerable.

2.3 Inter-Observer Variation

Table 1 ΔE^*_{ab} between individual preferred skin colour centre and mean preferred skin colour centre

Observer													average
1	7.0	5.4	6.0	2.2	5.6	3.3	6.2	3.6	9.6	7.3	3.9	2.0	5.2
2	5.8	5.6	2.8	6.3	3.4	2.1	1.6	5.7	3.1	5.3	3.0	0.6	3.8
3	6.7	2.3	2.6	2.6	2.5	6.8	6.1	3.1	3.6	5.4	2.0	2.6	3.9
4	5.3	6.7	6.0	2.2	1.9	4.9	5.0	2.2	8.9	7.3	3.8	4.6	4.9
5	4.6	5.1	5.6	6.7	2.7	1.1	0.9	4.9	2.0	1.5	7.6	2.8	3.8
6	1.9	3.9	1.0	5.7	2.0	3.6	3.3	0.9	2.3	1.9	2.0	1.6	2.5
7	2.1	2.8	4.0	8.1	7.3	8.0	4.8	5.5	2.4	4.7	4.8	0.5	4.6
8	1.8	3.9	8.5	7.1	5.9	4.5	6.6	7.1	4.0	7.1	7.3	2.1	5.5
9	6.5	5.0	4.8	6.2	1.8	6.5	7.8	1.5	7.8	7.3	6.8	1.2	5.3
10	1.9	3.3	3.9	6.9	5.9	3.6	1.9	6.9	2.2	6.5	8.2	0.3	4.3
11	1.0	5.2	8.5	3.9	7.2	1.9	4.9	4.9	2.7	2.4	4.6	6.4	4.5
12	4.2	1.5	3.9	2.9	4.4	1.6	5.3	2.0	3.2	6.3	2.4	2.3	3.3
13	5.0	4.7	2.6	4.7	3.7	5.5	4.8	3.3	1.6	5.0	4.2	2.0	3.9
14	5.6	8.5	7.0	4.1	5.3	6.4	7.0	5.6	3.3	6.7	6.8	3.5	5.8
15	5.5	2.7	1.7	6.1	5.9	3.2	1.9	7.5	3.0	2.9	2.8	8.9	4.3
16	6.0	4.1	4.1	3.3	4.4	4.4	5.0	1.3	5.2	7.8	1.2	1.6	4.0
17	6.7	7.8	6.4	3.3	2.4	5.9	4.4	6.7	2.7	3.8	5.3	1.1	4.7
18	2.7	5.0	6.0	6.1	4.3	2.8	1.0	3.9	3.7	1.6	7.2	3.6	4.0
19	6.5	7.8	7.2	1.4	2.8	1.9	2.6	2.9	3.6	5.8	3.2	1.2	3.9
average	4.6	4.8	4.9	4.7	4.2	4.1	4.3	4.2	4.0	5.1	4.6	2.6	4.3

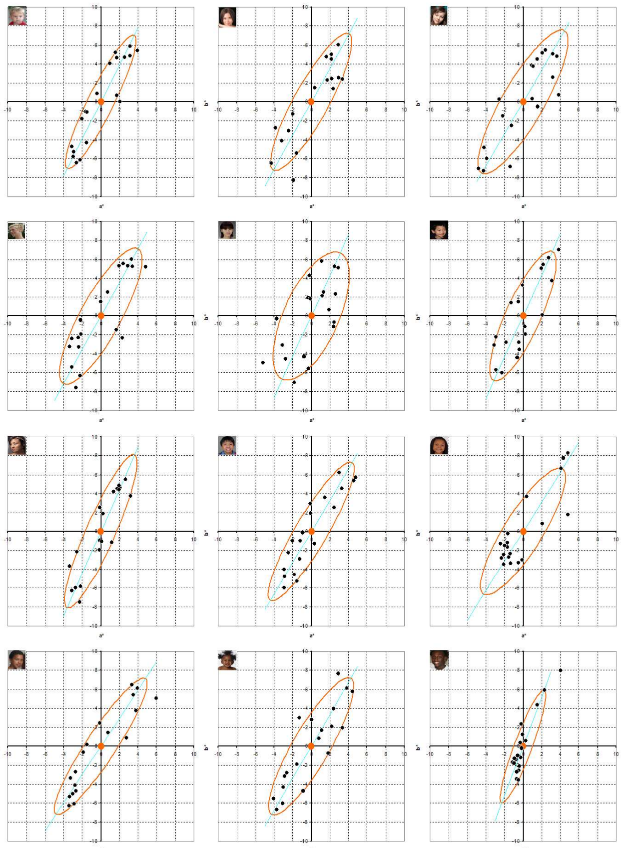


Fig. 7. Variations of observer judgements on each image

There are 228 points (19 by 12) that are the combinations of all observer preferred skin colour centres on all images. An ellipse covering 85% of the points of variances, (Δa^*, Δb^*), was generated (black ellipse in Fig. 8). The ellipse represents the inter-observer variation on skin colour preference. Again, the figure shows that the hue variation is smaller than the chroma variation. The result reveals that skin colour preferences from different observers are more consistent in hue and have a larger variation in chroma. Translating the ellipse centre to absolute a*b* coordinates (translating (0, 0) to the preferred centre (22.2, 22)), one will find that the orientation of the ellipse (the major axis tilts 62^0 off the a* axis) reveals an interesting hue preference: if a chroma higher than the mean chroma is preferred, the preferred hue angle is to be slightly higher than the average preferred hue angle (the hue preference is slightly more yellowish); if a chroma lower than the mean chroma is preferred, the preferred hue angle is to be slightly lower than the average preferred hue angle (the hue preference is slightly more pinkish).

3. Phase II of the Experiment

This above experiment (Phase I) was considered to be the first round of the evaluation. The results reveal a strong preference at an 'ideal' skin colour centre. This experiment was intended to narrow down the preferred skin colour region which would be explored in a finer resolution in a subsequent psychophysical experiment (Phase II). The intention was to produce a more accurate preferred skin colour centre.

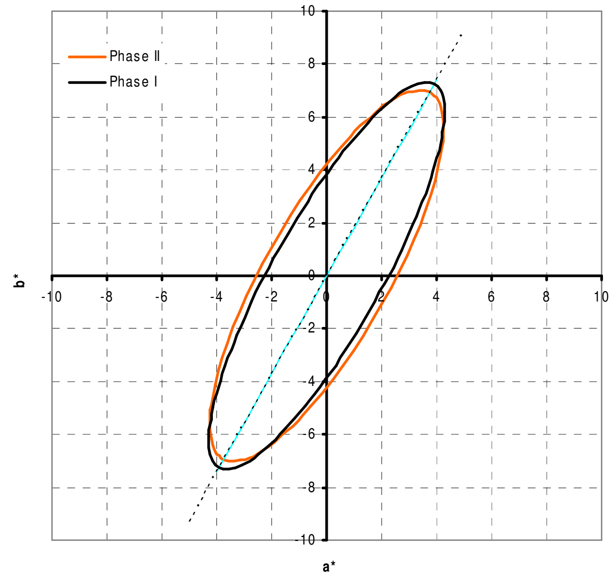


Fig. 8. Variations of all observer judgements on all images

In Phase II, a twice denser sampling of 9 skin colour centres around the preferred skin colour centre determined in Phase I were generated to process test images. A final preferred skin colour centre would be obtained from this experiment, and the inter-observer variation would be verified as well.

Monitor displays, and experimental set-up and procedures are the same as in Phase I. To avoid potential bias toward certain images, a complete different set of 12 images were carefully selected to cover various skin types. Quantitative analysis of the statistical colour distributions of the set of images and the image database verifies that the statistical distribution of the skin colour of the image set is well balanced. 20 observers participated the experiment. Five of them participated Phase I as well. Again, observers are Caucasian dominated.

The preferred skin colour centre judged by each observer on each image against the mean overall preferred skin colour centre was analysed and compared with the result of Phase I. Each observer's ΔE^*_{ab} averaged from all images are in the range of 2.7 and 5.6, which is very close to that in Phase I. The average ΔE^*_{ab} between each individual observer's preferred skin colour centre and the overall mean preferred skin colour centre is 3.9, while this is 4.3 in Phase I. Although the skin colour centres to be tested are twice denser in Phase II, the average observer variation was not reduced by half. Instead, they are almost the same. It is therefore concluded that 4 ΔE^*_{ab} units is about the observer variation of skin colour preference.

The ellipse of the distributions of the preferred skin colour centre offset (Δa^*, Δb^*) of each observer on each image was generated and analysed. Their sizes, shapes, and orientations are similar to those in Phase I. Again, all (Δa^*, Δb^*) of the 20 observer's preferred skin colour centres on 12 images were applied to generate an ellipse that covers 85% of the data points (see the orange ellipse in Fig. 7). The ellipses from both phases are almost the same, and mostly interestingly, the orientations of the two ellipses are the same (two major axes overlap with each other). Although the sampling density of skin colours was doubled in Phase II, test images are different, and observers are different, the variance of observer judgements is not changed. The result demonstrates that the variance of observer judgements is independent of the sampling of skin colour centres for judgement.

A display ICC profile was used to convert display RGB to CIE L*a*b* for computing skin colour centres. The display ICC profile was modelled with a matrix plus R, G, and B tone curves for colour transformation. To improve the colour accuracy, the RGB colour of the final preferred skin colour centre obtained in Phase II with different lightness were displayed on all five displays that were used for the experiment, and their colorimetric values were measured using an EyeOne spectrophotometer. The a*b* of the preferred skin colour were averaged from the measurements. The result a*b* are (21, 24) with a hue angle of 49^0.

Comparisons among the skin colour distribution of the skin image database, the skin measurements of RIT and University of Joensuu, and preferred skin colours of other reports can be found in reference 14.

4. Conclusions

Psychophysical experiments were conducted to evaluate skin colour preference. The preferred skin colour centre for mixed skin colours was found to be a*b* = (21, 24) in D50 illuminant. Preferred skin colours are more chromatic than real skin colours. The observer variance on preferred skin colours is about 4 ΔE^*_{ab} (ignoring L*). The variance may be used to determine a preferred skin colour region around the preferred skin colour centre, however, it should be realized that tolerances in different directions are different. The observer variance is larger in chroma than in hue. A preferred skin colour ellipse found from the experiment is confined within a small hue range and a relatively larger chroma range. The orientation of the ellipse shows that if an individual preferred skin colour is more saturated than the average preferred skin chroma, its hue angle tends to be slightly higher than the average preferred hue angle; on the other hand, if an individual preferred skin colour is less saturated than the average preferred chroma, its hue angle tends to be slightly lower than the average preferred skin hue angle.

Acknowledgement

The authors would like to thank Dr. Jan Morovic for his technical advice.

References

[1] Bartleson, C.J. (1959) “Some Observations on the Reproduction of Flesh Colors”, Photographic Science and Engineering, 3(3): 114-117.
[2] Bartleson, C.J. (1960) “Memory Colors of Familiar Objects”, J. Opt. Soc. Am., 50(1): 73-77.
[3] Bartleson, C.J. and Bray, C. P. (1962) “On the preferred Reproduction of Flesh, Blue-Sky, and Green-Grass Colors”, Photographic Science and Engineering, 6(1): 19-25.
[4] Sanders, C.L. (1959), “Color Preference for Natural Objects”, Illumination Engineering, 452-456.
[5] Hunt, R.W.G., Pitt, I.T., and Winter, L.M. (1974) “The Preferred Reproduction of Blue Ski, Green Grass and Caucasian Skin in Colour Photography”, J. Photographic Science, 22: 144-149.
[6] Sanger, D., Asada, T., Haneishi, H., and Miyake, Y. (1994) “Facial Pattern Detection and Its Preferred Color Reproduction”, IS&T/SID 2nd Color Imaging Conference, 149-153.
[7] Yano, T. and Hashimoto, K. (1997) “preference for Japanese Complexion Color under Illumination”, Color Res. Appl. 22(4): 269-274.
[8] Park, D.-S., Kwak, Y., Ok, H., and Kim, C.Y. (2006) “Preferred skin color reproduction on the display”, J. Electronic Imaging, 15(4): 041203.
[9] Kuang, J., Jiang, X., Quan, S., and Chiu, A. (2005) “A psychophysical study on the influence factors of color preference in photographic color reproduction”, Proc. SPIE/IS&T Electronic Imaging: Image Quality and System Performance II, Vol. 5668, 12-19.
[10] Bodrogi, P. and Tarczali, T. (2001) “Colour Memory for Various Sky, Skin, and Plant Colours: Effect of the Image Contest”, Color Res. Appl., 26(4): 278-289.
[11] Zeng, H. and Luo, R, “Skin Color Modeling of Digital Photographic Images”, to be published.
[12] M. Nielsen and M. Stokes, “The Creation of the sRGB ICC Profile”, Proc. IS&T/SID 6th Color Imaging Conf., 253-257 (1998).
[13] Bartleson CJ, Grum F. Optical Radiation Measurements: Visual Measurements, Vol. 5, Academic Press, Inc. 1984.
[14] Zeng, H. and Luo, R, “Colour and Tolerance of Preferred Skin Colours”, to be appeared in Color Res. Appl.