Received: 2 August 2021 | Revised: 7 September 2021 | Accepted: 16 September 2021

DOI: 10.1002/col.22737

RESEARCH ARTICLE

COLOR RESEARCH AND APPLICATION WILEY

Development of an image-based measurement system for human facial skin colour

Ruili He[1] | **Kaida Xiao**[1] | **Michael Pointer**[1] | **Yoav Bressler**[2] | **Zhen Liu**[1,3] | **Yan Lu**[1]

[1]School of Design, University of Leeds, Leeds, UK

[2]Stratasys Ltd., Rehovot, Israel

[3]Engineering College, Qufu Normal University, Rizhao, China

Correspondence
Kaida Xiao, School of Design, University of Leeds, Leeds LS29JT, UK.
Email: k.xiao1@leeds.ac.uk

Funding information
European Union's Horizon 2020, Grant/Award Number: 814158

Abstract

In this study, an image-based measurement system was developed for human facial skin colour, involving the development of a digital imaging system, collection of facial skin colour from 60 human subjects, generation of different colour characterization models, and performance evaluation. The factors that affect facial skin colour characterization, including different training datasets (two colour charts and the collected facial skin colour dataset), mathematical mapping methods (linear transformation, polynomial regression, root-polynomial regression and neural network) and camera image formats (JPG and RAW), were investigated and quantified not only by the conventional method of CIELAB colour difference, but also two newly introduced measures, facial colour contrast and skin colour gamut. The results indicate that the RAW image format for camera digital signals gave a more stable performance than the JPG format images, and the higher order polynomial regression with good predictive accuracy in terms of CIELAB colour difference did not perform well for the whole facial image. It is suggested to evaluate the model performance using the colour of both specific facial positions and the overall facial skin colour. Our comparative analysis in this study provides useful guidance for determining the colour characterization model for facial skin.

KEYWORDS

colour characterization model, facial image, predictive accuracy, skin colour

1 | INTRODUCTION

The colour of skin is one of the colours that we see most often in our daily lives, and it plays an important role in many multidisciplinary applications. Apart from the colour reproduction of skin in amateur and professional photography,[1] cinematography and printing,[2] these applications include the photographic recording of skin colour for medical diagnosis or treatment response,[3-5] skin-colour-based face detection for computer vision applications,[6,7] the identification of skin colour preference in the cosmetic and personal healthcare industries,[8-10] and more recently, the potential manufacture of facial prostheses in 3D colour printing.[11-13] For all these applications, a reliable technique to quantify the colour of skin objectively is of vital importance.

CIE colorimetry was developed as an objective colour quantification tool that represents human colour vision,[14] and instruments such as spectrophotometers and tele-spectroradiometers have been successfully used to measure skin colour in terms of CIE XYZ tristimulus values or CIELAB values. The variability of skin colour measurements for these two kinds of instruments was investigated by Wang et al. to quantify their dependency on the acquisition parameters.[15] Xiao et al. used spectrophotometric measurement to quantify skin colours along three colour dimensions: lightness, redness, yellowness, and CIELAB colour space was used to evaluate skin colour differences and variation of ethnicity and body locations.[16] In addition, Yun et al. conducted skin colour measurement using a portable spectrophotometer to investigate the correlations among the melanin and erythema indexes and CIELAB colour coordinates.[17]

The limitation of such instrument measurements, however, is that the measurement needs to be conducted consecutively and only a small area of skin can be measured each time due to the limited instrumental aperture. Moreover, facial skin colour is nonuniform and texturally uneven, a limited number of spot measurements cannot truly represent the colour appearance of the human face. Consequently, image-based colour measurement as widely used in the graphic-arts industry, has been applied to skin colour measurement. For instance, Miyamoto et al. developed a digital imaging system to quantify hyperpigmented spots on the face and used six colour chips from the GretagMacbeth ColorChecker to validate colour accuracy and reproducibility.[18] Kikuchi et al. developed an image evaluation method for skin colour distribution in facial images, with CIE XYZ tristimulus values calculated from captured images by a matrix.[19]

For the method of image-based colour measurement, the accuracy of colour prediction highly depends on camera colour characterization, which aims to determine a transformation model for converting camera image data to CIE colour coordinates. Although the methods for camera colour characterization have been extensively studied,[20-24] there is only limited comparative analysis available on skin colour prediction. Considering the factors that affect predictive accuracy, Xiao et al. used a skin-specific colour chart, with spectra closer to human skin spectra, to transform camera RGB values to spectral reflectance for skin images.[25] And in our preliminary study, a novel camera colour characterization model for the colour measurement of human skin was developed with higher predictive accuracy.[26]

Another concern about skin colour characterization is the lack of validation methods for quantifying predictive accuracy. In some studies,[27,28] the skin colour information in terms of CIE colour coordinates were derived from the image without reporting colour predictive accuracy of the digital imaging system. Typically, the conventional method used is the CIELAB colour difference, based on testing data of uniform colour patches rather than actual skin colour. It is usually taken for granted in most cases that good prediction of the testing colours will lead to good performance on a whole target, but this does not fairly represent what happens in practice. Additional measures related to the prediction of facial skin colour should be considered for model evaluation.

In this study, an image-based measurement system was developed for human facial skin colour, and the factors that affect colour prediction, such as different training datasets, mapping methods and image formats, were comprehensively investigated. Different colour characterization models were generated based on the developed digital imaging system, and two new measures were proposed to validate the model predictive accuracy in addition to CIELAB colour difference. Moreover, recommendations for developing an image-based measurement system and determining a suitable colour characterization model are provided for different objectives or applications of facial skin colour.

2 | MATERIALS AND METHODS

2.1 | Digital imaging system for facial image capture

To collect colour images of human faces, a specific digital imaging system was developed, and the schematic diagram is shown in Figure 1. This system consists of a high-resolution digital SLR camera and two THOUSLITE LED Cubes, which contain a computer-controlled array of LEDs such that a spectral power distribution with specific requirements can be generated. The two Cubes were aligned to be 30° to the normal to provide uniform illumination on the position of the human face, and the digital camera was located in the horizontal center of the two Cubes, Figure 1A. With the aim of quantitative analysis of skin colour in the captured images, it is essential to avoid specular highlights from the two LED cubes and to ensure the subject's face is illuminated evenly. Therefore, a cross linear polarizer was placed in front of each LED Cube to filter the emitted light, and a second linear polarizing filter was placed over the camera lens at a specified rotation angle to reduce specular highlights in the captured images. In addition, the camera was equipped with a matt black lens hood to prevent any stray light from the two LED Cubes from directly entering the camera lens during image capture.

In the digital imaging system, the illumination produced by the two LED Cubes was set to have a Correlated Colour Temperature (CCT) of 6500 K and the actual

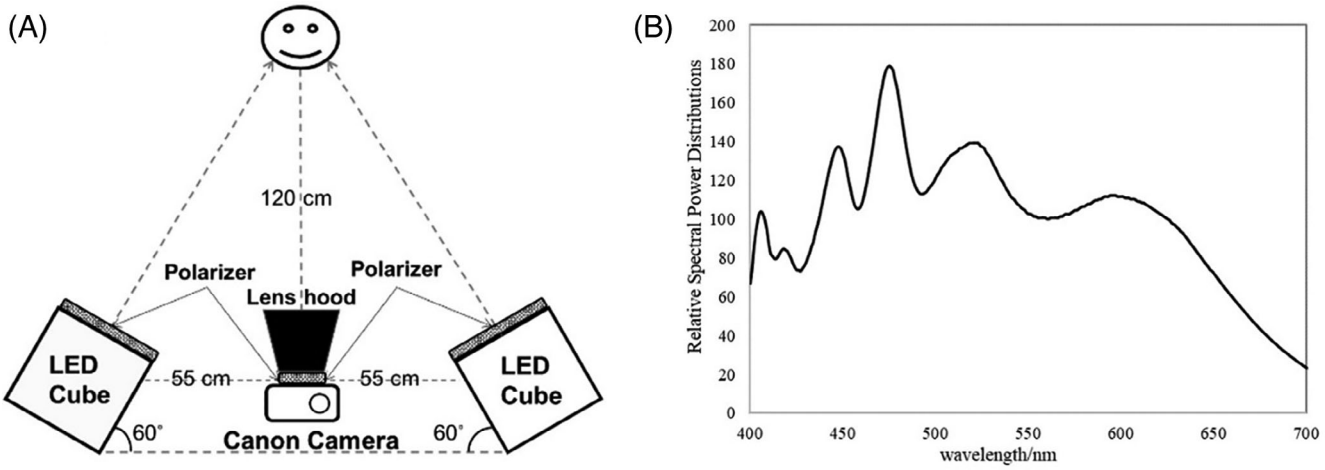

FIGURE 1 The digital imaging system for facial image capture: A, the schematic diagram and B, the relative SPD of the illumination[26]

parameters of CCT, colour rendering index (Ra), Duv and luminance, measured using a CS2000 tele-spectroradiometer, were 6544 K, 95, 0.0064, and 127 cd/m^2, respectively. Figure 1B shows the relative spectral power distribution (SPD) of the illumination used.

Image capture was carried out in a dark room and the LED lighting system produced the only illumination. A Canon EOS 6D Mark II RGB digital camera with a resolution of 6240 (width) × 4160 (height) pixels (equal to a 26 MP image with a 3:2 aspect ratio) was used to capture images of human faces against a black background under the 6500 K illumination, and the camera parameters were set at ISO 640, aperture size F/5.6, shutter speed 1/8 second, the white balance mode with a customized colour temperature of 6500 K. Human participants were asked to present themselves with their faces clean, without any makeup, and without bangs of hair on their foreheads. Each human subject was instructed to put their chin on a fixed chin rest which had a horizontal distance of 120 cm from the camera lens and the image of the face was then captured by the digital camera.

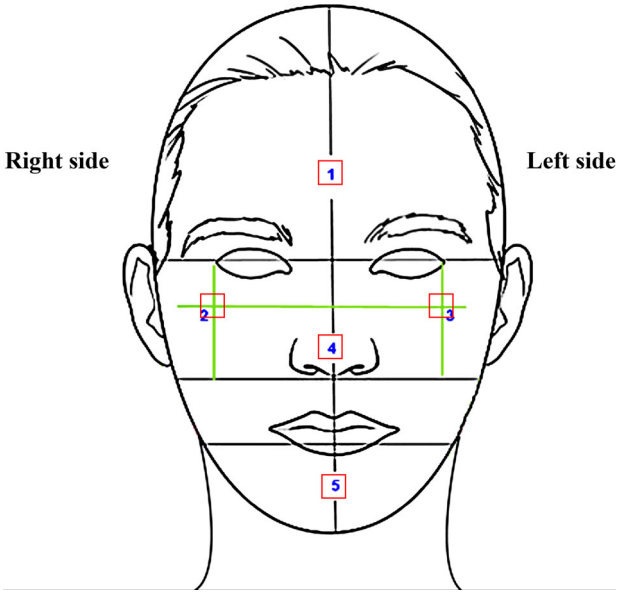

FIGURE 2 The five facial positions measured on each human face: forehead (FH), right cheekbone (CBR), left cheekbone (CBL), nose tip (NT) and chin (CH)[29]

2.2 | Skin colour measurement

To evaluate the performance of facial colour prediction, the skin colour of each participant was measured using a spectrophotometer and considered as the ground truth data. A Konica Minolta CM-2600d spectrophotometer with small aperture size (3 mm) and specular component included (SCI) was used to measure the colour of each of five facial locations, as shown in Figure 2. The forehead (FH), right cheekbone (CBR), left cheekbone (CBL), nose tip (NT) and chin (CH) respectively represents the top, right, left, middle and bottom area of the human face. The spectral reflectance data obtained ranged from 400 to 700 nm with an interval of 10 nm, and it was used to calculate the corresponding CIE XYZ tristimulus values and CIELAB values, with the CIE 1931 standard colorimetric observer colour matching functions (CMFs) and the measured SPD of the 6500 K illumination used in the digital imaging system.

2.3 | Determination of prediction models

A camera colour characterization model is generally determined by a training dataset and a mapping method to convert camera digital signals (image data) to CIE XYZ tristimulus values. In this study, two different image formats, RAW and JPG, were respectively used as camera

digital signals and combined with three different training datasets and seven mathematical methods to generate different models for skin colour prediction.

2.3.1 | Image formats

In the digital imaging system, two image formats, RAW images (.CR2) without postprocessing, and postprocessed JPG images (.JPG), were saved from the digital camera. For the captured JPG images, the RGB values (ranging from 0 to 255) of each pixel were extracted using the *imread* function in MATLAB software. For the RAW images, the open source raw-files converter *dcraw* was used to decode the RAW image without colour correction,[30] and then the RAW RGB values (ranging from 0 to 1) of each pixel were extracted from the *tiff* image in MATLAB.

2.3.2 | Training datasets

Two colour charts used in previous studies were selected separately as a training dataset for skin colour prediction. These were the X-Rite Digital ColorChecker SG chart with 140 colour patches,[12] and the silicon skin colour chart with 95 colour samples that covered a large range of human skin tones.[26] For both colour charts, the spectral reflectance of each sample was measured using the CM-2600d spectrophotometer and the corresponding CIE XYZ tristimulus values, as well as CIELAB values, were calculated with the CIE1931 CMFs and the measured SPD of the 6500 K illumination. Moreover, each colour chart was captured at the same position as the human face using the digital imaging system with the same settings, and the averaged RAW/RGB values of each sample were derived from the captured images.

In addition, a subset of the facial skin colour data collected in this study was also used as a training dataset, which includes the measurement data and image data from 200 facial locations of 40 human subjects. The measurement data refers to the CIE XYZ tristimulus values of each facial position measured using the spectrophotometer, and the image data refers to the averaged RAW/RGB values of each facial position, with a size of 50×50 pixels (about 9 mm^2 on the human face) located in the captured facial images. Therefore, three different datasets were used for the development of skin colour characterization models:

- CCSG dataset: 140 colour patches of the X-Rite Digital ColorChecker SG chart,
- SSCC dataset: 95 colour samples of the silicon skin colour chart,
- FSCD dataset: 200 skin colour data collected of 5 facial locations of 40 human participants.

2.3.3 | Mapping methods

The techniques of linear transformation, polynomial regression (PR), root-polynomial regression (RPR) and neural networks (NN) were used as the mapping methods to transform camera image data to CIE XYZ tristimulus values. For the PR and RPR methods, three degrees (first, second, third) were used with different extensions and the corresponding terms are listed in Table 1. It is noted that the first order for PR and RPR have the same terms and is named as first PR. For the neural network technique, the *fitnet* function in MATLAB was used with the *trainbr* method (Bayesian regularization backpropagation) to train an optimized mapping between input (RAW/RGB values) and output vectors (CIE XYZ tristimulus values). The architecture used has three hidden layers, where the first hidden layer size is 5, the second is 25, the third is 5, and the number of epochs is 1000.

The seven mathematical methods, linear, first PR, second PR, third PR, second RPR, third RPR and NN, were investigated in this study, and combined respectively with the three training datasets (FSCD, CCSG, SSCC) to generate 21 different prediction models for estimating skin colour, in terms of CIE coordinates, respectively based on the captured RAW and JPG facial images.

2.4 | Measures for model evaluation

2.4.1 | Colour difference

The conventional method to quantify the predictive accuracy of colour characterization is to use the average CIELAB colour-difference between instrument measurements and the corresponding predictions of testing colours, as expressed in Equation (1), where n is the number of testing colours (here $n = 100$), i indicates the i^{th} testing data,

TABLE 1 The corresponding terms for linear, first PR, second PR, third PR, second RPR, and third RPR techniques[26]

	Number	terms
Linear	3	r, g, b
First PR	4	$r, g, b, 1$
Second PR	10	$r, g, b, r^2, g^2, b^2, rg, rb, gb, 1$
Third PR	20	$r, g, b, r^2, g^2, b^2, rg, rb, gb, r^2g, r^2b, g^2b, rg^2, rb^2, gb^2, r^3, g^3, b^3, 1$
Second RPR	7	$r, g, b, \sqrt{rg}, \sqrt{rb}, \sqrt{gb}, 1$
Third RPR	14	$r, g, b, \sqrt{rg}, \sqrt{rb}, \sqrt{gb}, \sqrt[3]{r^2g}, \sqrt[3]{r^2b}, \sqrt[3]{b^2g}, \sqrt[3]{rg^2}, \sqrt[3]{rb^2}, \sqrt[3]{gb^2}, \sqrt[3]{rgb}, 1$

L_0^*, a_0^*, b_0^* indicate the CIELAB values calculated from instrument measurements on the real human face, and L_1^*, a_1^*, b_1^* represent the CIELAB values calculated from the predicted CIE XYZ tristimulus values by each model. Larger CIELAB colour-difference values indicate lower accuracy of the corresponding prediction model.

$$\Delta E_{ab}^* = \frac{1}{n}\sum_{i=1}^{n}\sqrt[2]{\left(L_{0i}^* - L_{1i}^*\right)^2 + \left(a_{0i}^* - a_{1i}^*\right)^2 + \left(b_{0i}^* - b_{1i}^*\right)^2} \quad (1)$$

2.4.2 | Facial colour contrast

Human skin colour is nonuniform and unevenly distributed. Colour heterogeneity is one of the important characteristics of facial appearance, and it needs to be well preserved in the facial image-based measurement system. In this study, the facial colour contrast between the five facial locations (Figure 2) is proposed to present the overall colour heterogeneity of the face. For each subject, the differences between each two facial locations were calculated (a total of 10 differences listed in Table 2) in terms of three colour components, CIELAB lightness L^*, redness $+a^*$ and yellowness $+b^*$.

In order to evaluate the degree of preservation of facial colour contrast in prediction models, the Pearson Correlation Coefficient (PCC) was used to quantify the relationship between the facial colour contrast of the objective measurements of an actual human face and those estimated from the corresponding facial image by the prediction models. The value of the PCC ranges from -1 to $+1$ where a value of $+1$ denotes total positive linear correlation, 0 denotes no linear correlation, and -1 denotes total negative linear correlation. The larger the PCC value, the better the preservation of facial contrast of the prediction model.

TABLE 2 Ten differences between each two different facial locations

	FH	CBR	CBL	NT	CH
FH					
CBR	Δ				
CBL	Δ	Δ			
NT	Δ	Δ	Δ		
CH	Δ	Δ	Δ	Δ	

2.4.3 | Skin colour gamut

Since previous methods to validate the predictive accuracy were only based on the specific facial locations, it is crucial to ensure the colour prediction of the whole facial image, especially for the image-based measurements in cosmetics and medical applications. Therefore, a skin colour gamut in terms of CIELAB values was introduced to verify the colour distributions estimated by each model. The skin colour gamut was approximately defined with the lightness L^* ranging from 40 to 75, the a^* value from 0 to 30 and the b^* value from 5 to 35, which was based on the range of the facial skin colours collected in this study (plotted in Figure 3) and the skin colour database developed for four ethnic groups (Caucasian, Chinese, Kurdish, Thai) by Xiao et al.[16]

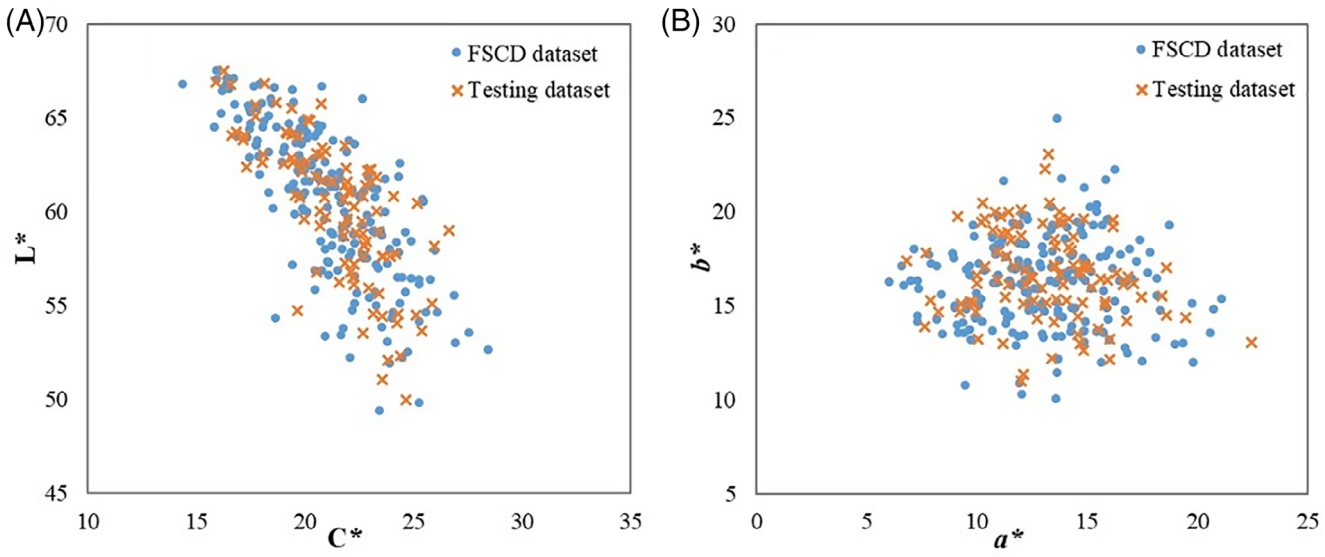

FIGURE 3 Colour distributions of the spectrophotometer measurements in the CIELAB L^*C^* (A) and a^*b^* (B) plane

It is desirable that the prediction model can convert most of the skin-colour pixels falling within the predefined gamut with the least colours outside the boundary. The percentage of pixel-based colours outside the gamut was calculated for each facial image, and the average result of all the testing images was used to validate the model performance, expressed in Equation (2), where m indicates the number of testing facial images (here $m = 20$), N is the total number of skin-colour pixels in the facial image, and n is the number of predicted colours that fall outside of the predefined colour gamut. A larger percentage value indicates that more colours are predicted outside the gamut and the prediction model has poorer predictive accuracy.

$$P = \frac{1}{m}\sum_{i=1}^{m}(n/N) \tag{2}$$

3 | RESULTS

3.1 | Facial images and skin colour data

In this study, 60 human subjects, including 15 Caucasian, 15 Chinese, 15 Indonesian and 15 Mexican, participated in the collection of facial images and skin colour measurement. These subjects included 22 males and 38 females with ages ranging from 19 to 29 years. Before their participation, all were given an introduction to this research and asked to sign a consent form, in accordance with the ethical review procedure of the University of Leeds.

An image of a white board, placed at the position of the human face, was used to assess the uniformity of the lighting. It was found that the lightness, L^*, values, from the image, ranged from 76.5 to 78.5, indicating that the lighting had good uniformity on the human face. In total, 120 colour images (60 JPG images and 60 RAW images) of human faces were captured from the 60 human subjects. In addition, 300 skin colours, in terms of CIE XYZ tristimulus values as well as CIELAB colour attributes, were collected using spectrophotometer measurement at five facial locations of each human face.[26] Figure 3 shows the distributions of the collected skin colour data in the CIELAB lightness-chroma L^*C^* plane (left) and the chromatic a^*b^* plane (right), where the dot symbols denote the 200 skin colours that were used as a training dataset, and the cross symbols denote the remaining 100 skin colours that were used as the testing data to evaluate model performance. These colour data generate an approximate colour gamut of human facial skin obtained in this study, with the lightness, L^*, ranging from 49.41 to 67.56, and the redness a^* and yellowness b^* varying from 6.03 to 22.43 and 10.04 to 24.95, respectively.

3.2 | Colour characterization models

Based on each training dataset with known CIE XYZ tristimulus values, each of the mapping methods were used to convert image data to the corresponding CIE XYZ values, and the prediction model was determined with optimal transformation results. In total, 42 colour characterization models (2 * 3 * 7) were investigated for the two image formats (RAW and JPG), three training datasets (FSCD, SSCC, CCSG) and seven mathematical mapping methods (linear, first PR, second PR, third PR, second RPR, third RPR, NN). Each model was applied to the facial image and the skin colours of the human face were estimated from RAW/RGB values.

3.3 | Model evaluation

The 100 facial skin colours collected from 20 human subjects were used as the testing dataset, and the performance of each prediction model was evaluated from the colour difference, facial colour contrast and skin colour gamut.

3.3.1 | Colour difference

The CIELAB colour-differences between the face measurements and each of the model predictions were calculated, and results are illustrated in Figure 4, where (A) shows the results obtained when using JPG image data as camera digital values, and (B) represents the results based on RAW image data. For each figure, the average colour-difference values are plotted and compared for each training dataset and mapping method. Error bars are included which were calculated using the SE of the mean for each model.

It can be clearly seen in Figure 4A that the CCSG training dataset (blue bars) has considerably larger colour-difference values than the other two datasets, except when a neural network was used as the mapping method. Moreover, the colour differences resulted from three different training datasets and seven mapping methods vary greatly. However, when the RAW images were used for skin colour prediction, Figure 4B, the variations become smaller, and the CCSG chart has similar CIELAB colour-difference values with other training datasets. In addition, when the FSCD training dataset was used, in most cases it gave the smallest colour

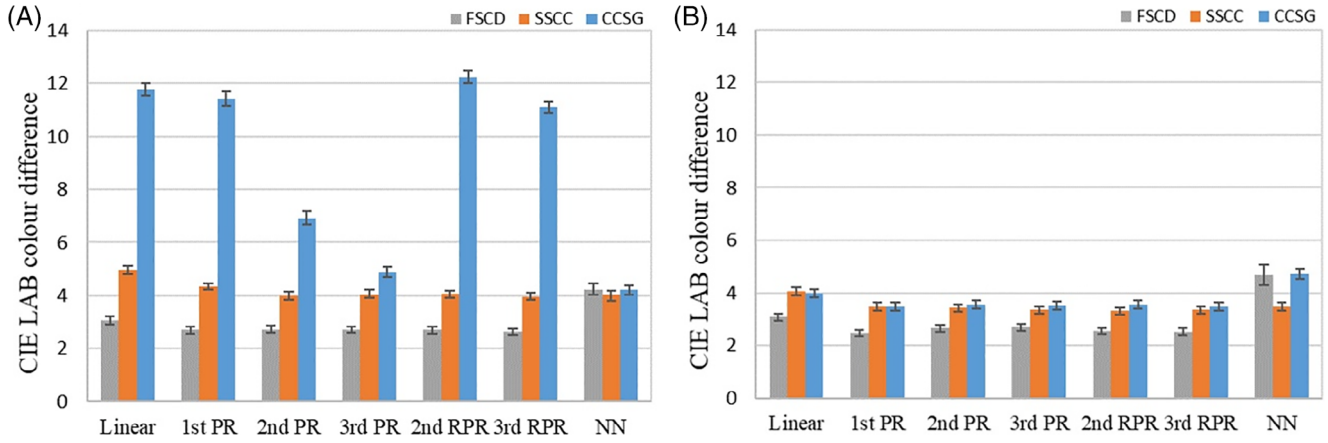

FIGURE 4 Histogram of the average CIELAB colour difference for each prediction model based on (A) JPG images and (B) RAW images

difference for both image formats except for the neural network model (NN).

Tables 3 and 4 respectively list the average and maximum CIELAB colour-difference values of each colour characterization model for the two image formats. With JPG images as camera digital signals, the third RPR mapping method predicted the smallest colour differences for the FSCD and SSCC datasets, 2.63 ΔE^*_{ab} and 3.96 ΔE^*_{ab}, respectively, followed by the first PR (2.69 ΔE^*_{ab}) and second PR method (3.97 ΔE^*_{ab}). For the CCSG training dataset, the smallest colour difference was achieved by the NN technique with the value of 4.21 ΔE^*_{ab}, followed by the third PR method (4.87 ΔE^*_{ab}).

For the results based on RAW image data, Table 3, the first PR, second RPR and third RPR methods predicted the smallest colour-difference value for the FSCD, SSCC and CCSG datasets, respectively. The colour differences predicted from the CCSG dataset were significantly reduced to less than 4 ΔE^*_{ab} from greater than 10 ΔE^*_{ab}, compared with the results (with the linear, first PR, second RPR, third RPR mapping methods) for JPG images. In addition, it can be observed in Table 4 that the maximum colour-difference value does not always correspond to the average colour-difference values, which suggests that more aspects need to be considered to make good predictions of facial skin colour based on images.

3.3.2 | Facial colour contrast

The results of the facial colour contrast PCC in terms of the three colour attributes, lightness, redness and yellowness are shown in Figure 5, where (A) and (B) show the results from JPG and RAW images, respectively. Regarding facial contrast for lightness, the FSCD dataset gives the best predictions, with PCC values greater than those for the SSCC and CCSG datasets. For facial contrast for redness and yellowness, different prediction models gave similar results, except for the NN model.

To compare the impact of different training datasets and mapping methods on the prediction of facial colour contrast, the PCC values of the three attributes were averaged and are given in Table 5. From the results from JPG images, the first PR method gives the maximum value of PCC, 0.64, for the FSCD training dataset. For the SSCC training dataset, the first PR, second RPR and NN methods have similar PCC values of 0.52. The model based on a neural network performed well for the CCSG dataset with the largest value of 0.55. In comparison, the NN method has the worst performance (0.44) for the FSCD dataset, the linear transformation (0.44) for the SSCC dataset and the second RPR method (0.40) for the CCSG dataset.

From the results based on RAW images, the first PR method has the largest PCC value, 0.62, for the FSCD dataset. For the SSCC and CCSG datasets the best performance was achieved with the second PR method and the linear transformation, respectively. The neural network model gave the worst predictive results when the FSCD dataset was used as the training data, with a PCC value of only 0.21.

3.3.3 | Skin colour gamut

The percentage of predicted colours outside the predefined skin colour gamut predicted by each model was calculated for each facial image, and the averaged results for 20 testing images are listed in Table 6. It is shown that different colour characterization models determined by three training datasets and seven mapping methods lead to obvious differences in estimating

TABLE 3 Average CIELAB colour-difference values based on JPG and RAW images (The smallest value for each dataset is shown in bold)

Images	Training	Linear	First PR	Second PR	Third PR	Second RPR	Third RPR	NN
JPG	FSCD	3.06	2.69	2.72	2.71	2.69	**2.63**	4.23
	SSCC	4.97	4.34	3.97	4.04	4.06	**3.96**	4.01
	CCSG	11.76	11.43	6.93	4.87	12.25	11.10	**4.21**
RAW	FSCD	3.08	**2.47**	2.65	2.70	2.55	2.53	4.68
	SSCC	4.06	3.48	3.43	3.35	**3.32**	3.35	3.48
	CCSG	3.98	3.48	3.56	3.51	3.54	**3.47**	4.73

TABLE 4 Maximum CIELAB colour-difference values based on JPG and RAW images (The smallest maximum value for each dataset is shown in bold)

Images	Training	Linear	First PR	Second PR	Third PR	Second RPR	Third RPR	NN
JPG	FSCD	9.53	6.51	6.26	6.05	8.10	**5.96**	16.62
	SSCC	8.35	7.52	8.68	8.78	7.80	**7.36**	8.55
	CCSG	17.00	17.16	14.63	**9.17**	17.73	15.97	14.69
RAW	FSCD	7.78	**5.75**	8.21	7.75	6.46	6.32	18.63
	SSCC	10.68	6.84	6.94	**6.75**	7.00	6.94	7.06
	CCSG	8.50	**7.69**	8.07	7.74	8.19	7.76	8.79

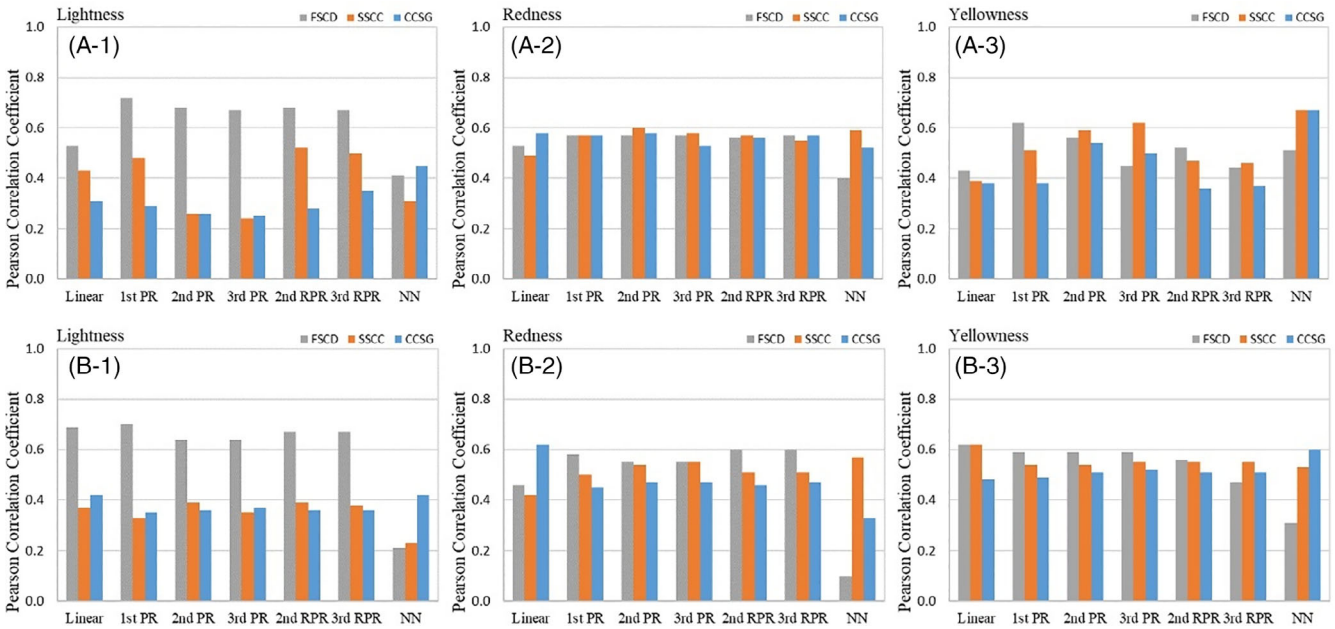

FIGURE 5 Histogram of facial colour contrast PCC in lightness, redness and yellowness based on (A) JPG images and (B) RAW images

pixel-based skin colours from JPG images. For instance, the third RPR method predicted 0.11% colours outside the skin colour gamut for the FSCD training dataset, but it resulted in 12.30% colour prediction when using the CCSG training dataset. In contrast, the linear transformation predicted the least number of colours outside the gamut for the CCSG dataset, and the same for the SSCC dataset. The minimum percentage value for the FSCD dataset was obtained using the first PR method (0.07%).

TABLE 5 Averaged PCC values for each prediction model based on JPG and RAW images

Images	Training	Linear	First PR	Second PR	Third PR	Second RPR	Third RPR	NN
JPG	FSCD	0.50	0.64	0.60	0.56	0.59	0.56	0.44
	SSCC	0.44	0.52	0.48	0.48	0.52	0.50	0.52
	CCSG	0.42	0.41	0.46	0.43	0.40	0.43	0.55
RAW	FSCD	0.59	0.62	0.59	0.60	0.61	0.58	0.21
	SSCC	0.47	0.46	0.49	0.48	0.48	0.48	0.44
	CCSG	0.51	0.43	0.45	0.45	0.44	0.44	0.45

TABLE 6 Percentages (%) of abnormal colours based on JPG and RAW image data (the smallest value for each dataset is shown in bold)

Images	Training	Linear	First PR	Second PR	Third PR	Second RPR	Third RPR	NN
JPG	FSCD	2.86	**0.07**	0.15	1.40	0.12	0.11	1.05
	SSCC	**1.42**	1.43	3.70	2.27	2.17	1.90	2.74
	CCSG	**0.06**	0.57	6.39	0.76	2.67	12.3	0.47
RAW	FSCD	0.00	**0.00**	0.60	0.84	0.43	0.09	0.95
	SSCC	0.80	1.16	0.86	0.72	0.69	**0.69**	1.03
	CCSG	**0.24**	1.45	0.94	1.38	1.27	1.65	0.87

In comparison, these models predicted fewer colours outside the predefined skin colour gamut based on RAW images, which indicates most skin colours were estimated by the model. Compared with the SSCC and CCSG datasets, the FSCD dataset almost has the smallest percentage of predicted colours that exceed the skin gamut, with the percentages in the range from 0.00% to 0.95%. The value of 0.00% means that the colour of each pixel in the facial image was predicted within the predefined skin colour gamut. In addition, the maximum percentage is 1.65% for RAW images, which was achieved by the prediction model determined with the CCSG dataset and the third RPR mapping method. This model also resulted in the maximum value (12.3%) for JPG images.

4 | DISCUSSION

To determine a colour characterization model in an image-based measurement system, it is necessary to select an appropriate training dataset, mapping method and image format as well. In this study, two image formats (RAW and JPG), three training datasets (FSCD, SSCC, CCSG) and seven mapping methods (linear, first PR, second PR, third PR, second RPR, third RPR, NN) have been investigated for the development of skin colour characterization model. Moreover, the predictive accuracy of the model was evaluated from three aspects.

4.1 | Determination of colour characterization model

Regarding the two image formats, JPG images result in large variation in the predictive accuracy of different models in terms of CIELAB colour differences and the percentages of colours outside the proposed gamut. This is probably because JPG images have a complex nonlinear power law relationship between digital signal response and light intensity, it is slightly compromised to find a best mapping between input and output vectors. In contrast, different models have similar predicted results for RAW images because of the linear relationship with light intensity, which is relatively simple to determine a mapping model. Moreover, it was confirmed that the predictive accuracy based on RAW images were basically better than that based on JPG images, indicating that RAW image data is recommended as camera digital signals for facial skin colour prediction. This agrees with the results of Zhang's study that also suggested that RAW images would be better than JPG images for predicting spectral reflectance.[23]

Three different training datasets have a significant impact on skin colour prediction, especially for JPG images. Although the CCSG colour chart has been widely used for camera colour characterization in various industrial applications,[31] it has the worst predictive accuracy in this study. The possible reason for this is that there is a very limited number (14) of skin-colour patches in the CCSG chart and this is an insufficient number for

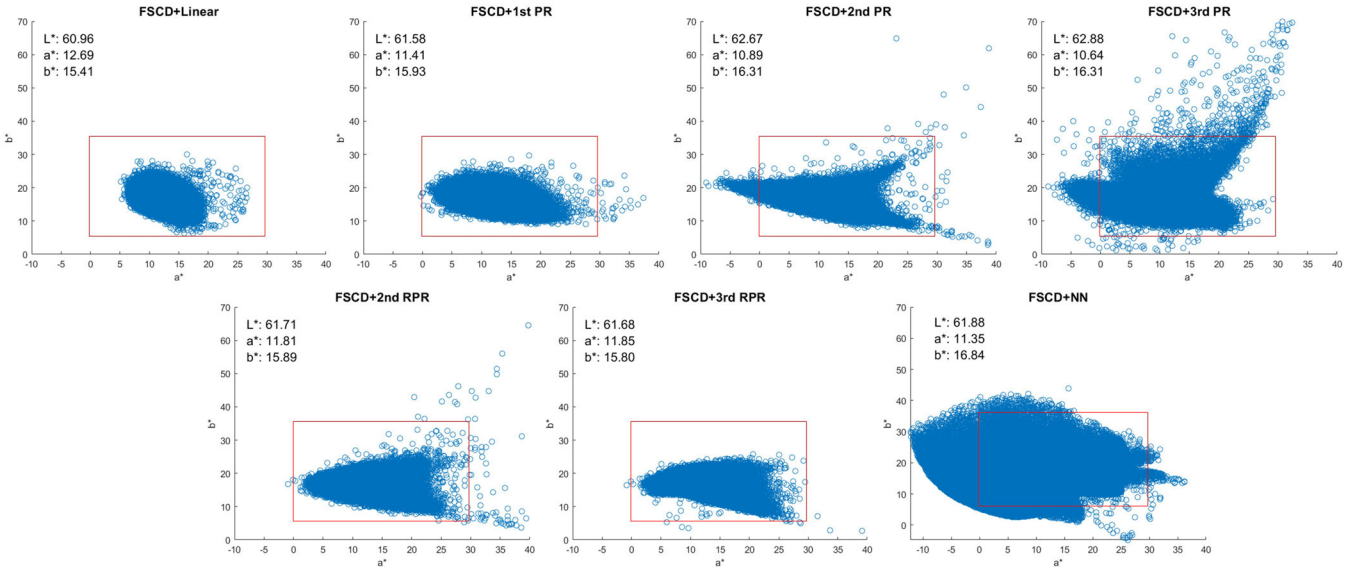

FIGURE 6 An example of colour distributions and averaged CIELAB values predicted using the seven mapping methods and the FSCD training dataset (the red squares indicate the predefined skin colour gamut in the a^*b^* plane)

training skin colours. Consequently, the CCSG chart is not the most suitable training dataset to predict skin colour from JPG images. In comparison, the SSCC training dataset provides a slightly better performance in terms of predictive accuracy than the CCSG chart, since it covers a wide range of skin tones. The FSCD training dataset has the best performance in predicting skin colours, revealing that real skin colour data is to be preferred as the training dataset to achieve more accurate colour prediction, specifically for whole facial image manipulation.

In terms of mapping methods, previous studies have shown that a higher order polynomial model performs better than a lower order model for higher predictive accuracy of camera colour characterization.[22] In this article, a similar result was found only when CCSG training dataset were used for JPG images. This is not surprising because most studies were based on the ColorChecker chart for camera colour characterization, and few studies focus on the investigation of different training datasets for skin colour prediction. While the linear transformation predicted most colours within the skin colour gamut, it has larger predicted errors in terms of CIELAB colour difference. In comparison, the first PR method has good performance on predictive accuracy from all three measures, it takes good care of facial points and areas.

In addition, the difference between the results of second, third RPR and that of second, third PR methods is not obvious, since the root-polynomial regression is a simple (low complexity) extension of polynomial regression,[32] but it has better results of predicted skin colours in terms of the out-of-bounds percentage. As shown in Figure 6, an example of skin colour distributions in a^*b^* plane and the average CIELAB values predicted by the seven different mathematical methods with the FSCD training dataset, where the red squares indicate the predefined skin colour gamut in the a^*b^* plane. Although all methods gave similar averaged CIELAB values, there is large variation in the colour scatter predicted by the third PR and NN mapping methods, which caused many predicted colours to be outside the skin colour gamut. It is suggested that the mapping method with complex extensions is not suitable for predicting pixel-based skin colours, since the complex mapping method may have an overfitting problem for the colours near to the gamut boundary. The neural network gave stable performance in terms of CIELAB colour difference for different training datasets and two image formats, but the results of predicting pixel-based colours are not satisfactory. This is probably because the number of samples in the training dataset is limited and much smaller than number in the testing data.

4.2 | Evaluation measures

The colour difference between the measured and predicted data is a simple and conventional approach to quantify the predictive accuracy, and most studies rely on this metric to develop a colour characterization model. However, it was found in this study that colour difference is not enough for the comprehensive evaluation of colour characterization. For instance, the colour difference predicted by the first PR and third PR mapping methods from JPG images was 2.69 and 2.71 ΔE^*_{ab}, respectively,

which are similar in value, but there were separately 0.07% and 1.40% predicted colours outside the skin colour gamut. In such a case, the first PR method is recommended to convert image data to CIE colour coordinates. Moreover, the colour difference is dependent on the absolute difference of testing points between the ground truth (instrument measurements) and the predictions, ignoring the characteristics of facial skin such as its colour heterogeneity. In this study, two additional measures are proposed: the facial colour contrast is concerned with the relative relationship between five different facial positions, and the skin colour gamut is predetermined to validate all pixel-based skin colours.

To achieve accurate image-based measurement for skin colour, it is expected that the facial colour contrast predicted from the facial image is well preserved with that of the actual human face. Regarding the results of facial colour contrast, different models have a similar PCC value, less than 0.65. This is probably due to: (1) the method for quantifying facial colour contrast, (2) the transformation model for predicting colours. It requires further investigation on facial colour contrast, which is an important factor for faithful skin colour reproduction of the human face.

The introduction of the proposed skin colour gamut allows for validation of all pixel-based skin colours predicted from the whole facial skin area, instead of focusing on just five facial locations. As shown in Figure 6, there is an evident difference between the skin colour distributions in the a^*b^* plane estimated using different prediction models, although the averaged CIELAB values were similar (1.5 ΔE^*_{ab}). This indicates that the average values are not adequate to quantify the predictive accuracy objectively of human facial skin colour. Although it was confirmed that the skin colour gamut is an effective measure to evaluate the prediction accuracy of the overall detail colours, it is simply defined from three dimensions of L^*, a^*, b^*, which generates a cubic gamut in orthogonal CIELAB colour space. It is recognized that an ellipsoid may better represent skin colour distributions and thus the specification of the boundary of the skin colour gamut should be further investigated.

4.3 | Recommendation

To develop an image-based measurement system for facial skin colour, three components are indispensable: the digital imaging system for human faces, the colour characterization model, and the measures for model evaluation. In the digital imaging system, it is necessary to provide diffuse illumination on the human face to estimate accurately skin colour from the captured image. Moreover, a linear polarizer filter can be applied to the light source and camera lens to reduce the specular light in the captured facial images.

Regarding the determination of the colour prediction model, careful selection the training dataset and the mapping method is necessary. In general, the skin-colour related colour chart is recommended instead of ColorChecker chart as the training dataset for higher predictive accuracy in most cases, and the skin colour data collected from human faces is preferred to perform image-based measurement for facial skin colour. Moreover, the RAW image format has more stable performance as camera digital signals than the JPG images. As for the mapping method, the first polynomial regression is suggested for predicting colours in both specific facial positions and wide facial skin area. When mainly focusing on the colour at specific facial positions, the third PR and third RPR methods also have good performance.

In addition, there is no doubt that colour difference is a simple way to quantify predictive accuracy, but it only focuses on the absolute accuracy of individual facial locations. More aspects should be considered to validate the predictive performance of the model. The two newly introduced measures, facial colour contrast representing the relative relationship among five different facial locations and skin colour gamut for validating all pixel-based skin colours, play a role in the validation of a prediction model.

5 | CONCLUSIONS

In conclusion, this study developed an image-based measurement system for human facial skin colour and investigated the effects of two camera image formats, three training datasets and seven mapping methods on skin colour characterization. The comparative analysis on the performance of 42 prediction models revealed interesting differences between different image formats, training datasets and mapping methods, and provided support for the determination of the prediction model for facial skin colour.

ACKNOWLEDGMENTS

We would like to thank and acknowledge the ApPEARS Appearance Printing European Advanced Printing School and the funding support from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement no. 814158.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Zenodo at https://zenodo.org/record/5532176#.YVHgCprMKM8

ORCID

Kaida Xiao https://orcid.org/0000-0001-7197-7159
Michael Pointer https://orcid.org/0000-0003-0063-1753

REFERENCES

[1] Zeng H, Luo R. Colour and tolerance of preferred skin colours on digital photographic images. *Color Res Appl*. 2013;38(1):30-45.

[2] Pedersen M, Bonnier N, Hardeberg JY, et al. Attributes of image quality for color prints. *J Electron Imag*. 2010;19(1):011016.

[3] Herbin M, Bon FX, Venot A, et al. Assessment of healing kinetics through true color image processing. *IEEE Trans Med Imaging*. 1993;12(1):39-43.

[4] Choi KM, Kim SJ, Baek JH, Kang SJ, Boo YC, Koh JS. Cosmetic efficacy evaluation of an antiacne cream using the 3D image analysis system. *Skin Res Technol*. 2012;18(2):192-198.

[5] Ly BCK, Dyer EB, Feig JL, Chien AL, del Bino S. Research techniques made simple: cutaneous colorimetry: a reliable technique for objective skin color measurement. *J Invest Dermatol*. 2020;140(1):3-12.

[6] Jones MJ, Rehg JM. Statistical color models with application to skin detection. *Int J Comput Vis*. 2002;46(1):81-96.

[7] Brancati N, De Pietro G, Frucci M, et al. Human skin detection through correlation rules between the YCb and YCr subspaces based on dynamic color clustering. *Comput Vis Image Understand*. 2017;155:33-42.

[8] Caisey L, Grangeat F, Lemasson A, Talabot J, Voirin A. Skin color and makeup strategies of women from different ethnic groups. *Int J Cosmet Sci*. 2006;28(6):427-437.

[9] Fink B, Grammer K, Matts PJ. Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evol Hum Behav*. 2006;27(6):33-442.

[10] Lee M, Han J, Kim E. An evaluation of the effects of makeup on perceived age based on skin color in Korean women. *J Cosmet Dermatol*. 2019;18(4):1044-1051.

[11] Xiao K, Zardawi F, van Noort R, Yates JM. Developing a 3D colour image reproduction system for additive manufacturing of facial prostheses. *Int J Adv Manuf Technol*. 2014;70(9–12):2043-2049.

[12] Xiao K, Wuerger S, Mostafa F, et al. New trends in 3D printing: colour image reproduction for 3D printing facial prostheses. *IntechOpen*. 2016;89-109.

[13] Sohaib A, Amano K, Xiao K, Yates JM, Whitford C, Wuerger S. Colour quality of facial prostheses in additive manufacturing. *Int J Adv Manuf Technol*. 2018;96(1):881-894.

[14] CIE. *CIE 15:20184 Colorimetry*. Vienna: CIE; 2018.

[15] Wang M, Xiao K, Luo MR, Pointer M, Cheung V, Wuerger S. An investigation into the variability of skin colour measurements. *Color Res Appl*. 2018;43(4):458-470.

[16] Xiao K, Yates JM, Zardawi F, et al. Characterising the variations in ethnic skin colours: a new calibrated data base for human skin. *Skin Res Technol*. 2017;23(1):21-29.

[17] Yun IS, Lee WJ, Rah DK, Kim YO, Park BY. Skin color analysis using a spectrophotometer in Asians. *Skin Res Technol*. 2010;16(3):311-315.

[18] Miyamoto K, Takiwaki H, Hillebrand GG, Arase S. Development of a digital imaging system for objective measurement of hyperpigmented spots on the face. *Skin Res Technol*. 2002;8(4):227-235.

[19] Kikuchi K, Mizokami Y, Egawa M, Yaguchi H. Development of an image evaluation method for skin color distribution in facial images and its application: aging effects and seasonal changes of facial color distribution. *Color Res Appl*. 2020;45(2):290-302.

[20] Hong G, Luo MR, Rhodes PA. A study of digital camera colorimetric characterization based on polynomial modeling. *Color Res Appl*. 2001;26(1):76-84.

[21] Cheung V, Westland S, Connah D, Ripamonti C. A comparative study of the characterisation of colour cameras by means of neural networks and polynomial transforms. *Colorat Technol*. 2004;120(1):19-25.

[22] Li C, Cui G, Luo MR. The accuracy of polynomial models for characterising digital cameras. *Proceedings of AIC2003 Bangkok: Color Communication and Management*; 2003:166-170.

[23] Zhang X, Wang Q, Li J, Zhou X, Yang Y, Xu H. Estimating spectral reflectance from camera responses based on CIE XYZ tristimulus values under multi-illuminants. *Color Res Appl*. 2017;42(1):68-77.

[24] Liang J, Xiao K, Pointer MR, Wan X, Li C. Spectra estimation from raw camera responses based on adaptive local-weighted linear regression. *Opt Express*. 2019;27(4):5165-5180.

[25] Xiao K, Zhu Y, Li C, Connah D, Yates JM, Wuerger S. Improved method for skin reflectance reconstruction from camera images. *Opt Express*. 2016;24(13):14934-14950.

[26] He R, Xiao K, Pointer M, et al. A novel camera colour characterisation model for the colour measurement of human skin. *Electronic Imaging (EI)*. Vol 222; 2021 (5).

[27] Kikuchi K, Masuda Y, Yamashita T, Kawai E, Hirao T. Image analysis of skin color heterogeneity focusing on skin chromophores and the age-related changes in facial skin. *Skin Res Technol*. 2015;21(2):175-183.

[28] De Rigal J, Des Mazis I, Diridollou S, et al. The effect of age on skin color and color heterogeneity in four ethnic groups. *Skin Res Technol*. 2010;16(2):168-178.

[29] He R, Xiao K, Pointer M, Westland S. Assessing skin tone heterogeneity under various light sources. *London Imaging Meeting (LIM)*. Society for Imaging Science and Technology; 2020:5-9.

[30] Coffin D, *Decoding raw digital photos in Linux*; 2018 https://www.dechifro.org/dcraw.

[31] Fairchild MD, Wyble DR, Johnson GM. Matching image color from different cameras. *International Society for Optics and Photonics: Image Quality and System Performance*. Vol 68080E; 2008.

[32] Finlayson GD, Mackiewicz M, Hurlbert A. Color correction using root-polynomial regression. *IEEE Trans Image Process*. 2015;24(5):1460-1470.

AUTHOR BIOGRAPHIES

Ruili He is a PhD candidate in the School of Design, University of Leeds, UK, and a Marie Sklodowska-Curie ITN Early Stage Researcher in ApPEARS Appearance Printing European Advanced Printing School. She is engaging in skin colour reproduction, colour appearance modeling and 3D colour printing.

Kaida Xiao is an associate professor in the School of Design, University of Leeds, UK and a Visiting Professor in the University of Science and Technology Liaoning, China. His research interests are related to 3D colour image reproduction, 3D colour printing, medical image capture and analysis, colour appearance modeling and image quality enhancement.

Michael Pointer received a Ph.D. from Imperial College, London in 1972, working with David Wright. He then worked in the Research Division of Kodak Limited. After periods at the University of Westminster and the National Physical Laboratory, he is now a Visiting Professor at the University of Leeds and technical advisor to the Color Engineering Laboratory at Zhejiang University, Hangzhou, China. He is UK Associate Editor of the journal, *Color Research and Application*.

Yoav Bressler received an M.Sc. with distinction from the Interdisciplinary Department of the School of Engineering, Tel-Aviv University, Israel in 1986. His interest in colour technology started in 1988 at Scitex Corporation and continued in large companies, including Kodak and HP as well small start-ups including Shira Computers and RealTimeImage. Since 2007, Yoav has taken a special interest in UV cured ink-jet printers, cumulating in 3D Poly-jet full colour printing in Stratasys Ltd. for the last 6 to 7 years.

Zhen Liu, received a Ph.D. in graphic communication engineering from Wuhan University, China, in 2012. Currently she is working in the Institute of Color Vision and Imaging, Qufu Normal University, China. His work focuses on multispectral imaging, colour restoration of faded relics.

Yan Lu received a BEng in electrical engineering and automation (2016) and an MS in electric-optical measurement and control technology (2019) from Fudan University, China. Currently she is a PhD researcher at the University of Leeds. Her work focuses on modeling facial colour appearance and facial attractiveness for human complexions.

How to cite this article: He R, Xiao K, Pointer M, Bressler Y, Liu Z, Lu Y. Development of an image-based measurement system for human facial skin colour. *Color Res Appl*. 2022;47(2): 288-300. doi:10.1002/col.22737