Designing a Simple 3-Channel Camera for Skin Detection

*J. Birgitta Martinkauppi**
University of Oulu, Oulu, Finland

Graham D. Finlayson
University of East Anglia, Norwich, United Kingdom

Abstract

Skin detection is an important preprocessing step for many applications. In some cases, reliable detection is needed under the real-world's challenging illumination conditions, that is, when the prevailing illumination does not correspond to the one used in calibration. Our particular goal in this paper is to design and study a three-sensor camera for these kinds of illumination conditions. This is done in three stages. First, a representative set of illuminants is selected from a given color temperature range. In the second stage, different illumination normalization methods are tested for the camera channel model. Simple bell-shaped sensors are tested with a chosen normalization method in the last stage. Different sensor combinations are evaluated based on their gamut ratios for skin and Munsell reflectances.

Introduction

Color information about skin is essential for several applications. For example, it can be used as a cue for finding skin areas from images in human-computer interaction or face tracking. Unfortunately, color information is sensitive to illumination changes – which are common in many practical situations.

Skin related research has been intense lately. For example, Imai et al.[1] has studied skin reflectances and Martinkauppi and Soriano[2] skin color signals (light modified by the reflectance of skin). Störring et al.[3] has shown in normalized color coordinates with one white balancing that the track of skin "body" chromaticities follows the curve of Planckian illuminants' chromaticities. The area of possible skin chromaticities for the one white balancing condition and the color temperature range was called the skin locus. Later, this skin locus was extended to several white balancing cases.[4] More about facial skin color modeling for real cameras can be found in ref. 5.[5]

While the earlier research of skin detection has concentrated on skin already perceived by a camera, in this study we examine the perception phase of sensors themselves. The goal is to find, theoretically, a set of three sensors which would be optimal for skin detection under challenging illumination conditions. Of course, different sensors have been modeled and studied before, for example by Vora and Trussell.[6] The unique aspect in our approach is that the sensors are optimized for detecting skin under varying illumination. This approach consists of three separate stages.

In the first stage, a representative set of illuminants is selected from a chosen color temperature (CT) range, 2000 K – 10000 K. The illuminants are modeled using a spectral power distribution (SPD) of blackbody radiators[7] (also called Planckian radiators). The problem in illuminant selection is caused by the nonlinear relationship between SPD and CT. Equal CT change at different places of the range produces different amounts of SPD change. To improve the situation, the use of an inverse color temperature range has been suggested and the reciprocal CT values have been shown to correspond better to the way humans see SPD variations.[8] Here, we study the chromaticities of the SPDs for a Sony camera and their relation to the color temperature range and its inverse. Note that the chromaticities are now presented in a camera dependent space instead of a human color space like u'v'-space.[9]

Next, the effect of different illumination normalization methods is studied for the common camera channel model. Some kind of illumination normalization is necessary because the level and power of illuminants varies greatly. The most common method has been to scale the illuminants so that they have equal values at the wavelength of 560 nm. Another method is to scale using Euclidean rule to obtain one coefficient for the whole SPD, like Romero et al.[10] The third possible method is to use the white response as a scaling factor.

Finally, the sensors are simulated by using a simple function. The selected illuminant from the first stage is normalized using the chosen method from the second stage and then used in the calculation of the sensors' responses to skin and Munsell reflectances. The size of gamuts are calculated and utilized in the evaluation of different sensor combinations.

Methods

We have divided the study into three separate stages with different goals. The goal of the first stage is to find a representative set of Planckian illuminants from the range of 2000 K -10000 K for further processing. Next, we study the effect of normalization methods for illuminants and the camera model. The purpose is to select one normalization method for the filter simulation. In the last stage, we test different simple filters and their combinations for skin detection. The results are compared against those from a real camera.

Stage one: The Set of Planckian Illuminants

Planck's law[7] for illumination spectral power distribution (SPD) was used because the normalization step can be applied separately to the data. The law models the light emitted from a heated black body and it has been shown to correspond quite well with real illuminants in practice[11]. According to it, the SPD of the emitted light depends on the temperature of the black body which is often referred to as the color temperature.

The color temperature is not linearly related to the SPD and this is the problem. For example, the SPD of the light changes much more between 2000 K and 3000 K than between 9000 K and 10000 K. Therefore, it is difficult to select a representative set of illuminants based on their color temperature. To solve this problem, it has been suggested we use inverse color temperature values. We study both approaches with the Sony camera in normalized color coordinates (NCC). The responses of the Sony camera are calculated using eq. 1 which gives plain responses of a channel without white balancing:

$$X = \sum I_n(\lambda) * x(\lambda) * o(\lambda) \tag{1}$$

where X = output for a channel (R, G, or B), = wavelength in nm, I = illumination SPD, n = normalization (0 = unnormalized), x = spectral response of a channel (r, g, or b, corresponding to X), and o = spectral reflectance of an object. Because the chromaticity response for SPDs is sought, the object is white with a constant value over the wavelength range. The output from eq. 1 is converted to the NCC chromaticity coordinates (eq. 2):

$$NCC\ X = \frac{X}{R+G+B} \tag{2}$$

Note that in Eq. 2 X, R, G, and B are calculated under the same illumination for the same object. From the obtained three coordinates only two chromaticity coordinates are needed because NCC r+NCC g+NCC b=1. These two are used to calculate a Euclidean measure m for the SPD, for example with NCC r and NCC g in eq. 3.

$$m = \sqrt{(NCC\ r)^2 + (NCC\ g)^2} \tag{3}$$

This equation is used to calculate one number value for a chromaticity pair as well as the difference between chromaticities of two SPDs.

Stage 2: Normalization of Illuminants

The channel response of a camera depends on three factors, as shown in Eq. 1. In general, the output for a channel is white balanced and scaled:

$$X = v * \frac{\sum I_{np}(\lambda) * x(\lambda) * o(\lambda)}{\sum I_{nc}(\lambda) * x(\lambda)} \tag{4}$$

where p refers to prevailing, n normalization, and c to calibration illumination; v is a constant used to scale the results at the wanted output range, like 0-255. As it is easy to observe from Eq. 4, the normalization of the illumination effects the results.

Some sort of normalization is typically applied to the SPD because the level and power of the illumination SPD can vary drastically. The main purpose is to scale the output so that the results are comparable under different illuminants. Here, we consider four different cases. The unnormalized spectral power distribution will, here, be marked by setting n=0:

$$I_0 = SPD \tag{5}$$

The most common normalization method is to divide all values by the value from an arbitrarily selected wavelength.

The wavelength is typically selected to be 560 nm. For this normalization method, n=1:

$$I_1 = 100 * \frac{SPD}{SPD(560\,nm)} \tag{6}$$

The basic idea of this method is to produce equal power at a wavelength. To avoid arbitrary wavelength selection, the Euclidean normalization has been suggested (n=2):

$$I_2 = \frac{SPD}{\sqrt{\sum SPD^2}} \tag{7}$$

A totally different approach for normalization is to use the response of the camera to a white object. The output is scaled using the sum of channel response to white WP (I_3):

$$\frac{1}{WP} = \frac{\sum r(\lambda) I_{0p}(\lambda)}{\sum r(\lambda) I_{0c}(\lambda)} + \frac{\sum g(\lambda) I_{0p}(\lambda)}{\sum g(\lambda) I_{0c}(\lambda)} + \frac{\sum b(\lambda) I_{0p}(\lambda)}{\sum b(\lambda) I_{0c}(\lambda)} \tag{8}$$

Of course, even this coefficient can be scaled to the wanted range. When the prevailing illumination is the same as the calibration one, the scaling factor WP=1 and it has no effect on the results. The WP scales down (or up) only when the two SPD does not match and is constant for all channels under the conditions. Other possibilities for a scaling factor would be the sum of response for two channels or for maximum channel value. In general, the effect of illumination normalization should disappear when the

prevailing illumination is the same as the calibration one (p=c, $I_{nc}=I_{np}$):

$$X = v * \frac{F_{nc} I_p(\lambda) * x(\lambda) * o(\lambda)}{F_{nc} I_c(\lambda) * x(\lambda)} = v * \frac{F_{nc}}{F_{nc}} \frac{I_p(\lambda) * x(\lambda) * o(\lambda)}{I_c(\lambda) * x(\lambda)} \quad (9)$$

where F is a normalization constant calculated for normalization method n. If the prevailing and calibration illumination differs, then a factor appears which depends on the illumination conditions:

$$X = v * (\frac{F_{pn}}{F_{cn}}) * \frac{\sum I_p(\lambda) * x(\lambda) * o(\lambda)}{\sum I_c(\lambda) * x(\lambda)} \quad (10)$$

The normalization dependent factor for cases I_1 and I_2 is not always enough to scale the output to the range [0, 255]. To obtain more realistic results, some other options need to be considered. The most simplest way is to set the maximum output value to be 255. However, this may cause severe distortion. Maybe a more sophisticated approach is to the use "knee"-function which is found in many real cameras. The knee function means that after a certain value ("knee point") the slope for the response to the input is lower. For example, we determined the following "knee-function" from the Sony manual:

$$Output\ for\ channel\ X = \begin{cases} X, 0 < X < 244 \\ 244(1-a) + aX, 244 < X \le 1109 \\ 255, X > 255 \end{cases} \quad (11)$$

where a is a knee factor (from the graph of Sony's manual we obtained the value 0.0127). Another option is not to use any normalization and use the white coefficient WP to scale down the results.

Illumination normalization or white coefficient can be considered as a constant scaling factor and it will cancel out if NCC chromaticities are calculated using responses from Eq. (4). While it might be tempting to use Eq. (1) with the white coefficients, it is not suitable for calculating the chromaticities under different calibration / prevailing illumination conditions. If the white balancing has been modeled with white coefficient WP, it will cancel out in chromaticities because the constant white coefficient WP will disappear.

Stage 3: Simple Sensors and Their Gamuts

In the last stage, the sensors are modeled with Matlab's gbellmf-function. It produces simple curves as shown in Fig. 1. The curves used in simulation have one peak and bandwidth from 30 nm (sensor 3) to 150 nm (sensor 15). Additionally, we use an impulse function which has value one at a wavelength and is referred to as sensor 1. The position of a sensor in the visible wavelength range is referred to by its peaks position.

The camera simulation also uses the results obtained from the two earlier stages. The representative illumination set is obtained from the first stage and the illumination normalization method from the second stage. Using this information, the responses of different sensors and their positions are simulated for skin and Munsell reflectances. The gamut of skin and Munsell sets is calculated for different sensor combinations using Matlab's convhulln-function. The evaluation of sensor combinations is based on the size of the gamut. The goal is to find three sensors which have a minimum skin gamut/Munsell gamut ratio. The results are compared to those from the Sony camera.

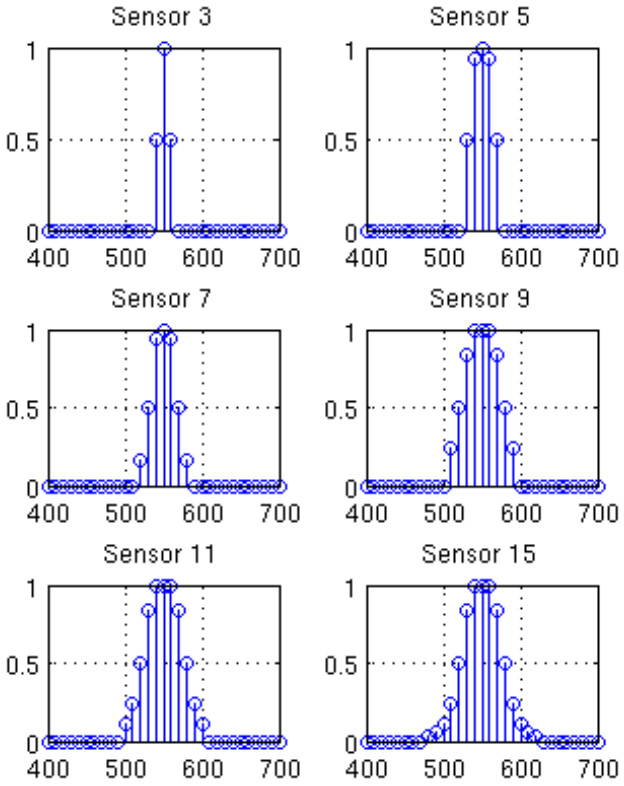

Figure 1. Simple curves generated by Matlab's gbellmf-function. The sensors numbers 3, 5, 7, 9, 11 and 15 are shown. The bandwidth of sensors 3-15 can be obtained by multiplying the sensor number by 10 nm.

Results

Stage 1: The Set of Planckian Illuminants

The responses of a Sony camera are calculated using Eq. 1 for SPDs from 2000 K to 10000 K (by 100 K steps). The obtained RGB values are converted to NCC chromaticities using Eq. 2. The upper row of Fig. 2 shows the NCC r,g, and b chromaticities as a function of color temperature and its inverse. As can be observed from the upper left image, the chromaticities are a nonlinear function of color temperature.

Using the inverse range, the nonlinearity decreases but still not enough. For example, the difference in the inverse of 9000 K and 10000 K is 0.0116 for NCC r, 0.0071 for NCC b and 0.0208 in their combinations in Euclidean distance. With the same amount of change from the inverse of 2000 K, the values are 0.0097, 0.0040 and 0.0105, respectively. Note that the difference between 2000 K and 3000 K is much bigger, 0.1665 for NCC r, 0.0930 for NCC b and 0.1912 in the Euclidean sense. Therefore, neither color temperature range nor its inverse is used for illumination selection. To solve this situation, the NCC chromaticities can be directly used for determining the wanted illuminants. The left image of the lower row of Fig. 2 shows the NCC b vs. NCC r. The linearity is not surprising because the SPDs' changes occur mainly in the red and blue regions. The right lower image shows the chromaticity pair combined into one coefficient value using Eq. 3. We use the coefficient from NCC g and b because it increases as a function of color temperature. The set of ten illuminants is selected so that they have an approximately equal distance between their coefficient values. The color temperature values for these illuminants are 2000 K, 2200 K, 2500 K, 2800 K, 3300 K, 3900 K, 4800 K, 5900 K, 7500 K and 10000 K and they are referred to by numbers 1-10.

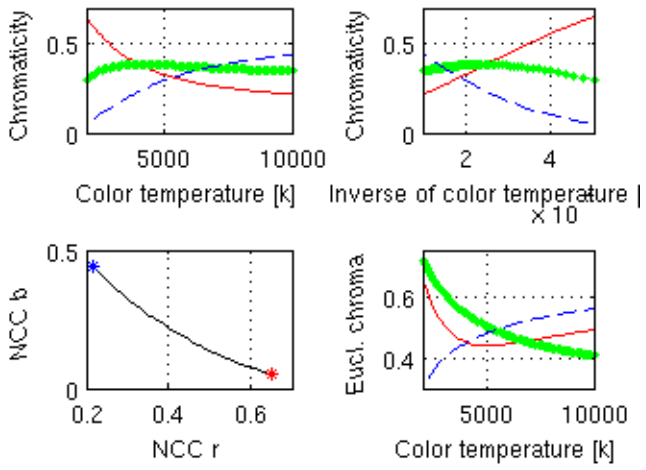

Figure 2. NCC chromaticity response of Sony camera for different Planckian illuminants. The upper row shows the chromaticity value as a function of color temperature (left image) and its inverse (right image). Neither relation is linear. The left image on the lower row shows the SPD change in NCC r and NCC b coordinates and the right one displays the Euclidean measure calculated for different chromaticity pairs.

Stage 2: Normalization of Illuminants

First, we compare chromaticities obtained from real camera measurements and camera simulations with different normalization methods. For this purpose, we have taken images of a Macbeth chart with a Sony camera under two Planckian like light sources, Horizon (H) 2300 K and A 2856 K. The comparison was made only with those color patches which had 90% of their values between 10 and 245 for all four prevailing light and calibration light cases. The results are shown in Table 1. The mean and standard deviation are calculated from absolute differences. The worst results were obtained by using white coefficient WP and Eq. 1 which was not used in any further tests. The performance of the other three normalizations (Eq. 4 and I_1, I_2, and WP) was equal.

For the three remaining normalization methods, we compared the gamuts which they produced for the Munsell reflectances under ten illuminants with color temperatures obtained from the previous stage. For the common normalization (I_1) and Euclidean normalization (I_2), the knee-function as presented in Eq. 11 was used to limit the output at the wanted range of 0-255. The evaluation was based on their gamut sizes from Matlab's convhulln-function (implementation of Qhull). The results for the normalization by 560 nm are presented in Fig. 3. It was noticed that the biggest relative size does not occur for the calibration illuminant as one might expect. This was also true for the Euclidean illumination normalization method (see Fig. 4). But when the white coefficient WP normalization was utilized, the calibration illuminant had the biggest volume size, as demonstrated in Fig. 5.

Table 1. Simulated vs. Real Chromatic Responses.

Case	Light: prevailing / calibration	Difference	
		Mean	Standard deviation
Eq. 1, WP	AA	0.148	0.033
	AH	0.117	0.028
	HH	0.250	0.034
	HA	0.269	0.044
Others	AA	0.038	0.013
	AH	0.048	0.016
	HH	0.041	0.018
	HA	0.041	0.018

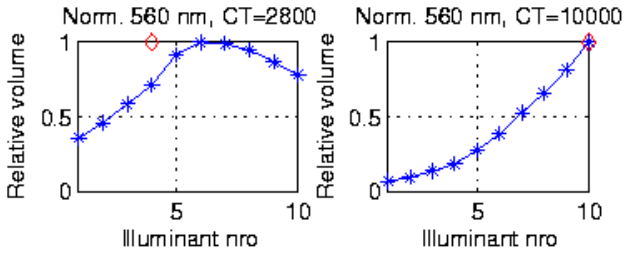

Figure 3. Relative sizes of Munsell gamuts are visualized for the common normalization. The calculations were done under one white balancing conditions marked with an open diamond in the images. The biggest relative size does not always co-incide with the calibration illuminant (CT).

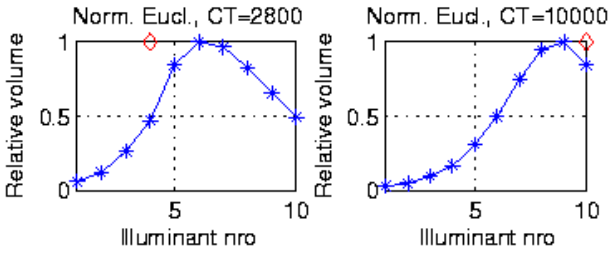

Figure 4. The relative sizes of Munsell gamuts for Euclidean normalization produce the same phenomena as the common normalization: the calibration illuminant does not have the biggest gamut size.

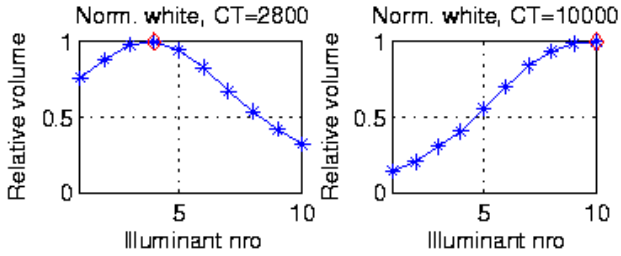

Figure 5. Normalization by the white coefficient produced a match between the calibration illuminant and the biggest volume size.

Because the mismatch between the biggest volume size and calibration was unexpected I_1 and I_2, we visualized their gamuts in RGB coordinates. As Fig. 6 displays, the gamut size increases in some cases, when the calibration illuminant is not the same as the prevailing one. For the Euclidean normalization, the sizes of the gamuts appeared similar.

We also computed the size of skin and Munsell gamuts for each calibrated case (canonical cases) and their non-calibrated cases. Figure 7 shows the results for the WP normalization. The I_1 and I_2 normalizations predicted the biggest volume for the lower color temperatures. These two normalizations were not used for the last stage because of this and because they need the knee-function (the general applicability of the knee is not known).

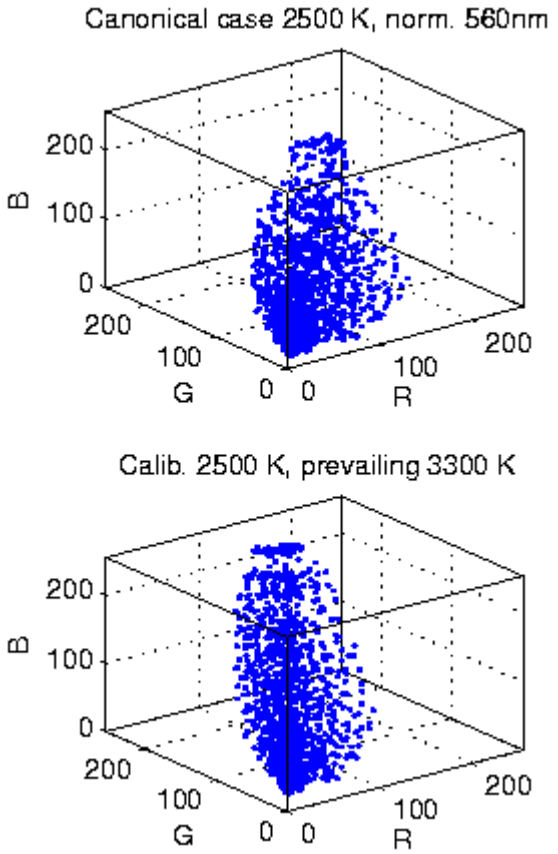

Figure 6. The gamut in the upper image for canonical case is smaller than the one in the lower image for which the prevailing and calibration illuminant are not the same. They are both obtained using the common normalization. Note also the effect of knee in the lower image.

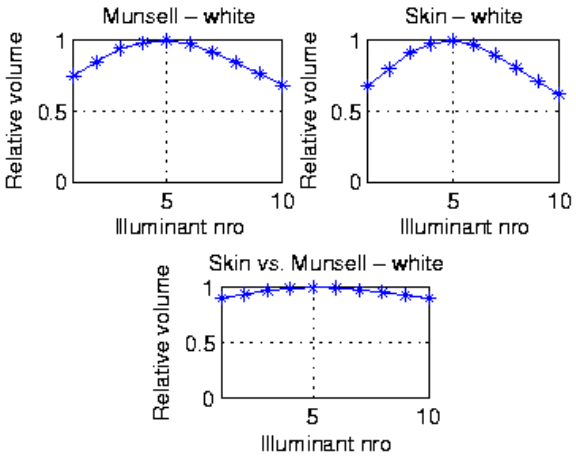

Figure 7. The gamut are calculated for each calibration sets. One set consists of one canonical case and nine uncanonical cases. The maximum sizes for Munsell and skin (upper row) is predicted to be at illuminant 5 (3300 K). The ratio between gamut sizes is shown in the lower row and it is quite constant.

Stage 3: Simple Sensors and Their Gamuts

The gamut sizes for different sensor combinations are calculated for one calibration case (one canonical case and nine uncanonical cases) and for all calibrations. Table 2 presents the triplets with maximum gamut sizes of Munsell and skin sets for different sensors (the position of the sensor is fixed with their peak position wavelength) when only one calibration is used. For the maximum size, the color temperature seems be the same with a Sony camera and most of the simulated sensors. The minimum value for the ratio between skin and Munsell gamuts were examined and the sensor combinations of these are presented in Table 3. When minimum value occurs at sensors very near each other, it might be better to use some other criteria.

The gamut results for all cases are shown in Tables 4 and 5. Note peaks' position for minimum gamut values.

Table 2. Sensor Combinations for Maximum Gamut Size.

Sensor	*Position of sensor triplets for skin at CT*	*Position of sensor triplets for Munsell at CT*
1	400 590 700, 3900 K	420 530 700, 3300 K
3	400 530 700, 3300 K	430 540 700, 3300 K
5	400 540 700, 3300 K	430 540 700, 3300 K
7	400 550 700, 3300 K	430 540 700, 3300 K
9	400 550 700, 3300 K; 400 560 700, 3300 K	430 540 700, 3300 K
11	400 550 700, 3300 K	430 540 700, 3300 K
13	400 550 700, 3300 K	430 540 700, 3300 K
15	400 550 700, 3300 K	430 540 700, 3300 K

Table 3. Sensor Combinations for Minimum Ratio Value.

Sensor	Positions of sensor triplets and CT
1	420 550 560, 2000 K
3	420 550 560, 2000 K
5	530 540 550, 2800 K; 530 540 550, 3300 K
7	520 540 550, 2200 K
9	510 520 530, 2800 K, 3300 K, 3900 K
11	520 530 540, 2500 K, 2800 K, 3300 K, 3900 K, 4800 K
13	520 530 540, 2000 K, 2200 K, 2500 K, 2800 K, 3300 K, 3900 K, 4800 K, 5900 K, 7500 K
15	520 530 540, 2000 K, 2200 K, 2500 K, 2800 K, 3300 K, 3900 K, 4800 K, 5900 K

Table 4. Sensor Combinations for Maximum Gamut Size.

Sensor	*Sensor triplets for skin*	*Sensor triplets for Munsell*
1	470 610 700	430 530 700
3	470 610 700	430 540 700
5	470 600 700	440 550 700
7	470 600 700	440 550 700
9	470 590 700	440 550 700
11	470 590 700	440 550 700
13	470 590 700	440 550 700
15	470 590 700	440 550 700

The optimal sensor combination in maximum gamut size sense depends on whether one or several calibrations are used (Tables 2 and 5). The minimum gamut size by itself might not be the best evaluation criteria for selecting optimal sensors for skin detection.

Conclusion

The design of a simple, 3-sensor camera was studied in three stages: selection of illuminants, selection of illumination normalization method, and evaluation of the obtained sensor responses.

The illuminants for the simulation were selected based on their Euclidean chromaticity value which should guarantee their representativeness in the color temperature range of 2000 K-10000 K. The illumination normalization issues were considered and one based on a white scaling factor was used in sensor simulations for Munsell and skin.

The combination of sensors which produce the maximum gamut size was found to be different for skin and Munsell sets and depends on whether one or several calibrations are in use. The minimum ratio value between gamuts seems to suggest very close positions for sensors and this criteria by alone might not be good for selecting an optimal sensor set for skin detection.

Table 5. Sensor Combinations for Minimum Gamut Size.

Sensor	Sensor triplets
1	550 570 580
3	540 550 560
5,7	530 540 550
9	440 450 460
11	450 460 470
13, 15	440 450 460

References

* Currently at InFotonics Center, University of Joensuu, Joensuu, Finland

1. F.H. Imai, N. Tsumura, H. Haneishi and Y. Miyake, "Principal component analysis of skin color and its application to colorimetric reproduction on CRT display and hardcopy", The Journal of Imaging Science and Technology 40(5): 422-430, 1996.
2. B. Martinkauppi and M. Soriano, "Basis functions of the color signal of skin under different illuminants". Proc. 3rd International Conference on Multispectral Color Science, Joensuu, Finland, 21-24, 2001
3. M. Störring, H.J. Andersen and E. Granum, "Physics-based modelling of human skin colour under mixed illuminants", Journal of Robotics and Autonomous Systems, 35(3-4): 131-142, 2001.
4. M. Soriano, B. Martinkauppi, S. Huovinen and M. Laaksonen, "Adaptive skin color modeling using the skin locus for selecting training pixels", Pattern Recognition, 36(3): 681-690, 2003.
5. B. Martinkauppi and M. Pietikäinen, "Facial skin color modeling", In: S.Z. Li and A.K. Jain (eds) Handbook of Face Recognition, Springer, 121-145, 2004.
6. P.L. Vora and H.J. Trussell, "Mathematical methods for the design of color scanning filters", IEEE Transactions on Image Processing, 6(2): 312 – 320, 1997.
7. G. Wyszecki and W.S. Stiles, "Color science: concepts and methods, quantitative data and formulae (2nd edition), 2000.
8. D.B. Judd, "Sensibility to color-temperature change as a function of temperature, Journal of the Optical Society of America, 23: 7-14, 1933.
9. R.W.G. Hunt, "Measuring colour", 1987.
10. J. Romero, A. García-Beltrán and J. Hernández-Andrés, "Linear bases for presentation of natural and artificial illuminants", Journal of the Optical Society of America A, 14(5): 1007-1014.
11. G.D. Finlayson and G. Schaefer, "Solving for colour constancy using a constrained dichromatic reflection model", International Journal of Computer Vision, 42(3): 127-144, 2001.