Performance of Human Skin Detection in Images According to Color Spaces

Jun-Yup Kim*, Yongtae Do**
* Computer and Communication Engineering Department, kimjunyup@nate.com
** Electronic Engineering Department, ytdo@daegu.ac.kr
Daegu University, 712-714, Korea

Abstract – Skin region detection in images is an important process in many computer vision applications targeting humans such as hand gesture recognition and face identification. It usually starts at a pixel-level, and involves a pre-process of color spae transformation followed by a classification process. A color space transformation is assumed to increase separability between skin classes and other classes, to increase similarity among different skin tones, and to bring a robust performance under varying imaging conditions, without any complicated analysis. In this paper, we examine if the color space transformation actually brings those benefits to the problem of skin region detection on a set of human hand images with different postures, backgrounds, people, and illuminations. Our experimental results indicate that color space transfomation affects the skin detection performance. Although the performance depends on camera and surrround conditions, normalized [R, G, B] color space may be a good choice in general.

Keywords: human skin detection, color space, hand gesture

1 Introduction

Human is an important target in computer vision. Particularly, human hands and faces are with great importance for commercial applications. A human in images has been considered difficult to deal with compared to vehicles and solid targets. However, recent fast advances in computing and imaging machines enhance active research on human hands and faces for recognition, idenfitication, and understanding [1, 2].

Skin detection is a basic step for vision processing targeting huamds. It can be defined as the process of finding which pixels of a given image correspond to human skin. The process of skin detection generally involves a pre-process of color space transformation and classification. Many studies describing skin detection have applied color space transformation for the following benefits. First, a certain color space transformation is assumed to increase the separability between skin and non-skin classes thus improving the classification process. Secondly, it is assumed to achieve the illumination invariance. Varying illumination presents additional challenges to the task of skin detection.

This paper is organized as follows. It describes color space transformations in Section 2. Experimental results for skin detection in various color spaces are given in Section 3. Section 4 concludes the paper.

2 Color space transformations

Color of a pixel in an image is defined as [C_0, C_1, C_2]. The color space transformation is a function that converts [C_0, C_1, C_2] to [C_0', C_1', C_2']. In most off-the-shelf camera systems, images are captured in Red, Green, Blue (RGB) space. We have evaluated eleven color spaces by following four transformations: XYZ, Y, IQ, YIQ, H, HS, SV, HSV, NRG (normalized RG), NRB (normalized RB) and NGB (normalized GB) [3].

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.619 & 0.177 & 0.204 \\ 0.299 & 0.586 & 0.115 \\ 0.000 & 0.056 & 0.944 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1.a}$$

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.586 & 0.115 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1.b}$$

$$r = \frac{R}{R+G+B} \qquad g = \frac{G}{R+G+B} \qquad b = \frac{b}{R+G+B} \tag{1.c}$$

$$H = \cos^{-1}\left(\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}} \right)$$

$$S = 1 - \frac{3}{(R+B+G)}[\min(R,G,B)] \qquad (1.d)$$

$$V = \frac{1}{3}(R+G+B)$$

Note that values of each component in [R, G, B] are in the range of [0, 255] while [r, g, b] (lower-cased) are in [0, 1]. In [r, g, b] space, one value can be computed by other two values, and only arbitrary two among three components are independent. By normalization, the brightness information is lost, and only comparative intensities are available. In [H, S, V] space, H and S components are also independent of brightness, which is represented separately by V component.

3 Experiements

3.1 Datasets and cameras used

We collected a total of 372 images from three different cameras for experimental datasets. Figure 1 shows example images, which are acquired from different cameras, in different illumination, with different backgrounds, for different people, and in different postures.

The first of three cameras used is the SK-2146AIN from Sunkwang Electron, which uses a 1/3" Sony Color CCD image sensor having 0.3 million pixels. AGC and BLC functions are available, and the lowest luminous intensity for picturing is 3 Lux. The second camera is CCN-261 IA from SECURA. It has 1/3" IT CCD image sensor and 0.41 million pixels. AGC and BLC functions are also available and the lowest luminous intensity for picturing is 0.8 Lux. The third camera is the PYRO 1394 Web Cam of the ADS Technologies. It has 0.3 million pixels. AGC and WBS functions are available and the lowest luminous intensity for picturing is 2.7 Lux.

3.2 Pixels classified for hand regions and non-hand regions

Pixels of human hands and non-hand regions are collected for learning and testing. For the ease of collection and classification, we set a small box inside hand region, which is represented as *region a* in Figure 2. *Region c* is used to collect non-hand pixels while pixels of *region b* are not used. For pixels of hand regions, mean (μ) and a standard deviation (σ) are computed. Then, pixels satisfying Equation (2) are selected as skin pixels.

Figure 1. Experimental images: Images at the left clolumn are those acquired by SK-2146AIN, while images at center and right columns are acquired by CCN-261 IA and 1394 Web Camera respectively. Images at the top row are used for learning parameters while other images are all for testing; Images of second row are taken in dark brightness, Images of third row are taken at different background, Images of fourth row are hands of different people, Images of bottom row are hands in different posture.

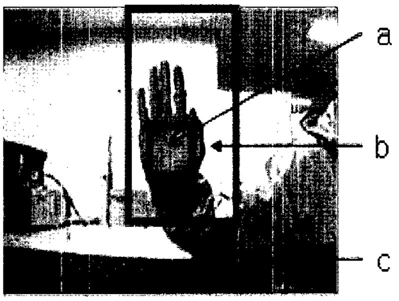

Figure 2. Hand region and other region in image

$$\mu - 2\sigma \le I(x,y) \le \mu + 2\sigma \qquad (2)$$

Figure 3 shows the results of experiment. Black pixels are those determined as hand skin. Twelve color spaces are tested including [R, G, B]. To remove noise effect, morphological operation has been applied and the results are given in Figure 4.

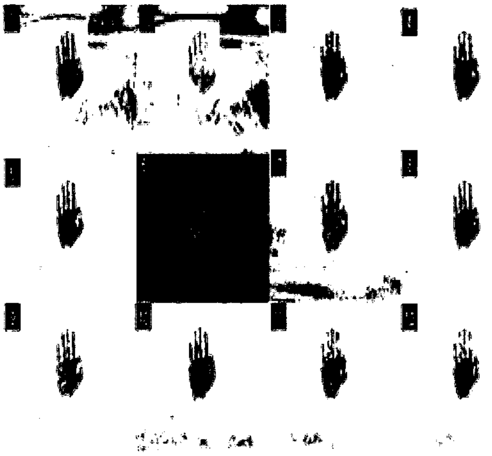

Figure 3. Skin region detection results for different color spaces : 1(RGB), 2(XYZ), 3(I), 4(IQ), 5(YIQ), 6(H), 7(HS), 8(SV), 9(HSV), 10(rg), 11(rb), 12(gb)

Table 1. Skin pixel detection success rates

Images	Learning image	Different back-ground 1	Different back-ground 2	Different bright-ness 1	Different bright-ness 2	Different hand posture1	Different hand posture2	Different person 1	Different person 2
SK – 2146 AIN									
RGB	96.1	55.8	46.8	48.9	48.8	74.2	65	88.2	97.2
XYZ	95.2	53.3	45.3	47.8	46.2	73.3	63.7	90	97.1
YIQ	92.8	51.5	49.8	50	50	71	68.3	92.2	87.5
HSV	92.2	55.4	48	49.6	49.8	70.8	67.7	85.1	91.5
S_rg	96.5	89.6	92.1	97.9	68	87.1	88.1	94.5	93.3
S_rb	95.9	81	87.9	97.3	66.2	83.8	92.6	91.6	92.8
S_gb	96.4	80.2	88.4	97.8	67.7	83.6	86.8	92.4	91.9
CCN – 261 IA									
RGB	96.4	67	69	71.7	75	75	78.6	95.5	89.6
XYZ	93.9	61.9	66.5	67.5	73.1	71.9	72.7	94.4	81.3
YIQ	96.1	71.1	71.8	74.7	53.2	73.3	80.9	97.7	73.9
HSV	95.3	75.1	76.1	77.4	51.5	74.4	75.8	97.9	81.8
S_rg	92.5	90.8	82.5	91.2	50.1	82.1	93.6	94.3	85.4
S_rb	87.5	88.6	80.7	82.4	49.2	78.1	91	93.1	85.1
S_gb	87.4	93.3	87.7	83.8	49.5	81	89.8	94.7	92.8
1394 Web Cam									
RGB	91.2	47.8	45.8	46.7	43.6	69.7	72.3	86.2	92.4
XYZ	85.5	43.7	42.3	44.8	36	62.9	65.7	74.6	85.9
YIQ	93.2	51.1	50	50.1	50.1	71.1	72.9	70.4	92.3
HSV	86.1	49.8	47.8	48.6	48.8	69	71.9	76.4	85.7
S_rg	93.9	91.1	96.5	95.4	80.9	77.7	81.9	76.8	89.9
S_rb	90	88.7	90.8	88.3	78.9	72	72.5	64.5	83.6
S_gb	90	86.3	91	88.2	77	73.2	71.7	62.9	81.1

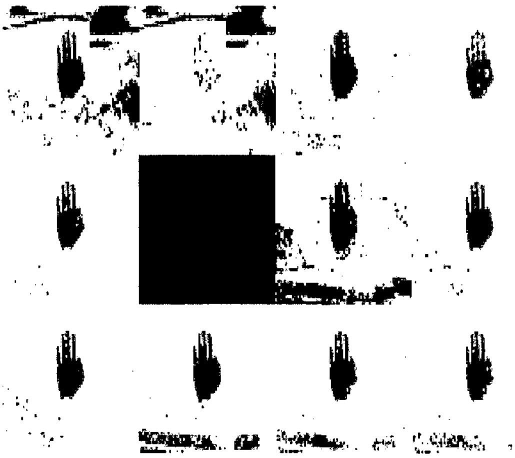

Figure 4. Results after morphological operation

3.3 Color space transformation and hand region detection efficiency

Table 1 shows the results of experiments done for different situations. In conclusion, the color space transformation gives a help to skin pixel detection although the efficiency is different for cameras. For example, SK-2146AIN and 1394 web camera showed good efficiency for most situations in [r, g, b] while CCN-261 IA camera was good in [Y, I, Q].

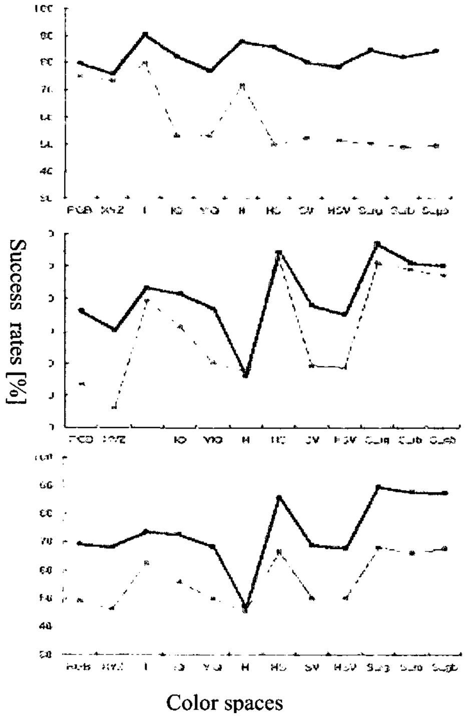

Figure 5. The detection success rate for different color spaces and cameras: (Top) SK-2146AIN, (Middle) CCN-261 IA, (Bottom) 1394 Web Camera. Darker upper line shows the average performance while brighter lower line shows the worst for each color space.

4 Conclusions

Skin detection is an important process in many of computer vision tasks targeting humans. A color space transformation is assumed to increase separability between skin and non-skin classes, to increase similarity among different skin tones, and to bring a robust performance under varying surrounding conditions, without any sound reasoning. In this paper, we examined if a color space transformation brings those benefits on a dataset of 372 images with different cameras, different illumination, different backgrounds, different people and different hand postures. We found that a color space transformation may make more efficient skin pixel detection than detection in [R, G, B] space, and the efficiency varies depending on camera and condition. In general, however, normalized [R, G, B] color space can be a first choice because it showed good results in many cases in our experiments.

References

[1] Vladimir I. Pavlovic, Rajeev Sharma and Thomas S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, July 1997

[2] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja, "Detecting faces in images: A survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, Jan., 2002

[3] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing 2/E, Prentice Hall, September 2003