ADAPTIVE SKIN DETECTION USING MULTIPLE CUES

Jinfeng Yang, Zhouyu Fu, Tieniu Tan and Weiming Hu

National Laboratory of Pattern Recognition(NLPR)
Institute of Automation, Chinese Academy of Sciences. P.O.Box 2728, Beijing, P.R.China
{jfyang, zyfu, tnt, wmhu}@nlpr.ia.ac.cn

ABSTRACT

This paper presents an adaptive approach to skin detection. First, we propose a nonlinear relationship among R, G and B components and use a closed curve to identify the skin cluster region. Then, a split machine is designed that aids the extraction of the pixels with similar low-level features from images. Finally, a nonlinear skin color classifier with an adaptive threshold is developed by analyzing the properties of the extracted pixels in the HSL, YCbCr, YUV and YIQ color spaces. Experimental results show that our proposed method works very well in skin detection.

1. INTRODUCTION

Skin information contributes much to object recognition in many cases. Generally, detecting skin information in images is concerned with two main problems: the selection of the suitable color space and the modelling of the skin color distribution. Besides the RGB color space, other color spaces transformed from RGB have been selectively used for modelling the skin color distribution, such as normalized RGB, HSV/HSI/HSL, TSL, YCrCb, YES, YUV and YIQ [1, 5, 6], etc. However, it is still an open problem that which color space is the best one in distinguishing human skin colors from non-skin colors. An accurate description of skin color distribution is still not achieved at present, especially with respect to complicated backgrounds and illumination, although researchers have proposed many methods, such as single Gaussian [1], mixtures of Gaussians [13], multiple Gausian clusters [16], elliptic boundary [15], Bayes classifier [2, 5] and self organizing map (SOM) [14], etc. Fortunately, previous work also gives us some hints on understanding the skin color property. Albiol et al. testified in theory that the difficulty in discriminating skin and non-skin classes was independent of color spaces [9]. Shin et al. found that color space transformation does not increase the separability between the skin and non-skin classes [7]. Jones and Rehg found their non-parametric model to be superior in both accuracy and computational cost according to the comparison results between parametric skin models and a statistical color model in RGB space [2]. Note that due to the overlap between the skin and non-skin colors, the pixel-based skin and non-skin color models do not work well when the discriminability between the skin and non-skin colors does not make much difference according to the adopted decision rules. Hence, some adaptive schemes [10, 11, 12] have been proposed for better skin segmentation in images with complex backgrounds. Here, we introduce a new method to solve the problem in skin segmentation with respect to the complexities of static images.

2. SKIN COLOR DISTRIBUTION

Human skin colors cluster spatially in whatever color space used [6]. Although the R, G and B channels correlate with each other, the RGB space is still reliable in describing the characteristic of skin colors [7]. The spatial clustering property of the skin colors indicates the existence of a specific relationship among the R, G and B components. Here, we propose a nonlinear relationship among the three color components as follows

$$\begin{cases} \rho = ((\alpha_1 R - \alpha_2 G)^2 + (\alpha_2 G - \alpha_3 B)^2)^{1/2} \\ \theta = \begin{cases} \theta_{t_1}, & \alpha_1 R - \alpha_2 G \geq 0 \\ \pi + \theta_{t_2}, & \alpha_1 R - \alpha_2 G < 0, \alpha_2 G - \alpha_3 B \leq 0 \\ 2\pi - \theta_{t_2}, & \alpha_1 R - \alpha_2 G < 0, \alpha_2 G - \alpha_3 B > 0 \end{cases} \end{cases} \quad (1)$$

where $\theta_{t_1} = \arccos((\alpha_2 G - \alpha_3 B)/\rho)$, $\theta_{t_2} = \arccos(|\alpha_2 G - \alpha_3 B|/\rho)$. In practice, we choose the normalized r, g and b to replace α_1, α_2 and α_3 respectively since each normalized component can be viewed as a 'color purity' of its corresponding component in RGB space [1]. The statistical skin and non-skin distributions computed from our constructed pixel datasets are demonstrated in Fig. 1 respectively. Here the skin pixel set contains 40,803,483 samples, and the non-skin pixel set contains 54,710,462 samples. From Fig. 1, we see that the skin colors cluster in the proposed (ρ, θ) plane prominently, while the non-skin colors spread along the θ axis. Although the skin colors cluster very well in the proposed plane, our experiments show modelling this distribution with single Gaussian cannot achieve better skin detection results. Fig. 1(b) shows that the skin cluster region is geometrically circumscribable, so we identify the skin clus-

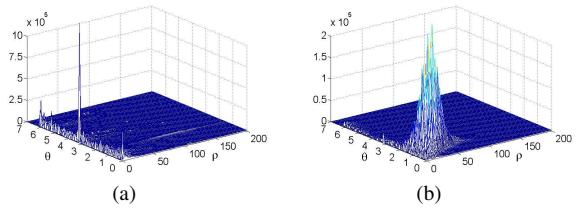
0-7803-8554-3/04/$20.00 ©2004 IEEE.

ter region with a closed curve, as shown in Fig. 2(a). This closed curve consists of three piecewise sub-curves, which enclose 98.4% of skin pixels in our skin pixel set. The skin pixels outside the closed curve are viewed as noise here. The Equation of the closed curve is expressed as

$$S(\rho,\theta)=\begin{cases} b^2(\tilde{\rho}_1-\rho_0)^2+a_1^2(\tilde{\theta}_1-\theta_0)^2-a_1^2b^2, f_1\le 0 \\ b^2(\tilde{\rho}_2-\rho_0)^2+a_2^2(\tilde{\theta}_2-\theta_0)^2-a_2^2b^2, f_2\le 0 \\ b^2(\tilde{\rho}_3-\rho_0)^2+a_3^2(\tilde{\theta}_3-\theta_0)^2-a_3^2b^2, f_3\ge 0 \end{cases} \quad (2)$$

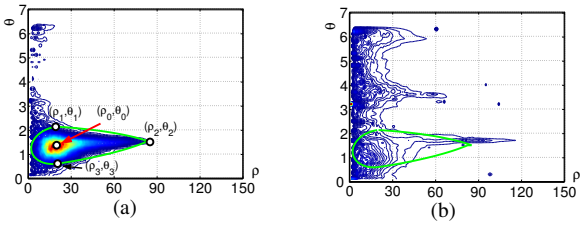
$$\begin{cases} f_1=\rho-\rho_0 \\ f_2=(\rho^2-\rho(\rho_1+\rho_2)+\rho_1\rho_2)(\theta-\theta_2) \\ f_3=(\rho^2-\rho(\rho_2+\rho_3)+\rho_2\rho_3)(\theta-\theta_2) \\ \tilde{\rho}_i=\rho\cos\beta_i-\theta\sin\beta_i \\ \tilde{\theta}_i=\rho\sin\beta_i+\theta\cos\beta_i. \end{cases} \quad (3)$$

The three sub-curves constrained by f_1, f_2 and f_3 are segmented piecewise from three concentric yet differently oriented ellipses centered at (ρ_0,θ_0). $\beta_i(i=1,2,3)$ denotes the orientation of the *i*th ellipse in the (ρ,θ) plane. Some parameter values in Eq. (2) are listed in Table 1.



(a) (b)

Fig. 1. The skin and non-skin color distributions. (a) The Non-skin distribution. (b) The skin distribution.



(a) (b)

Fig. 2. A closed curve fitting the skin cluster region. (a) The selected skin cluster region. (b) The non-skin region enclosed by this curve.

Table 1. The values of some parameters

(ρ,θ)	0	1	2	3	$\beta_1(a_1)$	$\beta_2(a_2)$	$\beta_3(a_3)$
$\rho\_$	20.2	19.3	85	21.1	$\pi/36$	$-\pi/36$	$\pi/36$
$\theta\_$	1.37	2.15	1.45	0.62	(20.2)	(64.2)	(64.2)

Obviously, the region enclosed by this closed curve contains a certain number of non-skin pixels, as shown in Fig. 2(b). In fact, color space transformation cannot completely eliminate the overlap between skin and non-skin colors in the statistical manner. Also, in many cases, whether a pixel is interpreted as skin color or not is dependent on the imaging environments. Hence, in Section 3, we introduce an adaptive method of skin segmentation for static images.

3. ADAPTIVE METHOD FOR SKIN SEGMENTATION

3.1. Training pixels in an image

Since not every region in an image contains skin information, we design a split machine (SM) which can quickly generate an adaptive grid over potential skin regions by recursively dividing a grid cell into four equal parts under predefined criterion, as shown in Fig. 3.

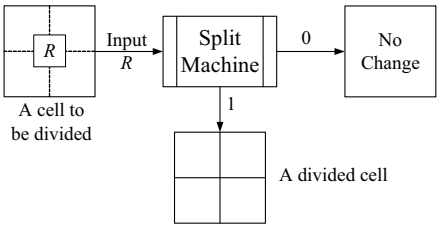


Fig. 3. The split machine (SM). The information in *R* determines whether a cell is divided or not.

In practice, it is unnecessary to exploit the low-level information within a whole grid cell to determine whether a cell is divided or not. Hence we select a square region *R* occupying one ninth of a cell as its representative, as shown in the top-left of Fig. 3. Suppose Q_1 is the set of all pixels in *R*, and Q_2 is the set of detected skin pixels in *R*, we can calculate the approximate probability density of the grey levels in the two sets: $p_k(i)=h_k(i)/N_k(i=0,1,...G_k-1;k=1,2)$, where N_k and G_k are the total numbers of pixels and intensity levels in the *k*th set, and $h_k(i)$ is the number of pixels in the *i*th intensity level. Thus, we compute the following statistical values used in texture analysis [17].

$$\begin{cases} \mu_k=\sum_{i=0}^{G_k-1} ip_k(i) \\ \sigma_k^2=\sum_{i=0}^{G_k-1}(i-\mu_k)^2p_k(i) \\ \mu_{4k}=\sigma_k^{-4}\sum_{i=0}^{G_k-1}(i-\mu_k)^4p_k(i)^{-3} \end{cases} \quad (4)$$

Here, we are only interested in μ_{4k} since μ_{4k} is a measure of flatness of the histogram. Let $g_1=S_{skin}/S_R$ and $g_2=\mu_{42}/\mu_{41}$, where S_{skin} denotes the detected skin area in *R*, and S_R denotes the area of *R*, the criterion for the SM is defined as

$$C_{SM}(R)=\begin{cases} 1, & g_1\ge\beta_1, g_2\ge\beta_2 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where $\beta_1=0.3$ and $\beta_2=0.65$. The introduction of g_2 aims to discard the regions with high texture information, since the skin region always shows little texture. For a cell, if $C_{SM}(R)=1$, the SM splits it into four equal parts, otherwise it stays unchanged. In general, the fine meshes can be developed over skin regions after an initial grid is updated three times by the SM, as shown in Fig. 4.

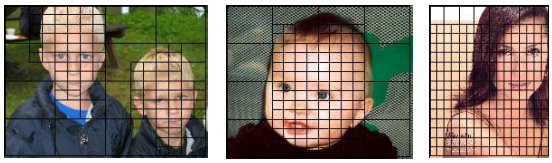


Fig. 4. Adaptive grids over some images.

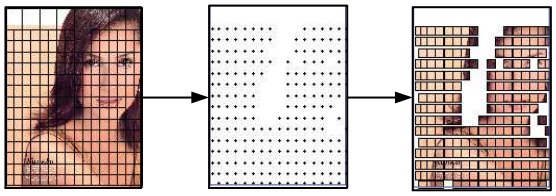


Fig. 5. The process of extracting the pixel set ω in an image.

Naturally, the SM can also generate fine meshes over the background region of an image if the skin and non-skin regions have similar low-level features, as shown in the right of Fig. 4. Hence this method is unable to obtain better skin segmentation in these cases. We must develop a technique that can adaptively explore the differences between the skin and non-skin pixels for different images. Based on the grid over an image, we extract the nodes with skin color information, and then cut out a square patch centered at each extracted node from the image, as shown in the middle and right of Fig. 5. Assume that ω is the set of all pixels in these patches, next we have to develop a method which can describe the differences of the pixels in ω.

3.2. Adaptive histogram

The proposed mapping expressed by Eq. 1 is unable to explore the potential differences of pixels in ω. From our experiments, we find that the pixel features exclusively based on the chrominance components are lack of separability. Hence, we take the luminance component into consideration in describing the attributes of the pixels in ω. Because the HSL color space works well for natural illumination, and YCbCr color space is perceptually uniform, we use the S, L, Cb and Cr components to capture pixel characteristics. Assume $d_1 = S/L$ and $d_2 = Cb/Cr$, we define

$$x = 100 \cdot \sqrt{d^2 + \alpha^2} \tag{6}$$

where $d = (d_1^2 + d_2^2)^{1/2}$ and $\alpha = \arccos(d_1/d)$. The pixels in ω are used to populate a histogram with bin width of 1 in x, as shown in Fig. 6. The bi-modal histogram shown in Fig. 6(a) illustrates the fact that the x values between the two peaks are not common in ω. Hence determining the threshold as the x_a level that has a minimum histogram value between the two peaks must meet minimum segmentation error requirements. For a uni-modal histogram, as shown in Fig. 6(b), we think that the closed curve $S(\rho, \theta)$ is enough to describe skin distribution and no further operation is needed. Note that the shapes of histograms hold the only two mentioned fashions, since the SM details the regions with similar low-level information in an image exclusively.

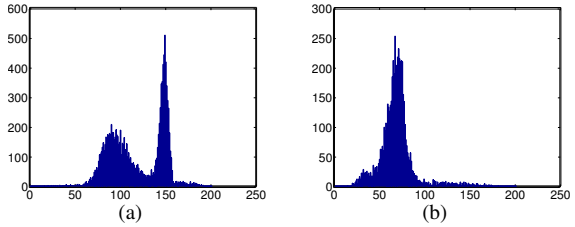


Fig. 6. The histograms. (a) A bi-modal histogram. (b) An uni-modal histogram

Now, we need to determine which peak of a bi-modal histogram represents the real skin information. Here we use the YUV and YIQ color spaces together to address this problem. Assume that $g_1(p) = I(p)/Q(p)$ and $g_2(p) = U(p)/V(p)$, where p is a pixel. Let $f_1 = g_1(p_1) - g_1(p_2)$ and $f_2 = g_2(p_1) - g_2(p_2)$, where p_1 and p_2 are two pixels falling into the two bins corresponding to the two peaks of the bi-modal histogram respectively. If $f_1 > 0$ and $f_2 < 0$ hold simultaneously, the peak that corresponds to p_1 is formed by the skin pixels, otherwise the other is formed by the skin pixels. Note that only using f_1 and f_2 to detect the skin information in images can result in very high false positive rates. Let $\tilde{x}_p$ denote the bin for the peak of skin pixels. We define a non-linear skin color classifier as

$$F_s(q) = \begin{cases} S(\rho_q, \theta_q), & \text{uni-modal} \\ S(\rho_q, \theta_q) \cdot G_1(x_q), & \tilde{x}_p < x_a \\ S(\rho_q, \theta_q) \cdot G_2(x_q), & \tilde{x}_p > x_a \end{cases} \tag{7}$$

where q is a pixel, $G_1(x_q) = x_a - x_q$ and $G_2(x_q) = x_q - x_a$. Note that we let $G_i(x_q) = 1$ for $S(\rho_q, \theta_q) > 0$. Thus, if $F_s(q) \leq 0$, the pixel q is labeled as skin.

4. EXPERIMENTS

In our experiments, we use the Compaq image database [3] to test the performance of the proposed skin color classifier. The database contains 4675 skin images with corresponding masks and 8965 non-skin images. We extract 2000 random samples from the skin and non-skin images respectively. Based on the corresponding masks of skin images, the background color of all the skin images were changed to white. Thus, the image samples contain 62,834,397 skin pixels and 195,438,297 non-skin pixels altogether. The skin pixel detection results are illustrated in Table 2. Here, Jones-Rehg's method is tested at a threshold of 0.4 and the results of ECU are reported by its "SkinColorDetector" [4]. Table 2 shows that our proposed method can reduce the false positive rate significantly as well as increase the true positive rate.

Due to space limitation, we only give some skin segmentation results in Fig. 7. In these images, the background colors and illuminations have negative effects on skin segmentation. The segmentation results show that the proposed

skin color classifier has a good performance in segmenting the skin regions from backgrounds.

Table 2. Skin and non-skin pixel detection

Methods	True positive(%)	False positive(%)
Jones-Rehg	90.88	24.31
ECU	93.26	20.78
Proposed	96.46	6.48

5. CONCLUSIONS

In this paper, we have proposed a nonlinear mapping that aims to describe the inherent relationship among pixel color components and a closed curve to identify the skin cluster region. Considering the varieties of images, we have designed a split machine that can generate adaptive grids over skin regions. For an image, an optimal threshold was acquired from the training pixel samples extracted from the local neighborhood of skin colored grid nodes. Finally, we have developed a nonlinear skin color classifier adapted to images varied with imaging environments. Experimental results showed that our approach has a good performance in skin segmentation.

6. ACKNOWLEDGEMENTS

This work is partly supported by NSFC (Grant No. 60105002, 60335010, 60121302), Natural Science Foundation of Beijing (Grant No. 4041004).

7. REFERENCES

[1] J. Yang, W. Lu and A. Waibel, Skin color medeling and adaptation. *ACCV*, 687-694, 1998.

[2] M. J. Jones and J. M. Rehg, Statistical color models with application to skin detection. *CVPR*, 247-280, 1999.

[3] M. J. Jones and J. M. Rehg, Compaq skin database. *http://www.crl.research.digital.com/publications/tech-reports/ abstracts/98_11.html*.

[4] S. L. Phung, ECU Face Detection Database. *http://www.soem.ecu.edu.au/∼sphung/face_detection/database/*

[5] D. Chai and A. Bouzerdoum, A Bayesian approach to skin color classification in ycbcr color space. In *Proc. IEEE Region Ten Conference* (TENCON'2000), 421-424, 2000.

[6] V. Vezhnevets, V. Sazonov and A. Andreeva, A survey on pixel-based skin color detection techniques. *Graphicon-2003*, Moscow, Russia, 2003.

[7] M. C. Shin, K. I. Chang, L. V. Tsap, Does colorspace transformation make any difference on skin detection? In *Proc. IEEE Workshop on Application of Computer Vision*, 2002.

[8] J. Brand and J. S. Mason, A comparative assessment of three approaches to pixel-level human skin-detection. *ICPR*, 1056-1059, 2000.

[9] A. Albiol, L. Torres and E. J. Delp, Optimum color spaces for skin detection. *ICIP*, 122-124, 2001.

[10] K-M. Cho, J-H. Jang and K-S. Hong, Adaptive skin-color filter. *PR*, 34:1067-1073, 2001.

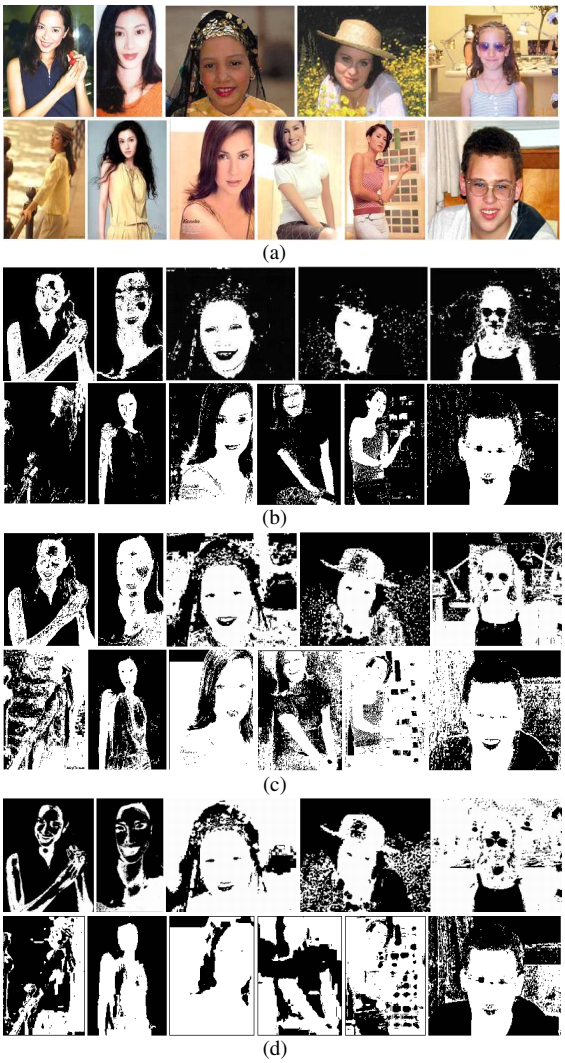


Fig. 7. Skin segmentation results. (a) Original images. (b) The proposed method. (c) ECU's method. (d) Jones-Rehg's method.

[11] S. L. Phung, D. chai and A. Bouzerdoum, Adaptive skin segmentation in color images. *ICASSP*, 353-355, 2003.

[12] L. M. Bergasa, M. Mazo, M. A. Sotelo and L. Boquete, Unsupervised and adaptive Gaussian skin-color model. *IVC*, 18:987-1003, 2000.

[13] S. J. McKenna, S. Gong and Y. Raja, Modelling facial colour and identity with Gaussian mixtures. *PR*, 31(12):1883-1892, 1998.

[14] D. Brown, I. Craw, J. Lewthwaite, A SOM based approach to skin detection with application in real time systems. *BMVC*, 2001.

[15] J. Y. Lee and S. I. Yoo, An elliptical boundary model for skin color detection. *ICISST*, 2002.

[16] S. L. Phung, A. Bouzerdoum and D. Chai A novel skin color model in ycbcr color space and its application to human face detection. *ICIP*, 289-292, 2002.

[17] A. Materka and M. Strzelecki, Texture analysis methods - A review, *COST B11 report*, Brussels, June 1998.