ADAPTIVE SKIN SEGMENTATION IN COLOR IMAGES

Son Lam Phung *Douglas Chai* *Abdesselam Bouzerdoum*
School of Engineering and Mathematics - Edith Cowan University
Perth, Australia

ABSTRACT

A new skin segmentation technique for color images is proposed. The proposed technique uses a human skin color model that is based on the Bayesian decision theory and developed using a large training set of skin colors and nonskin colors. The proposed technique is novel and unique in that texture characteristics of the human skin are used to select appropriate skin color thresholds for skin segmentation. Two homogeneity measures for skin regions that take into account both global and local image features are also proposed. Experimental results showed that the proposed technique can achieve good skin segmentation performance (false detection rate of 4.5% and false rejection rate of 4.0%).

1. INTRODUCTION

Skin segmentation plays an important role in recent color-based approaches to human face detection and segmentation [1-4]. In these approaches, regions of the input image that have skin colors are segmented to give an initial estimate of the face locations. These segmented skin regions will be further processed in the later stages of face detection. Skin segmentation also has been applied to hand segmentation and hand-gesture analysis [5], and filtering objectionable Web images [6]. Clearly, in all of these applications, results of skin segmentation have a strong influence on subsequent stages of image analysis, and consequently on the final outcome.

In this paper, we propose a skin segmentation technique for color images. The proposed technique uses a human skin color model to differentiate skin colors and nonskin colors. The model is based on the Bayesian decision rule for minimum cost, and is developed using a very large training set of skin and nonskin colors. Compared to many existing skin segmentation approaches, the proposed technique is novel and unique in that it takes into account the homogenous property of the human skin texture. This property is used to select a skin color threshold that can separate each skin region from its nearby background, even when the background has skin colors. The threshold is adapted to local image region.

The organization of the paper is as follows. The human skin color model is described in Section 2. The adaptive skin segmentation algorithm is presented in Section 3. Implementation and experimental results are discussed in Section 4. Conclusions and further work are given in Section 5.

2. HUMAN SKIN COLOR MODEL

A human skin color model is used to decide if a pixel is skin color or nonskin color. A human skin color model is characterized by a classification algorithm and a color space that is used to represent pixel color. Existing classification algorithms use neural networks (multilayer perceptions [7], self-organizing maps [8]), probabilistic classifiers based on density estimation (Gaussian [9, 10], mixture of Gaussians [11], histogram [12]), and fixed decision boundaries (rectangle [1], set of planes [2], an ellipse [13]). Color spaces that have been used in skin color detection include RGB [10, 12], YCbCr [1, 14], HSV [13], CIE Lab [15], CIE Luv [11], Farnsworth UCS [3], and normalized RGB [16].

The human skin color model used in our work is based on the Bayesian decision for minimum cost. Let x be the feature vector of a pixel. Let ω_1 be the skin color class and ω_2 be the nonskin color class. Let $p(x|\omega_1)$ and $p(x|\omega_2)$ be the class-conditional probability densities of class ω_1 and ω_2 respectively.

For each pixel x, a skin color score (also known as likelihood ratio) is defined as follows:

$$s(x) = \frac{p(x \mid \omega_1)}{p(x \mid \omega_2)}. \tag{1}$$

The higher the score $s(x)$, the more likely x can be considered as a skin color. Classification decision usually involves a fixed threshold τ:

$$\begin{aligned}&\text{if } s(x) \geq \tau \text{ then } x \in \omega_1\\ &\text{else } x \in \omega_2\end{aligned}. \tag{2}$$

There exists an optimum threshold τ in a sense that it minimizes the overall classification cost. Let $C_{ij} \geq 0$, (i, j) = {1, 2}, be the cost of deciding that a pixel x is in class ω_i while x is in fact in class ω_j. Let $P(\omega_i|x)$ be *a posteriori* probability of class ω_i; that is, $P(\omega_i|x)$ is the probability of a given sample x belonging to class ω_i. Let

$P(\omega_i)$ be *a priori* probability of class ω_i; that is, $P(\omega_i)$ is the probability of observing a sample from class ω_i.

The classification cost $R_i(x)$ of assigning x to ω_i is:

$$R_i(x) = C_{i1}P(\omega_1 \mid x) + C_{i2}P(\omega_2 \mid x) . \quad (3)$$

The feature vector x is assigned to the class with minimum classification cost:

$$x \in \omega_1 \quad \text{if} \quad R_1(x) \leq R_2(x) . \quad (4)$$

Substituting (3) into (4), and rearranging terms gives:

$$x \in \omega_1 \quad \text{if} \quad \frac{P(\omega_1 \mid x)}{P(\omega_2 \mid x)} \geq \frac{C_{12} - C_{22}}{C_{21} - C_{11}} . \quad (5)$$

From (5), we obtain, by applying the Bayesian formula:

$$P(\omega_i \mid x) = \frac{p(x \mid \omega_i)P(\omega_i)}{p(x)} , \quad (6)$$

the following relation:

$$x \in \omega_1 \ \text{if} \ s(x) = \frac{p(x \mid \omega_1)}{p(x \mid \omega_2)} \geq \frac{C_{12} - C_{22}}{C_{21} - C_{11}} . \frac{P(\omega_2)}{P(\omega_1)} . \quad (7)$$

Therefore, the optimum threshold τ in (2) is

$$\tau_o = \frac{C_{12} - C_{22}}{C_{21} - C_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)} . \quad (8)$$

Unfortunately, the above theoretical threshold is hard to estimate because *a prior* probabilities $P(\omega_i)$ are usually unknown. In practice, the threshold τ is often chosen experimentally to reflect a trade-off between false detections and false rejections. A false detection occurs when the classifier considers a non-skin color as skin color; a false rejection occurs when the classifier considers a skin color as non-skin color. From (2), a larger value of τ leads to more false rejections while a smaller value of τ leads to more false detections.

To use the decision rule summarized in (1) and (2), we need an accurate estimate of the class-conditional probability densities $p(x|\omega_i)$, $i = \{1, 2\}$. Parametric approaches that model $p(x|\omega_i)$ as a Gaussian distribution [9, 10] or a mixture of Gaussians [11] have been reported. However, we propose that in this problem a non-parametric approach using histogram is sufficient. The histogram approach is viable in terms of memory storage because the feature vector x has very low dimension (at most 3). Moreover, the histogram approach can give a very accurate estimate of conditional densities provided that a very large and representative set of labeled samples are available. In fact, a major task in developing our skin color model is to collect a large labeled set of skin colors and nonskin colors.

The histogram approach for computing conditional densities can be described briefly as follows. From a set of labeled skin colors and nonskin colors, we obtain two histograms $H_1(x)$ and $H_2(x)$. $H_1(x)$ is the count of skin color pixels that have feature vector value x, and $H_2(x)$ is the count of nonskin color pixels that have feature vector x. The class-conditional densities are simply normalized histograms:

$$p(x \mid \omega_i) = \frac{H_i(x)}{\sum_x H_i(x)} . \quad (9)$$

In our work, the 24-bit RGB (Red-Green-Blue) color space is used so $x = [\text{R G B}]^T$. Note that the proposed classification algorithm can be applied to any color space. For example, in [14, 17], the feature vector $x = [\text{Cb Cr}]^T$ is used, whereby Cb and Cr are the chrominance components of the YCbCr color space.

3. ADAPTIVE SKIN SEGMENTATION

In the previous section, we describe a simple yet very effective algorithm for classifying skin colors and nonskin colors. However, for skin segmentation purposes, pixel-level skin color detection is not enough because the image background (nonskin regions) may also have skin colors. For this reason, skin segmentation approaches that rely solely on pixel-level classification often have very high false detection. This motivates us to propose a skin segmentation algorithm that addresses the shortcomings of pixel-level segmentation.

The single threshold τ in (2) is often fixed by the designer. We find that a threshold in [1, 2] can detect most skin colors while keeping false detection low. This result is consistent with the result published in [14]. However, a fixed threshold is very limited in separating skin and nonskin regions in different images. Our experiments with a range of images showed that in most cases, we can manually select a threshold τ on the skin score $s(x)$ that can separate well skin regions from the background. What we need is an automatic mechanism to select the appropriate threshold.

The selection mechanism is based on the observation that the human skin has relatively smooth texture. Perceptually, a skin region of a person shows quite strong homogeneity. In contrast, a skin-colored image region that is mixture of an actual skin region and the nearby background tends to be non-uniform. We further observe that, in most natural images, the skin scores $s(x)$ of a skin region differ significantly from the skin scores of its nearby background regardless of the nearby background having skin colors or nonskin colors. Based on these observations, we suggest that an appropriate skin segmentation threshold τ can be chosen by raising it from some initial threshold step-by-step until the segmented skin-colored regions become homogenous.

3.1. Skin Segmentation Algorithm

The major steps of the proposed segmentation algorithm are shown in Fig. 1. From an input RGB image **C**, a skin score map $\mathbf{S}_\mathrm{C}$ is generated by finding the skin score for every image pixel in **C** (i.e. through table look-up). Next, an averaging filter of size 3×3 is applied on $\mathbf{S}_C$ in order to reduce noise. This filtering step is based on the fact that nearby pixels should have similar skin scores.

The skin score map is segmented using an initial threshold $\tau = 1$. The result of segmenting a skin score map is a binary mask of 0s (for nonskin) and 1s (for skin). For each sufficiently large and connected region in the binary mask B_{region}, we test if the region is homogenous. If the region is not homogenous, the threshold τ is increased by a factor of 1.1, and the skin score map of the region is segmented using the new threshold. This process continues repeatedly until the skin regions become homogenous. Once a homogenous region is found, its bounding box is identified. The region is grown to fill the bounding box by adding skin color pixels connected to it as long as the grown region remains homogenous. Finally, the region is added to the skin list. The algorithm shown in Fig. 1 can be implemented in a recursive fashion.

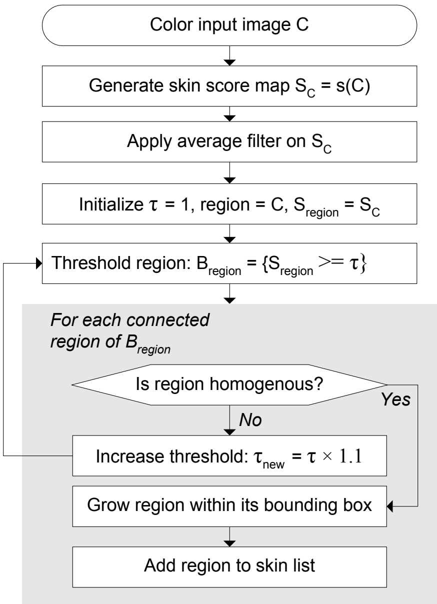


Figure 1: Adaptive skin segmentation algorithm.

3.2. Region Homogeneity Measures

We propose two measures to indicate the homogeneity of a region in both global and local senses. The first measure is the standard deviation σ of the pixel intensities in the region. It can be expected that a non-uniform region has higher σ value than a uniform region. Because all pixel intensities are used its computation, the standard deviation σ reflects the global variation in the region.

The second homogeneity measure is the number of edge pixels in the region. Because edge pixels are determined by local changes in image intensity, the second measure can indicate correctly that a region is non-uniform when the first measure fails. In our work, the Sobel edge detector is used because of its computational simplicity (see Table 1). The edge map **E** needs to be computed only once for the entire input image.

For each region, we calculate the number of edge pixels N_e *within* the region (edge pixels on the region boundary are not counted), the number of "1" pixels N_s in the region, and the maximum dimension N_d (width or height) of the region's bounding box. A region is considered homogenous if:

$$(\sigma \le 40)\ \textbf{AND}\ \{(\frac{N_e}{N_d} \le 1.5)\ \textbf{OR}\ (\frac{N_e}{N_s} \le 0.02)\}. \quad (10)$$

Table 1: Sobel edge detector.

Intensity image **I**

Sobel masks:

$$S_x = \frac{1}{8}\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad S_y = \frac{1}{8}\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

Gradient

Horizontal $\mathbf{G}_x = \mathbf{I} \otimes S_x$

Vertical $\mathbf{G}_y = \mathbf{I} \otimes S_y$

Magnitude $\mathbf{G} = \sqrt{\mathbf{G}_x^2 + \mathbf{G}_y^2}$

Edge map

$$\mathbf{E}(i,j) = \begin{cases} 1 & \text{if } \mathbf{G}(i,j) \ge \theta_e \\ 0 & \text{otherwise} \end{cases} \qquad \theta_e = 20$$

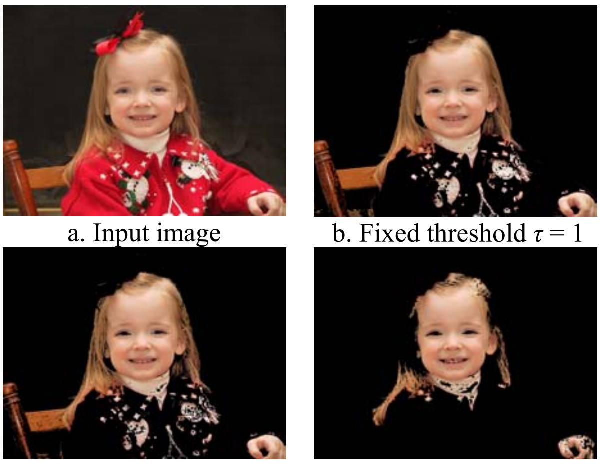
4. IMPLEMENTATION AND RESULTS

4.1. ECU Face Detection Database

In this work, we use the ECU face detection database, which was prepared as part of our face detection project at Edith Cowan University. The database consists of more than 3,000 color images; to the best of our knowledge, it is one of the largest online databases that support the many tasks involved in color-based face detection. All images are manually segmented for skin regions as well as for face regions. The database is made available online at:
http://www-soem.ecu.edu.au/~sphung/face_detection/.

4.2. Implementation and Results

Images 1 to 2500 in the ECU database were used for training. From these images, a training set of more than 120 million skin color samples and 150 million nonskin color samples was prepared. This training set, which consists of a wide range of skin colors (black, white, yellow, and brown), was used to construct the human skin color model described in Section 2. The various parameters of the proposed technique (Section 3) were found using these training images.



a. Input image b. Fixed threshold $\tau = 1$

c. Fixed threshold $\tau = 2$ d. Proposed technique

Figure 2: Sample results of skin segmentation.

Table 2: Skin segmentation comparison.

Method	False Detection	False Rejection
Fixed $\tau = 1$	13.6%	1.6%
Fixed $\tau = 2$	9.5%	2.3%
Proposed method	4.5%	4.0%

The proposed technique of using adaptive threshold τ was compared with the segmentation technique of using a fixed threshold τ. Images 2501 to 2600 of the ECU database were used as a test set. Two fixed thresholds $\tau = 1$ and $\tau = 2$ were tested (these values were chosen because they give reasonable classification of skin colors and nonskin colors). False detection and false rejection were computed by comparing pixel-wise the segmentation outputs with the ground-truth (i.e. the manually segmented) images. Test results in Table 2 show that the proposed technique indeed improves over the fixed-threshold approach.

Sample results of skin segmentation are shown in Fig. 2. The output images in Fig. 2b and Fig. 2c were obtained by applying fixed thresholds $\tau = 1$ and $\tau = 2$ on skin score map **S**, respectively. The output image in Fig. 2d is produced by the proposed skin segmentation technique. What we obtain by applying the new segmentation technique are homogenous image regions that have very high probability of being skin regions.

5. CONCLUSIONS AND FURTHER WORK

The proposed skin segmentation technique performs better than fixed threshold pixel-level skin color segmentation. This improvement is achieved by taking into account the texture characteristics of the human skin and using appropriate homogeneity measures for skin regions. We also report a robust human skin color model, which is built from a large and representative set of skin colors and nonskin colors. Our next focus is to develop a skin-color-based face detection algorithm that uses the skin segmentation technique described in this paper.

6. REFERENCES

[1] D. Chai and K. N. Ngan, "Face segmentation using skin color map in videophone applications," *IEEE Trans. CSVT*, vol. 9, no. 4, pp. 551-564, Jun. 1999.

[2] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE Trans. on Multimedia*, vol. 1, no. 3, pp. 264-277, Sep. 1999.

[3] H. Wu, Q. Chen, and M. Yachida, "Face detection from color images using a fuzzy pattern matching method," *IEEE Trans. PAMI*, vol. 21, no. 6, pp. 557-563, Jun. 1999.

[4] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *IEEE Trans. PAMI*, vol. 24, no. 5, May 2002.

[5] X. Zhu, J. Yang, and A. Waibel, "*Segmenting hands of arbitrary color*," IEEE Int. Conf. on Automatic Face and Gesture Recognition, Grenoble, France, 2000.

[6] M. Fleck, D. Forsyth, and C. Bregler, "*Finding naked people*," European Conference on Computer Vision, 1996.

[7] S. L. Phung, D. Chai, and A. Bouzerdoum, "*A universal and robust human skin color model using neural networks*," INNS-IEEE Int. Joint Conf. on Neural Networks, Washington, DC, USA, Jul. 2001.

[8] D. Brown, I. Craw, and J. Lewthwaite, "*A SOM-based approach to skin detection with application in real time systems*," British Machine Vision Conference, 2001.

[9] B. Menser and M. Wien, "*Segmentation and tracking of facial regions in color image sequences*," SPIE VCIP'2000, Perth, Australia, Jun.2000.

[10] G. Xu and T. Sugimoto, "*Rits Eye: A software-based system for realtime face detection and tracking using pan-tilt-zoom controllable camera*," Int. Conf. on Pattern Recog., 1998.

[11] M.-H. Yang and N. Ahuja, "*Gaussian mixture model for human skin color and its applications in image and video databases*," SPIE: Storage and Retrieval for Image and Video Databases, San Jose, Jan. 1999.

[12] M. J. Jones and J. M. Rehg, "*Statistical color models with application to skin detection*," Cambridge Research Laboratory, Technical Report CRL 98/11, December 1998.

[13] N. Bojic and K. K. Pang, "*Adaptive skin segmentation for head and shoulder video sequence*," SPIE VCIP'2000, Perth, Australia, Jun. 2000.

[14] H. Wang and S. F. Chang, "A highly efficient system for automatic face detection in MPEG video," *IEEE Trans. CSVT*, vol. 7, no. 4, pp. 615-628, Aug. 1997.

[15] J. Cai, A. Goshtasby, and C. Yu, "*Detecting human faces in color images*," Int. Workshop on Multimedia Database Management Systems, Aug. 1999.

[16] J. Yang and A. Waibel, "*A real-time face tracker*," IEEE Workshop on Applications of Computer Vision, Sarasota, Florida, USA, 1996.

[17] D. Chai and A. Bouzerdoum, "*A Bayesian approach to skin color classification in YCbCr color space*," IEEE TENCON'2000, Kualu Lumpur, Malaysia, Sep. 2000.