A BAYESIAN SKIN/NON-SKIN COLOR CLASSIFIER USING NON-PARAMETRIC DENSITY ESTIMATION

D. Chai, S. L. Phung and A. Bouzerdoum

Edith Cowan University
School of Engineering and Mathematics
100 Joondalup Drive, Joondalup WA 6027
Perth, AUSTRALIA

ABSTRACT

This paper addresses an image classification technique that uses Bayesian decision rule for minimum cost to determine if a color pixel has skin or non-skin color. Our proposed approach employs non-parametric estimation of class-conditional probability density functions of skin and non-skin color with feature vector that consists of all three components of the RGB color space. Experimental results demonstrate that the classifier can achieve good classification performance. Furthermore, its simplicity is an attractive feature for real-time applications. It is a useful tool for image processing tasks such as human face detection, facial expression and hand gesture analysis.

1. INTRODUCTION

Automatic detection of human faces in digital images has long been a difficult problem in computer vision that has found applications in areas such as face recognition for biometric identification and content-based facial image coding. Earlier face detection algorithms work on gray-scale images [1-3]. New face detection algorithms published in recent years, however, utilize color images, and they are designed to take advantage of the information stored in the chrominance components [4-7]. The first step of these face detection algorithms is often a skin color detection stage in which the input image is scanned pixel-wise for skin color pixels. Skin color image regions form initial estimates of facial regions, which are then refined using more sophisticated image segmentation and face/non-face classification techniques.

The result of the skin color detection stage depends heavily on the accuracy of the skin/non-skin color classifier. Classifiers based on linear decision boundaries [4, 5], multilayer perceptrons [8], self-organizing maps [9], Gaussian models [10, 11] and Gaussian mixtures [12] have been studied in the past. In this paper, we propose a simple but effective skin/non-skin color classifier that uses the Bayesian decision rule for minimum cost [13] and employs a non-parametric estimation of class-conditional probability densities for skin and non-skin color. The paper is organized as follows. Section 2 presents our proposed approach. The experimental results are provided in Section 3 and conclusion can be found in Section 4.

2. SKIN/NON-SKIN COLOR CLASSIFICATION

2.1 Problem Formulation

Skin/non-skin color classification can be formulated as a two-class classification problem. The two classes are skin color (denoted by ω_1) and non-skin color (denoted by ω_2). The feature vector used in making classification decision is $\mathbf{x} = (r\ g\ b)$ where r, g and b are the respective red, green and blue components of a color pixel.

In RGB color space, a pixel is defined by three components: red, green and blue. These components can be transformed into other color spaces such as HSV and YCbCr whereby their luminance and chrominance signals are decoupled. Many existing skin color detection algorithms use only partial information of the pixel. For example, only chrominance values stored in Cb-Cr are used in [8, 14] and H-S in [15], and two normalized RGB components in [16]. The rationale behind these approaches is based on the observation that different skin colors share a marked similarity in certain 2-D chrominance plane. However, we have shown in [17] that skin color detection can be improved if the full three components of a pixel are utilized for classification. This leads to our decision for using a feature vector that consists of all three red, green and blue components.

2.2 Bayesian Classifier

Suppose $P(\omega_i|\mathbf{x})$ is the probability of observing a pixel belonging to class ω_i given that its color is $\mathbf{x}$. Thus, $P(\omega_1|\mathbf{x})$ and $P(\omega_2|\mathbf{x})$ are the respective *a posteriori* probabilities for skin and non-skin color classes. Let C_{ij} be the cost of deciding a feature vector $\mathbf{x} \in \omega_i$ when in fact $\mathbf{x} \in \omega_j$. For $i, j = \{1, 2\}$, the costs of correct classifications are represented by C_{11} and C_{22}, while the costs of false detection and false rejection are represented by C_{12} and C_{21}, respectively.

Let $R_i(\mathbf{x})$ be the class-conditional cost of deciding a feature vector $\mathbf{x} \in \omega_i$. It can be calculated for $i = \{1, 2\}$ as

$$R_1(\mathbf{x}) = C_{11} \cdot P(\omega_1 \mid \mathbf{x}) + C_{12} \cdot P(\omega_2 \mid \mathbf{x}), \tag{1}$$

and

$$R_2(\mathbf{x}) = C_{21} \cdot P(\omega_1 \mid \mathbf{x}) + C_{22} \cdot P(\omega_2 \mid \mathbf{x}). \tag{2}$$

0-7803-7761-3/03/$17.00 ©2003 IEEE

An optimum classifier will assign $\mathbf{x}$ to the class that gives the minimum cost. In other words, the decision rule is

$$\begin{cases} \text{If } R_1(\mathbf{x}) < R_2(\mathbf{x}) \text{ then } \mathbf{x} \in \omega_1 \\ \text{else} \quad\quad\quad\quad\quad\quad \mathbf{x} \in \omega_2 \end{cases} . \tag{3}$$

Combining with (1) and (2), we can rewrite the decision rule as

$$\begin{cases} \text{If } \dfrac{P(\omega_1 \mid \mathbf{x})}{P(\omega_2 \mid \mathbf{x})} > \dfrac{C_{12} - C_{22}}{C_{21} - C_{11}} \text{ then } \mathbf{x} \in \omega_1 \\ \text{else} \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \mathbf{x} \in \omega_2 \end{cases} . \tag{4}$$

By applying the Bayesian formula

$$P(\omega_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_i) P(\omega_i)}{p(\mathbf{x})} \tag{5}$$

to (4), the decision rule can be expressed as

$$\begin{cases} \text{If } \dfrac{p(\mathbf{x} \mid \omega_1)}{p(\mathbf{x} \mid \omega_2)} > \tau \text{ then } \mathbf{x} \in \omega_1 \\ \text{else} \quad\quad\quad\quad\quad\quad \mathbf{x} \in \omega_2 \end{cases} , \tag{6}$$

where

$$\tau = \frac{C_{12} - C_{22}}{C_{21} - C_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)} . \tag{7}$$

In the above equations, $p(\mathbf{x}|\omega_i)$ is the class-conditional probability density function of skin color (when $i = 1$) and non-skin color (when $i = 2$); $P(\omega_i)$ is the *a priori* probability of class ω_i; and τ represents the adjustable threshold. Note that the costs of false classifications are manipulated by C_{12} and C_{21}, while the costs of correct classifications (*i.e.* C_{11} and C_{22}) are typically set to zero.

There are three approaches to the estimation of $p(\mathbf{x}|\omega_i)$, namely parametric, non-parametric and semi-parametric [18]. The parametric approach assumes a functional form of $p(\mathbf{x}|\omega_i)$, which can be customized by a set of parameters. Estimating $p(\mathbf{x}|\omega_i)$ becomes essentially finding a set of parameters that fits the functional form to the training data. The non-parametric approach, on the other hand, does not express $p(\mathbf{x}|\omega_i)$ in a parametric form but allows it to be determined entirely by the data. Semi-parametric approach, most notably neural networks, uses very general functional form that can have a variable number of adjustable parameters.

Both parametric (*e.g.* Gaussians [10] or mixtures of Gaussians [12]) and semi-parametric (*e.g.* multilayer perceptrons [8], self-organizing map [9]) approaches have been reported in the past for skin/non-skin classification. In this paper, we propose a non-parametric approach, which can be easily implemented in RGB color space with good results. In the next section, we describe the estimation of class-conditional density functions $p(\mathbf{x}|\omega_i)$ using the histogram method.

2.3 Non-parametric density estimation

The class-conditional probability density functions $p(\mathbf{x}|\omega_1)$ and $p(\mathbf{x}|\omega_2)$ are estimated as follows:

Step 1: A training set T that contains a large number of labeled skin and non-skin pixels is first prepared.

Step 2: To reduce memory requirement, the feature vector $\mathbf{x}$ in T is quantized by a quantization step Q to become $\mathbf{x}_q$ in T_q, *i.e.*

$$\mathbf{x} = (r, g, b) \in T \rightarrow \mathbf{x}_q = (r_q, g_q, b_q) \in T_q, \tag{8}$$

where $r_q = \lfloor r/Q \rfloor$, $g_q = \lfloor g/Q \rfloor$, and $b_q = \lfloor b/Q \rfloor$.

Step 3: Initialize two 3-D matrices of size $R \times R \times R$ for skin and non-skin pixel counts as:

$$H_1(r_q, g_q, b_q) = 0, \tag{9}$$

$$H_2(r_q, g_q, b_q) = 0, \tag{10}$$

where R is the range of each quantized color component.

Step 4: To construct histograms, update pixel counts for each pixel $\mathbf{x}_q \in T_q$:

$$H_1(r_q, g_q, b_q) = H_1(r_q, g_q, b_q) + 1 \quad \text{if } \mathbf{x} \in \omega_1, \tag{11}$$

$$H_2(r_q, g_q, b_q) = H_2(r_q, g_q, b_q) + 1 \quad \text{if } \mathbf{x} \in \omega_2. \tag{12}$$

Step 5: Normalize histograms to obtain the pdf estimates:

$$p(\mathbf{x} \mid \omega_1) \cong h_1(\mathbf{x}_q) = \frac{H_1(\mathbf{x}_q)}{N_1} \tag{13}$$

$$p(\mathbf{x} \mid \omega_2) \cong h_2(\mathbf{x}_q) = \frac{H_2(\mathbf{x}_q)}{N_2}, \tag{14}$$

where N_1 and N_2 are the total numbers of skin color and non-skin color pixels in the training set T, respectively.

These non-parametric density estimations are stored in a lookup table, which are then used by the classification decision rule shown in (6).

2.4 Smoothing of Density Estimates

The density estimates obtained using the algorithm presented in the previous section may have many irregularities. This is evident when the training set does not provide sufficient coverage of the feature space. The resulting classifier will produce pepper-like noise in detected skin color regions. This problem can be alleviated by applying a 3-D smoothing filter on the density estimates. A 3×3×3 averaging filter is used in our work to obtain regularized estimates $h_i^*(r_q, g_q, b_q)$:

$$h_i^*(r_q, g_q, b_q) = \frac{1}{27} \sum_{(r,g,b) \in N_3} h_i(r, g, b) . \tag{15}$$

Here N_3 is the 3×3×3 neighborhood (including the center) of (r_q, g_q, b_q) in the quantized color space.

2.5 Memory Requirement

The memory requirement of the algorithm is analyzed in this section. The RGB color space adopted here is 8 bits per component (*i.e.* $L = 256$ levels). If the quantization step is Q, then the number of levels for each quantized color component is $R = L/Q$. Hence, each histogram h_i has $(L/Q)^3$ entries. Assuming each entry takes 8 bytes (double floating-point), we need $16(L/Q)^3$ bytes to store both the skin and non-skin histograms.

In our work, we select the quantization step Q = 2, which is a trade-off between pdf estimation accuracy and memory requirement. The storage requirement for a set of full matrices of pdf estimates is 33MB. This requirement is attainable with today's desktop computer.

3. EXPERIMENTAL RESULTS

3.1 Training Data

The training set T for pdf estimation was prepared by manually labeling skin and non-skin color pixels in a set of 600 training images. The images are of various sizes and they contain subjects of different skin colors including black, yellow and white. This leads to the use of more than 145 million pixels to train the classifier. The RGB distribution of skin color pixels from the training set is shown in Figure 1.

The training set and all other data sets used in our work were taken from ECU Face Detection database, which is available online at `http://www-soem.ecu.edu.au/~sphung/face_detection.html` .

3.2 Testing Data

The following two different test sets were prepared in order to obtain data that has a representative mixture of skin and non-skin color. Test set 1 that consists of only human skin was used to evaluate the correct detection and false rejection rates. Test set 2 that contains no human skin was used to evaluate the correct rejection and false detection rates.

Test set 1 consists of more than 3.5 million skin color pixels that were manually extracted from color photographs of the 731 soccer players from 32 countries participating in the FIFA World Cup Korea/Japan 2002. This test set is a representative of a wide range of skin colors including black, white, yellow and brown. Test set 2, on the other hand, consists of more than 12 million pixels from non-skin image regions.

3.2 Results and Discussion

The size of the training and testing data sets are given in Table 1, while the classification rates for threshold $\tau = 1$ are shown in Table 2. This threshold corresponds to a symmetrical cost function and equal priors [13].

The classification performance at various threshold values was evaluated, and the results are plotted in Figures 2 and 3. Figure 2 shows a reduction in the amount of false detection as higher τ value was used. However, the increase in τ has resulted in more false rejection. Figure 3 shows that the total error rate reaches a minimum of 4.64% at threshold $\tau = 1.1$. In addition, the receiver operating characteristics (ROC) curve of the classifier is shown in Figure 4.

The classifier is very good at detecting skin regions in images, and therefore it is capable of forming good initial estimates for other facial image processing tasks such as face detection. The drawback of the classifier is that it will pick up regions in the background that have skin color appearance, ie false detection. However, these regions can be eliminated by post-processing steps such as those discussed in [4]. These steps depend on the type of applications that use the results of skin color detecion.

Our proposed algorithm was implemented in C++, which is made available at the previously mentioned URL. Using a Pentium III 600MHz PC with 512MB RAM, the program took between 200 to 300 ms to process a 640×480 bitmap file (including file input/output)

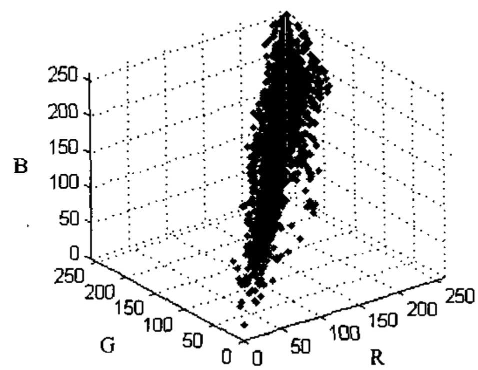

Figure 1: Skin color distribution in RGB color space.

Table 1: Training and testing sets.

Data set	Size (pixels)
Training set	145,289,257
Test set 1 (skin)	3,586,535
Test set 2 (non-skin)	12,165,120

Table 2: Classification result for $\tau = 1$.

Category	Rate (%)
False detection	8.55
False rejection	0.68

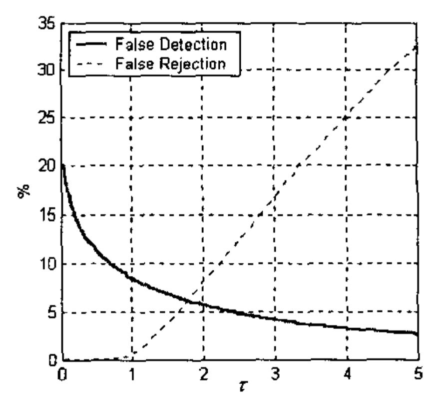

Figure 2: Misclassification at different threshold values.

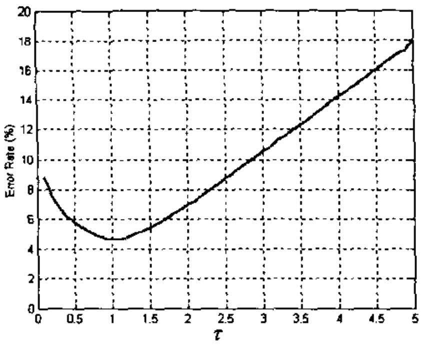

Figure 3: Total error rates at different threshold values.

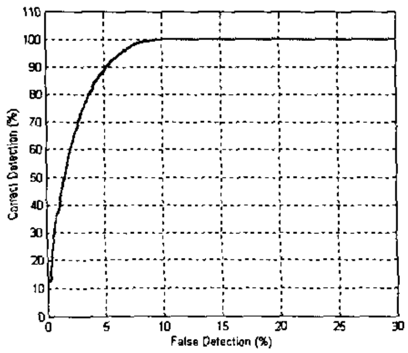

Figure 4: Receiver operating characteristics (ROC) curve of the skin/non-skin color classifier.

4. CONCLUSION

In this paper, we have presented a non-parametric approach for estimating the class-conditional probability density functions of skin and non-skin colors. These are used in the Bayesian decision rule for minimum cost to classify pixels in RGB color space as skin or non-skin color. Our analysis has shown that the non-parametric approach is viable in term of memory requirement to handle the full RGB color space. The major advantages of the proposed classifier are its simplicity (which is attractive to real-time applications) and accuracy (error rate of 4.64%).

5. ACKNOWLEDGMENT

The authors wish to thank Mr Fok Hing Chi Tivive for taking part in preparing the ECU face detection database.

6. REFERENCES

[1] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. on PAMI*, vol. 20, no. 1, pp. 39-51, Jan. 1998.

[2] H. A. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Trans. on PAMI*, vol. 20, no. 1, pp. 23-38, Jan. 1998.

[3] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: an application to face detection," *Computer Vision and Pattern Recognition 1997*, Puerto Rico, Jun. 1997.

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

[5] 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, Sept. 1999.

[6] 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.

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

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

[9] 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.

[10] B. Menser and M. Wien, "Segmentation and tracking of facial regions in color image sequences," *Visual Communication and Image Processing 2000*, Perth, Australia, Jun. 2000.

[11] G. Xu and T. Sugimoto, "Rits Eye: A software-based system for realtime face detection and tracking using pan-tilt-zoom controllable camera," *14th International Conference on Pattern Recognition*, 1998.

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

[13] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: John Wiley & Sons, Inc., 2001.

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

[15] N. Bojic and K. K. Pang, "Adaptive skin segmentation for head and shoulder video sequence," *SPIE Visual Communication and Image Processing 2000*, Perth, Australia, Jun. 2000.

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

[17] S. L. Phung, A. Bouzerdoum, and D. Chai, "A novel skin color model in YCbCr color space and its application to human face detection," *IEEE International Conference on Image Processing*, Rochester, USA, Sep. 2002.

[18] C. M. Bishop, *Neural networks for pattern recognition*: Oxford: Clarendon Press, 1996.