A NOVEL SKIN COLOR MODEL IN YCBCR COLOR SPACE AND ITS APPLICATION TO HUMAN FACE DETECTION

Son Lam Phung, Abdesselam Bouzerdoum, and Douglas Chai

Visual Information Processing Research Group
Edith Cowan University, Western Australia
{s.phung, a.bouzerdoum, d.chai}@ecu.edu.au

ABSTRACT

This paper presents a new human skin color model in YCbCr color space and its application to human face detection. Skin colors are modeled by a set of three Gaussian clusters, each of which is characterized by a centroid and a covariance matrix. The centroids and covariance matrices are estimated from large set of training samples after a *k*-means clustering process. Pixels in a color input image can be classified into skin or non-skin based on the Mahalanobis distances to the three clusters. Efficient post-processing techniques namely noise removal, shape criteria, elliptic curve fitting and face/non-face classification are proposed in order to further refine skin segmentation results for the purpose of face detection.

1. INTRODUCTION

Face detection, which aims to detect the presence and subsequently the position of human faces in an image, is often the first important step in automated facial image analysis. Results of face detection enable tasks such as face recognition and facial expression analysis to be performed on focused image regions. The major challenge in face detection is to cope with a wide range of variations in the human face pattern caused by factors such as different lighting, face orientation, face size, facial expression and people ethnicity. The presence of complex background or extra facial features such as glasses, beard, and moustache also adds to the complexity of the problem.

Several face detection algorithms including those that use neural networks [4, 10], support vector machines [1, 8], mixtures of factor analyzers [13] and wavelets [5] have been proposed. Interested readers are referred to two comprehensive reviews on face detection published recently by Yang *et al*. [14], and by Hjelmas and Low [6]. In recent years, face detection in color images through the detection of skin colored regions has gained special attention. This approach not only enables fast localization of potential facial regions but also proves to be highly robust to geometric variations in the face patterns. However, the success of this approach depends heavily on the accuracy and robustness of the human skin color model.

Many existing human skin color models operate only on chrominance planes (e.g. Cb-Cr [2, 3] or H-S [11]). They work based on the observation that chrominance of different skin colors (e.g. white, black, yellow and brown) shares a marked similarity, which can be used to distinguish skin colors from non-skin colors. These models work reasonably well with medium level of lighting.

However, as shown in Fig. 1, for low and high luminance Y the decision boundary between skin and non-skin in chrominance plane is reduced. Therefore, significant false alarms will incur if only one chrominance decision boundary is used for all levels of luminance Y. To address this issue, Hsu *et al*. [7] suggested a non-linear transform to the YCbCr color space that makes the skin color cluster in the chrominance plane luma-independent. Garcia and Tziritas [5] used a set of bounding planes to approximate the skin clusters in YCbCr and HSV spaces.

In this paper, we present a skin color-based face detection algorithm that employs a human skin color model, which takes into account the luminance Y in classifying skin and non-skin pixels. In our approach, the distribution of human skin colors in YCbCr space is modeled with three Gaussian clusters that correspond approximately to three levels of luminance: low, medium and high. The rest of the paper is organized as follows. The construction of the human skin color model in YCbCr color space is described in Section 2. The various stages of our face detection algorithm are presented in Section 3. Experimental results and discussions are provided in Section 4. Conclusion is given in Section 5.

2. HUMAN SKIN COLOR MODEL

Skin colors form a concentrated cluster in YCbCr space and this cluster has a highly irregular boundary (Fig. 1). We will approximate this cluster using three 3-D Gaussians.

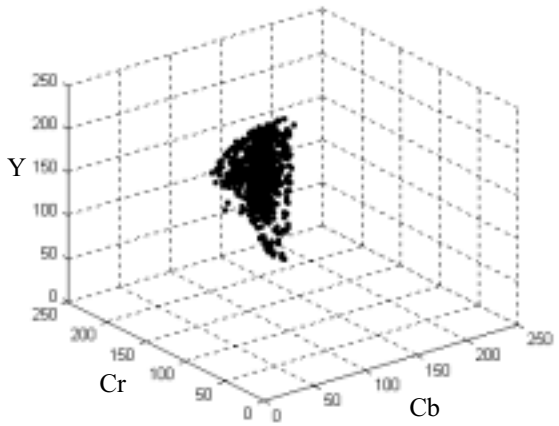

Figure 1. Skin colors in YCbCr color space. The decision boundary on Cb-Cr plane is large for medium Y and is reduced for low or high Y.

2.1 Training Phase

Let **E** be an ensemble of skin color samples: $\mathbf{E} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ where $\mathbf{x}_i = [\mathrm{Y}_i, \mathrm{Cb}_i, \mathrm{Cr}_i]^{\mathrm{T}}$.

Step 1: Assign the samples to k ($k = 3$ in this case) initial clusters: $\mathbf{E} = \{\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_k\}$:

$$\mathbf{E}_i = \{\mathbf{x}_j \in \mathbf{E} \mid \mathrm{Y}_{min} + (i\text{-}1)\Delta_Y \le \mathrm{Y}_j < \mathrm{Y}_{min} + (i\text{-}1)\Delta_Y\}, \quad (1)$$

where $i = \overline{1,k}$, $[\mathrm{Y}_{min}, \mathrm{Y}_{max}]$ is the range of Y in YCbCr color specification and $\Delta_Y = \dfrac{\mathrm{Y}_{max} - \mathrm{Y}_{min}}{k}$ is the interval.

Step 2: Compute the centroid and covariance matrix for each cluster ($i = \overline{1,k}$):

$$\boldsymbol{\mu}_i = \frac{1}{N_i} \sum_{\mathbf{x}_j \in \mathbf{E}_i} \mathbf{x}_j , \quad (2)$$

$$\mathbf{c}_i = \frac{1}{N_i - 1} \sum_{\mathbf{x}_j \in \mathbf{E}_i} (\mathbf{x}_j - \boldsymbol{\mu}_i)((\mathbf{x}_j - \boldsymbol{\mu}_i)^{\mathrm{T}} . \quad (3)$$

N_i is the size of cluster E_i.

Step 3: For each sample $\mathbf{x}_j$ ($j = \overline{1,N}$) compute the k Mahalanobis distances to the k clusters ($i = \overline{1,k}$):

$$M(\mathbf{x}_j, \boldsymbol{\mu}_i, \mathbf{c}_i) = (\mathbf{x}_j - \boldsymbol{\mu}_i)^{\mathrm{T}} \mathbf{c}_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i) . \quad (4)$$

Reassign the sample $\mathbf{x}_j$ to the cluster $\mathbf{E}_i$, which corresponds to the smallest distance:

$$x_j \rightarrow \mathbf{E}_i \text{ where } i = \arg \min_i M(\mathbf{x}_j, \boldsymbol{\mu}_i, \mathbf{c}_i) . \quad (5)$$

Step 4: Repeat steps 2-3 until a maximum allowable number of iterations is reached or there is no further change in the clusters. Store the cluster centroids and covariance matrices.

2.2. Classification Phase

Let $\mathbf{x} = [Y, Cb, Cr]^{\mathrm{T}}$ be a pixel in a color input image. The classification rule is as follows: **x** is skin color if it satisfies the following two tests:

(i) Its projection on the Cb-Cr plane is inside a predetermined rectangle:

$$Cb \in R_{Cb} \text{ and } Cr \in R_{Cr}, \quad (6)$$

where $R_{Cb} = [75, 135]$ and $R_{Cr} = [130, 180]$. The ranges R_{Cb} and R_{Cr}, which are found experimentally, are used to eliminate quickly non-skin colors.

(ii) The minimum Mahalanobis distance from **x** to the clusters is below a threshold:

$$\min_i \{M(\mathbf{x}, \boldsymbol{\mu}_i, \mathbf{c}_i)\} < \theta_d . \quad (7)$$

3. FACE DETECTION

Let $\mathbf{X} = [x_{ij}]_{W \times H}$ be the input color image. The output of skin color detection is a binary map $\mathbf{B} = [b_{ij}]_{W \times H}$:

$$b_{ij} = \begin{cases} 1 \text{ if } x_{ij} \text{ is skin color} \\ 0 \text{ otherwise} \end{cases} . \quad (8)$$

Each set of connected 1's in **B** is called an *object*, which is a potential face candidate. However, even with highly accurate skin color detection, two sources of errors remain to be addressed: (i) background pixels (i.e. non-face such as hands or scenery) can have skin colors and this leads to false alarms; (ii) some facial regions (e.g. eyes and mouths) do not have skin colors and this leads to false rejections. In the following discussion, let A_i, W_i, H_i and P_i denote the area, width, height of the bounding box, and perimeter of object i in **B** respectively.

3.1. Noise Removal

The binary map **B** may contain noise, which tends to scatter and have relatively small area. This suggests the following noise removal technique: object i is removed if

$$\frac{A_i}{W \times H} < \theta_{a1} \quad \text{and} \quad \frac{A_i}{A_{\max}} < \theta_{a2} . \quad (9)$$

$A_{\max}$ is the area of the largest object in **B,** $A = W \times H$ is the area of the input image, and θ_{a1} and θ_{a2} are two thresholds. Another useful technique is to remove, in each object, all cross sections (horizontal or vertical) that are relatively short compared to the longest cross sections of the object. Used in conjunction with the first technique, this will remove "spikes" connected to a facial region.

3.2. Shape Criteria

The human face resembles an oval shape even under different views and has a relatively constrained aspect ratio. For each object in the binary map, two shape measures are defined:

$$\text{Circularity:} \quad C_i = A_i / P_i^2, \tag{10}$$

$$\text{Aspect ratio:} \quad R_i = H_i / W_i. \tag{11}$$

The object is kept only if:

$$C_i > \theta_c \text{ and } R_i \in R_R, \tag{12}$$

where θ_c (= 0.05) is a threshold and R_R (= [0.8, 2.2]) is a range of valid aspect ratios for face.

3.3. Refining face region with elliptic curve fitting

Because the human face has an ellipse-like shape, face candidates can be refined by finding the best-fit ellipse for each object in the binary map. This can be done, for example, using Hough transform. We adopt the following technique for finding the best-fit ellipse.

Let $\mathbf{B}_i = \{(x_j, y_j)\}$, $j = \overline{1, N}$ be the set of pixel coordinates for object i in **B**. We first compute means μ_x and μ_y, variances δ_x^2 and δ_x^2, and covariance δ_{xy} for $\mathbf{B}_i$. The best-fit ellipse has the center at (μ_x, μ_y) and the major axis at an angle θ (clockwise) from the x-axis:

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\delta_{xy}}{\delta_x^2 - \delta_y^2}\right). \tag{13}$$

The ellipse major and minor radii are:

$$a = (M+N)^{\frac{1}{2}} \text{ and } b = (M-N)^{\frac{1}{2}}, \tag{14}$$

where

$$M = 4(\delta_x^2 + \delta_y^2) \text{ and } N = -\frac{4(\delta_x^2 - \delta_y^2)}{\cos 2\theta} \tag{15}$$

The ellipse interior is logical AND'ed with $\mathbf{B}_i$ to form the refined region for object i.

3.4. Face/Non-face Classification

For each object i in **B**, its bounding rectangle is identified and an intensity image that corresponds to this rectangle is extracted. In this final stage of face detection, we aim to verify if the intensity image contains face or non-face. The following distribution-based classification algorithm is used. A more detailed explanation of this algorithm can be found in [9].

Step 1: The intensity image is resized to a matrix of size 15×15, histogram-equalized, and then read column-wise to form a column vector $\mathbf{z}$ in $\mathbf{R}^{225}$ space.

Step 2: The vector $\mathbf{z}$ is projected onto face subspace spanned by a set $\mathbf{V}$ of d eigenfaces (d=20), and a feature vector $\mathbf{w}$ is extracted:

$$\mathbf{w} = \mathbf{V}^{\mathrm{T}}(\mathbf{z} - \mathbf{z}_m). \tag{16}$$

The set of eigenfaces is found by performing principal component analysis (PCA) [12] on a large set of face vectors $\mathbf{z}$; $\mathbf{z}_m$ is the average face vector. This step is essentially a dimension-reduction step.

Step 3: Each feature vector $\mathbf{w}$ in $\mathbf{R}^d$ space is classified as face or non-face by a distribution-based method, which models the class-conditional densities of face $p(\mathbf{w}|face)$ and non-face $p(\mathbf{w}|non\text{-}face)$ in $\mathbf{R}^d$ as two mixtures of Gaussians:

$$p(\mathbf{w} \mid face) = \sum_{i=1}^{G_f} \pi_{f_i} g(\mathbf{w}; \boldsymbol{\theta}_{f_i}, \mathbf{C}_{f_i}) \text{ and} \tag{17}$$

$$p(\mathbf{w} \mid nonface) = \sum_{i=1}^{G_{nf}} \pi_{nf_i} g(\mathbf{w}; \boldsymbol{\theta}_{nf_i}, \mathbf{C}_{nf_i}). \tag{18}$$

The subscripts f and nf denote face and non-face; π_i is a mixing factor, G is the number of Gaussian components in each mixture. Each Gaussian component is governed by a Gaussian function:

$$g(\mathbf{w}; \boldsymbol{\mu}, \mathbf{C}) = (2\pi)^{-\frac{k}{2}} |\mathbf{C}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{w}-\boldsymbol{\mu})\} \tag{19}$$

where $\boldsymbol{\mu}$ and $\mathbf{C}$ are the mean and covariance matrix.

The parameters of the Gaussian mixtures (i.e. mixing factors, means and covariance matrices) are found by applying the Expectation/Maximization (EM) algorithm on a training set of face and non-face feature vectors.

Once the class-condition densities are found, a feature vector $\mathbf{w}$ is classified as face if it satisfies the two following conditions:

$$\text{Condition 1: } \frac{p(\mathbf{w} \mid face)}{p(\mathbf{w} \mid nonface)} > \tau_p, \tag{20}$$

and

$$\text{Condition 2: } p(\mathbf{w} \mid face) > \tau_{face}. \tag{21}$$

τ_p and τ_{face} are two fixed thresholds whose values are determined experimentally.

4. RESULTS AND DISCUSSION

This section presents the testing results of two important aspects of our approach: the human skin color model and the face detector. The parameters of the skin color model were computed from a training set of near 4000 skin color samples that were manually selected from a set of color images. The three clusters found were centered at (75, 118, 143), (138, 110, 150), and (184, 111, 147) in YCbCr space.

The test set for the skin color model of near 300,000 samples was prepared in a similar way as for the training set: 100 test color images were used. The Receiver Operating Characteristics (ROC) of the model is shown in Fig. 2. Note that the threshold for the modified Mahalanobis distance should be chosen with an emphasis on reducing false rejections because false alarms of skin colors can be removed by later stages of face detection.

The test set for the proposed face detection algorithm consisted of images obtained from various sources. These test images contained people of different skin tones (e.g. white, black, yellow and brown) and had wide variations in illumination (e.g. indoor, outdoor). The testing results are shown in Table 1.

Table 1: Face detection results.

Face detection test set*	200 color images image size: 352× 288
Total faces	234 faces
Correct detection	204 faces (87.6%)
False rejections	30 faces (12.4%)
False alarms	22 faces

*The test set and related materials are available at: http://www-soem.ecu.edu.au/~sphung

5. CONCLUSION

In this paper, a new and improved human skin color model, which incorporates chrominance as well as luminance in classifying skin and non-skin colors, was proposed. The distribution of skin colors in YCbCr color space is approximated using three Gaussian clusters, which correspond relatively to three levels of luminance. Efficient face candidate verification techniques for face detection task were also discussed.

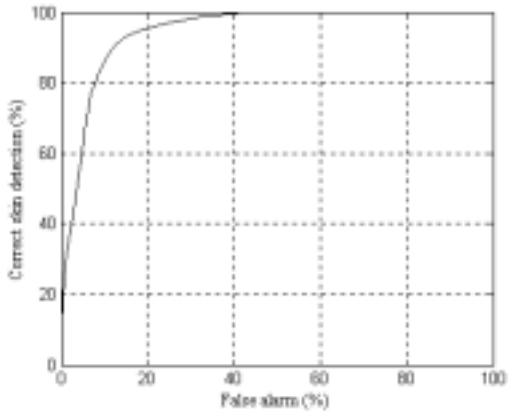

Figure 2. ROC for skin color detection for different threshold θ_d.

6. REFERENCES

[1] N. Bassiou, C. Kotropoulos, T. Kosmidis, and I. Pitas, "Frontal face detection using support vector machines and backpropagation neural networks," *IEEE ICIP '2001*, Thessaloniki, Greece, Oct. 2001.

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

[3] 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.

[4] R. Feraud, O. J. Bernier, J.-E. Viallet, and M. Collobert, "A fast and accurate face detector based on neural networks," *IEEE Trans. PAMI*, vol. 23, no. 1, pp. 42-53, Jan. 2001.

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

[6] E. Hjelmas and B. K. Low, "Face detection: a survey," *Computer Vision and Image Understanding*, vol. 83, no. 3, pp. 236-274, 2001.

[7] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *IEEE ICIP'2001*, Thessaloniki, Greece, Oct. 2001.

[8] 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.

[9] S. L. Phung, D. Chai, and A. Bouzerdoum, "A distribution-based face/non-face classification technique," to appear in *Australian Journal of Intelligent Information Systems*.

[10] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Trans. PAMI*, vol. 20, no. 1, pp. 23-38, Jan. 1998.

[11] K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking," *Signal Processing: Image Communication*, vol. 12, pp. 263-281, Jun. 1998.

[12] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-96, 1991.

[13] M.-H. Yang, N. Ahuja, and D. Kriegman, "Face detection using a mixture of factor analyzers," *IEEE ICIP '1999*, Kobe, Japan, 1999.

[14] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Trans. PAMI*, vol. 24, no. 1, pp. 34-58, Jan. 2002.