A Probabilistic Model for the Human Skin Color

T.S. Caetano and D.A.C. Barone

Universidade Federal do Rio Grande do Sul – Instituto de Informática
Av. Bento Golçalves, 9500, bloco IV – Porto Alegre – RS – Brazil

E-mail: {caetano,barone}@inf.ufrgs.br

Abstract

We present a multivariate statistical model to represent the human skin color. In our approach, there are no limitations regarding if the person is white or black, once the model is able to learn automatically the ethny of the people involved. We propose to model the skin color in the chromatic subspace, which is by default normalized with respect to illumination. First, skin samples from both white and black people are collected. These samples are then used to estimate a parametric statistical model, which consists of a mixture of gaussian probability density functions (pdf's). Estimation is performed by a learning process based on the expectation-maximization (EM) algorithm. In the following, experiments are carried out and receiving operating characteristics (ROC curves) are obtained to analyse the performance of the estimated model and compare it to outcomes of models that use a single gaussian density. Finally, conclusions are presented and future work is outlined.

1. Introduction

Automatic human face recognition is an emerging application in video surveillance and authentication systems. Such systems should be robust and reliable when matching a given face to an existing database to return status concerning the person's permissions. Although face recognition is a fundamental part in a fully automated facial analysis system, the first important step in recognizing faces in video sequences is to find where they are. This is called face *detection* or, specifically, face tracking in the case of video sequences. Most of the existing face detection methods use gray intensity values to detect faces [1-6]. However, it is a well-known fact that the majority of images acquired today are colored and the skin color feature should be an important source of information for discriminating faces from the background. Only a few face detection methods use color as a feature [7-9]. In these systems, color is modeled as a gaussian function in some color space where intensity plays no role and the whole information is provided by hue and saturation. The reason for using such color spaces is that human skin color varies more in intensity than in chromatic feature itself between different ethnic groups [10]. This fact gives rise to a relatively well-clustered cloud of skin samples even for training sets containing people from different races. As a result, there have been several attempts to model the sample distribution with a single gaussian probability density function [7-9].

In this paper, we propose to model the distribution obtained from white and black people skin samples with *two* gaussians, rather than with a single one. In spite of the fact that samples are well-clustered in the chromatic color subspace and one gaussian may be enough, we present results which indicate that a gaussian mixture modeling, where two different but nearly positioned clusters are used instead of one, can offer better classification performance in the case where high correct detection rate is needed.

2. Color space selection

The color model used is based on normalized RGB, or chromaticity space. The chromaticities are defined as

$$r = \frac{R}{R+G+B}$$
$$g = \frac{G}{R+G+B} \qquad \text{(1a-b-c)}$$
$$b = \frac{B}{R+G+B}$$

where it can be seen that they sum to one. Indeed, two chromaticities contain the whole information about the chromatic value.

The reason for using this color space is due to evidences that the human skin color is more compactly represented in it than it is in other color spaces, such as RGB, HSI, SCT and YQQ [11]. Here we use the chromaticities **r** and **b** to describe the color. The third chromaticity, in this case **g**, contains no information, once the other two are already specified.

In the next section, we describe the sample set used to estimate the statistical model of the chromaticities.

3. The sample set

In order to cover a wide range of skin chromatic characteristics, a set of 140.000 skin pixels has been sampled from 40 people, 20 black and 20 caucasians with white skin. Exactly half of the sample set came from each ethnic group. The images were taken both from random sites on the Internet and from the Web-available Stirling University face database. Care has been taken to avoid sampling areas of high luminance variations and highlights. Fig. 1 shows the distribution of the skin samples for the whole set, including black and white people.

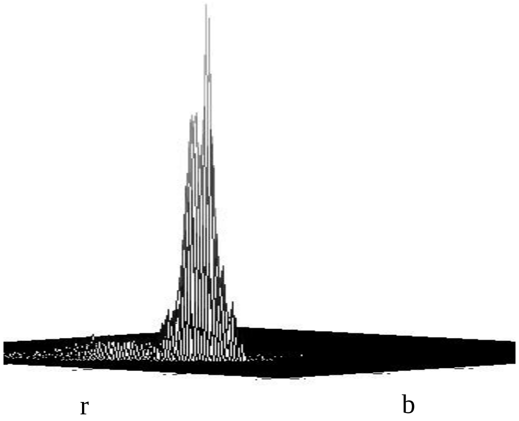

Figure 1. The distribution of the entire sample set.

The visual inspection of Fig. 1 suggests that a single gaussian may be a good model to fit the data. In the following section we describe the estimation of the parameters for a single gaussian modeling.

4. Multivariate gaussian model

It has been suggested in the literature that the skin color can be modeled by a single gaussian pdf, even for samples coming from different ethnic groups [7-9]. Considering independence between the samples, the model can be obtained from the maximum-likelihood criterion, which has analytical solution for a single gaussian pdf and results in the following estimates [12]

$$\hat{m} = \frac{1}{n}\sum_{k=1}^{n} x_k$$
$$\hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n} (x_k - \hat{m})(x_k - \hat{m})^T, \tag{2a-b}$$

where n is the number of samples, x_k is the vector representing the k*th* sample and $\hat{m}$ and $\hat{\Sigma}$ are, respectively, the mean vector and the covariance matrix of the gaussian pdf estimated. The resulting probability density that models the data is then

$$p(x) = \frac{1}{(2p)^{d/2}\left|\hat{\Sigma}\right|^{1/2}} \times \exp\left(-\frac{1}{2}(x-\hat{m})\hat{\Sigma}^{-1}(x-\hat{m})^T\right) \tag{3}$$

In the next section we describe our alternative model, which uses two gaussian clusters rather than one to fit the data set.

5. The mixture of gaussians model

The contribution of this work is to propose a new way to model the sample set presented in section 3 based on a mixture of gaussian pdf's. The motivation for this is that, in spite of the fact that the whole data is well-clustered and one gaussian can provide good performance, there are in fact two different ethnies and intuitively it is reasonable to say that an optimal result can be obtained if we address one gaussian to each ethny. We assume that the entire data set can be modeled by a probability density function given by

$$p(x \mid \Theta) = \sum_{l=1}^{2} p(x \mid l, \Theta_l) P(l) \tag{4}$$

where *l* denotes each individual gaussian, $p(x \mid l, \Theta_l)$ is the specific density function for gaussian *l* and *P(l)* is the scaling factor, which states how much of each gaussian is present in the mixture. The goal here is to estimate the parameter vector Θ applying the maximum-likelihood method, which consist in estimating the vector of parameters $\hat{\Theta}$ that maximizes $p(x \mid \Theta)$. However, there is no analytical solution for the maximization of (4) and the use of a numeric procedure is necessary.

The standard algorithm to learn the maximum-likelihood solution in this case, where more than one gaussian is present, is the iterative expectation-maximization (EM) procedure [13]. The EM algorithm, when applied to the problem of fitting a mixture of gaussians to multivariate data, results in the following update rules for the parameters:

$$\hat{P}_l = \frac{1}{N}\sum_{i=1}^{N} p(l \mid x_i, \hat{m}_l^{old}, \hat{\Sigma}_l^{old}) \tag{5a}$$

$$\hat{m}_l = \frac{\sum_{i=1}^{N} x_i\, p(l \mid x_i, \hat{m}_l^{old}, \hat{\Sigma}_l^{old})}{\sum_{i=1}^{N} p(l \mid x_i, \hat{m}_l^{old}, \hat{\Sigma}_l^{old})} \tag{5b}$$

$$\hat{\Sigma}_l = \frac{\sum_{i=1}^{N} p(l \mid x_i, \hat{m}_l^{old}, \hat{\Sigma}_l^{old})(x_i - \hat{m}_l)(x_i - \hat{m}_l)^T}{\sum_{i=1}^{N} p(l \mid x_i, \hat{m}_l^{old}, \hat{\Sigma}_l^{old})} \tag{5c}$$

for $l \in \{1,2\}$, where N, $\hat{m}_l^{old}$, $\hat{\Sigma}_l^{old}$, $\hat{m}_l$, $\hat{\Sigma}_l$ and $\hat{P}_l$ are, respectively, the sample size, the old mean and covariance, the mean vector, the covariance matrix and the mixture factor estimated for class l ; $p(l \mid x_i, \hat{m}_l^{old}, \hat{\Sigma}_l^{old})$ is the probability that data x_i belongs to class *l*. The above equations are used iteratively to update the parameters $\hat{P}_l$, $\hat{m}_l$ and $\hat{\Sigma}_l$ for each class *l*, until no noticeable change in the parameters is detected.

Next section describes the experiments that were accomplished and a comparison between the single-gaussian and the mixture of gaussians models.

6. Experiments and results

We have applied both models to a set of 100 images from the Stirling database and from the Internet, which contain people from white and black skin in different levels of background complexity and luminance variation. The procedure of applying each model to an image can be summarized as follows:

i) For each image pixel, obtain its chromaticity vector v=[r b] from eq. (1).

ii) Apply the vector as input in the model and measure the output L.

As a result, the measure L will be an indication of the likelihood of the pixel belonging to the skin. In fact, the output of a color image after applying the model is a gray-level image, where the brightness of each pixel indicates the probability of being a skin one. Fig. 2 shows some images from the test set and the results of the application of both models.

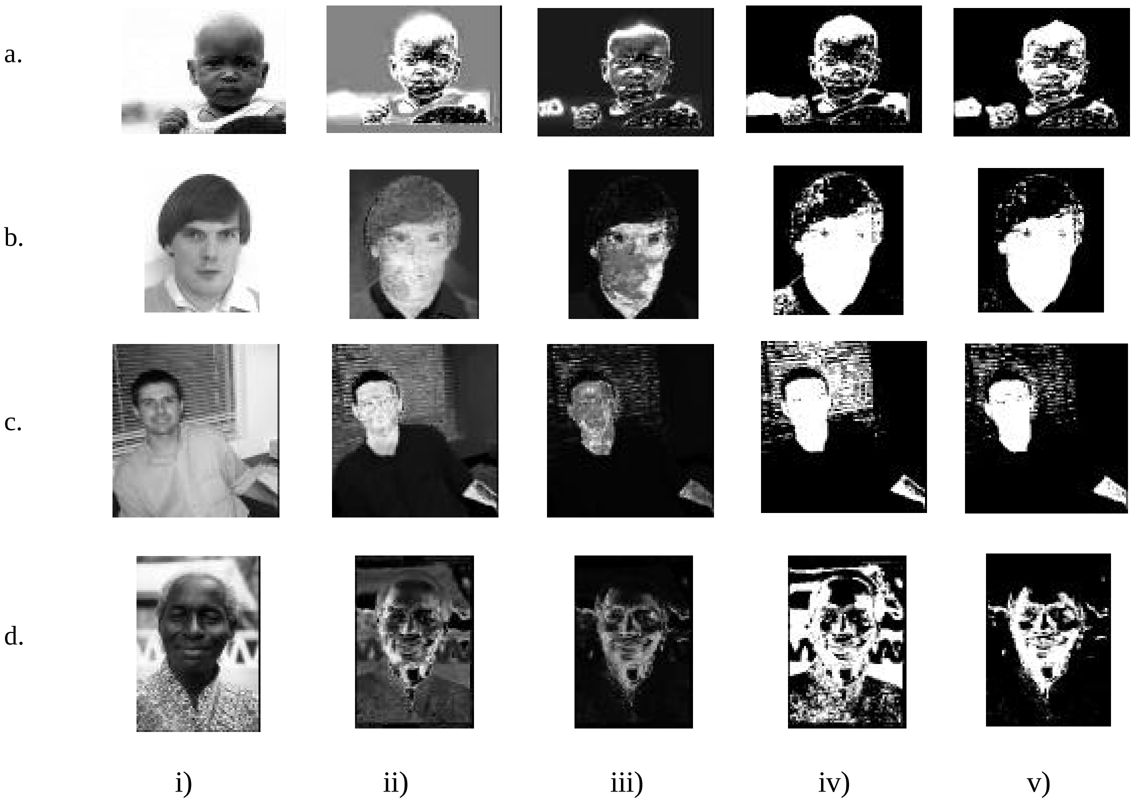

Figure 2. Performances obtained for the two models: i) Source image (original colored) ; ii) Result of single-gaussian model application; iii) Result of mixture of gaussians model application; iv) Image (ii) segmented at thresholds shown in figure 3; v) Image (iii) segmented at thresholds shown in figure 3.

In Fig. 2, it is possible to see the different effects obtained from applying one or other model to fit the experimental data set. While high correct skin detection is possible with the single-gaussian model, the cost incurred is a high false detection rate, as we can see in the segmented images in fig. 2 (iv and v):

- fig. 2.a) In the baby image, there are excessive false detections in the left bottom corner for the single-gaussian model (iv).

- fig. 2.b) In this image, a larger amount of pixels from the man's right shoulder and hair is wrongly detected by the single-gaussian model (iv) than by our model (v).

- fig. 2.c) Here, several pixels are detected as skin by the single-gaussian model (iv), despite the fact that they really belong to the window. Our model (v) gives less false detections for the same amount of correct detections.

- fig. 2.d) In the black man image, the performance of the single-gaussian model (iv) is very poor compared to ours (v).

From these observations, we can conclude that the mixture of gaussians model proposed is more selective.

For the same correct detection ratio, our model provides less false detected pixels. In order to perform a quantitative analysis to measure the differences between the results obtained from the two models, we can build receiving operating characteristics (ROC curves), selecting manually which areas in the image in fact contain skin and then varying a threshold to segment the gray-level image. The points in the curve can be determined by measuring the proportion of correctly and wrongly detected pixels with respect to the total number of skin pixels. The ROC curves in Fig. 3 were generated from the four examples above and illustrate the fact that the proposed mixture model has better performance when high detection is a need.

Here, correct detection ratio (cdr) and false detection ratio (fdr) are the relevant variables, which are defined as

- cdr = number of pixels detected as skin which **are really** skin pixels / total number of real skin pixels.
- fdr = number of pixels detected as skin which **are not** real skin pixels / total number of real skin pixels.

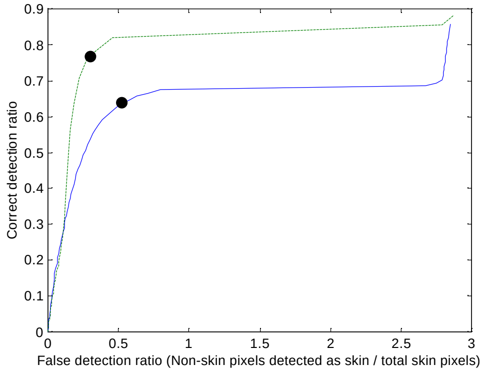

Figure 3a- ROC curve for the baby image (2a). Dashed line - proposed model; solid line - single-gaussian model. Black dots point where threshold values for fig. 2a (iv, v) were selected.

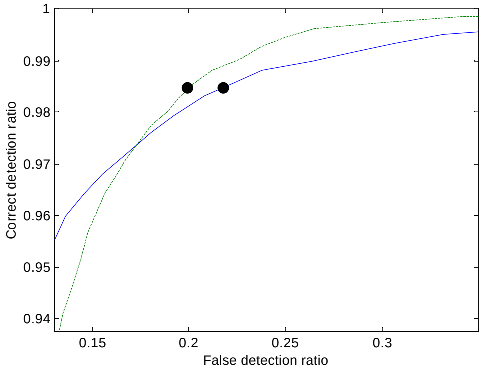

Fig. 3b- ROC curve for the adult image (2b). Dashed line - proposed model; solid line - single-gaussian model. Black dots point where threshold values for fig. 2b (iv, v) were selected.

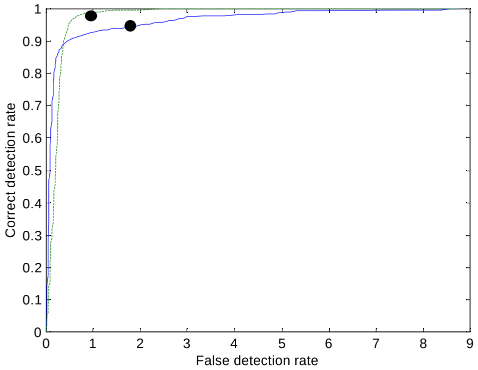

Figure 3c- ROC curve for the sat man (2c). Dashed line - proposed model;solid line - single-gaussian model. Black dots point where threshold values for fig. 2c (iv, v) were selected.

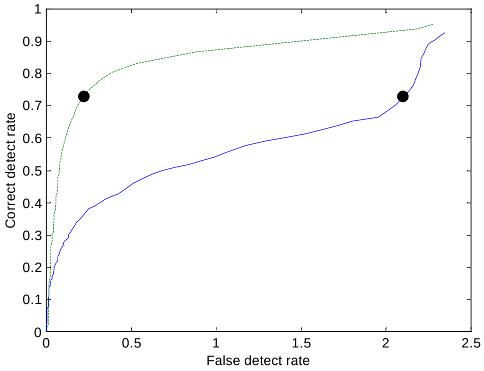

Fig. 3d- ROC curve for black adult image (2d). Dashed line - proposed model; solid line - single-gaussian model. Black dots point where threshold values for fig. 2d (iv, v) were selected.

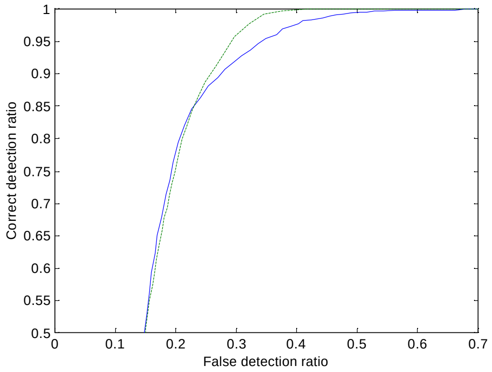

Figure 4. Average performance over 100 test images.

In all the cases in Fig. 3, we can observe that, for a given high correct detection rate, the false detection is lower in our model than in the one-cluster model. Indeed, this fact has been observed in *every* image from the test set – above a certain correct detection ratio, our model performs better than the single-gaussian model. We have averaged the performance values obtained in the 100 test images in order to obtain a general result. The resulting ROC curves are shown in Fig.4.

It can be easily seen that, in the case of our test set, the proposed model has better performance for detection rates above 85%. This result shows that our model is more appropriate to avoid excessive false detected areas while still remaining a high degree of correct detection.

7. Conclusions and future work

We have presented a gaussian mixture approach to model the human skin color in a race-independent way. A sample set from white and black people was obtained and we used the EM algorithm to find maximum-likelihood estimates for a two-class gaussian mixture model. Our results show that the mixture model has a better performance, compared to the model with a single gaussian, regarding the number of false detections for high correct classification. Major applications are face detection and tracking in surveillance, security and authentication systems.

We are currently working on a improved version of the model, where more specific ethnies are being added to the sample set and more than two gaussians are used to build the mixture.

Acknowledgements

The authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the financial support, necessary to accomplish this work.

References

[1] T.K. Leung, M.C. Burl, P. Perona, Finding Faces in cluttered scenes using random labeled graph matching, Proceedings of the 5th. Int. Conf. on Computer Vision, 1995, pp. 637-644.

[2] H. A. Rowley, S. Bluja, T. Kanade, Neural network-based face detection, IEEE Trans. on Patt. Anal. Mach. Int. 20 (1) 1998, 23-38.

[3] K.-K. Sung, T. Poggio, Example-based learning for view-based human face detection, IEEE Trans. Patt. Anal. Mach. Int. 20 (1) 1998, 39-51.

[4] K.C. Yow, R. Cipolla, Feature-based human face detection, Image and Vision Computing 15 (1997) 713-735.

[5] K.C. Yow, R. Cippola, A probabilistic framework for perceptual grouping of features for human face detection, Proceedings of the 2nd. Int. Conf. on Automatic Face and Gesture Recognition, Vermont, USA, 1995.

[6] M.C. Burl, T.K. Leung, P. Perona, Face locatization via shape statistics, Procee-dings of the First International Workshop on Face and Gesture Recognition, Zurich, Switzerland, 1995.

[7] J. Cai, A. Goshtasby, Detecting human faces in color images, Image and Vision Computing, 18 (1999) 63-75.

[8] E. Saber, A.M. Tekalp, Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions, Pattern Recognition Letters 19 (1998) 669-680.

[9] J-G. Wang, E. Sung, Frontal-view face detection and facial feature extraction using color and morphological operations, Pattern Recognition Letters 20 (1999) 1053-1068.

[10] Y.A. Waibel, A real-time face tracker, Proceedings of WACV'96, Sarasota, USA, 1996.

[11] E. Littmann, H. Ritter, Adaptive color segmentation. A comparison of neural and statistical methods, IEEE Trans. Neural Networks 8 (1) (1997) 175-185.

[12] R.O. Duda, P.E. Hart, Pattern Clas-sification and Scene Analysis, John Wiley and Sons, 1973.

[13] A. P. Dempster, N.M. Laird, D.B. Rubin, Maximum-Likelihood from incomplete data via the EM algorithm, Journal of Royal Stat. Soc. 39 (1) (1977) 1-38.