PERGAMON

Pattern Recognition 36 (2003) 3019–3021

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Rapid and brief communication

Do mixture models in chromaticity space improve skin detection?

Tibério S. Caetano[a,*], Sílvia D. Olabarriaga[b], Dante A.C. Barone[a]

[a] *Instituto de Informática, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, 91501-970, Caixa Postal 15064, Porto Alegre, RS, Brazil*

[b] *Image Sciences Institute, University Medical Center, E 01.335, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands*

Received 27 January 2003; accepted 6 February 2003

Abstract

This note reports an experiment where a single Gaussian model and several Gaussian mixture models were used to model skin color in the *rg* chromaticity space. By using training and test databases containing millions of skin pixels, we show that mixture models can improve skin detection, but not always. There is a relevant operating region where no performance gain is observed.

Keywords: Skin color model; Skin color segmentation; Skin detection; Gaussian mixture model; Performance evaluation

1. Introduction

The single Gaussian model (SGM) [1] has been widely used for modeling the skin color distribution. More recently, Gaussian mixture models (GMMs) have been proposed for this task (e.g. Refs. [2,3]), where it was suggested that they could outperform SGM. However, these works do not provide an experimental comparison of the two modeling approaches for skin color modeling. A detailed analysis on their relative performances is still lacking.

This note presents an experimental investigation which aims at comparing performances of modeling skin color with SGM and GMM in the *rg* chromaticity space, which is the most popular color space for skin color modeling [4] —what justifies our choice for it since the impact of a new result is larger.

Our experiment shows an interesting result. GMMs just improve performance in a very specific operating region, when true and false positive rates are high.

* Corresponding author. Tel.: +1-780-492-6084; fax: +1-780-492-1071.

E-mail address: caetano@inf.ufrgs.br (T.S. Caetano).

2. Methodology

The methodology used in the experiment is described as follows:

Training and test sets: The first step of the experiment consists of selecting representative skin samples for training and testing. In the *rg* chromaticity space, each color is represented by a vector $\mathbf{v} = (r, g)$, where $r = R/(R+G+B)$ and $g = G/(R+G+B)$. This work considers Internet images, which were acquired under unknown conditions. Skin pixels were selected by manually cropping skin regions with closed polygonal curves. Non-skin pixels were also stored for test purposes. All test pixels were *labeled* as being skin or not, once this is needed during the evaluation step. The sampling was systematic: (i) a total amount of 21,216,152 pixels was collected from people belonging to several ethnic groups, such as Caucasian, African, Hispanic and Asian;(ii) pixels were acquired from several regions of the human skin, including the face, arms and legs; (iii) images containing strong highlights and shadows in skin areas were discarded. The final sample was divided in two equal parts: one for training and the other for testing, so each containing 10,608,076 pixels. The non-skin pixels, which totalized 440,451,063, were added to the test set.

0031-3203/03/$30.00 © 2003 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/S0031-3203(03)00116-X

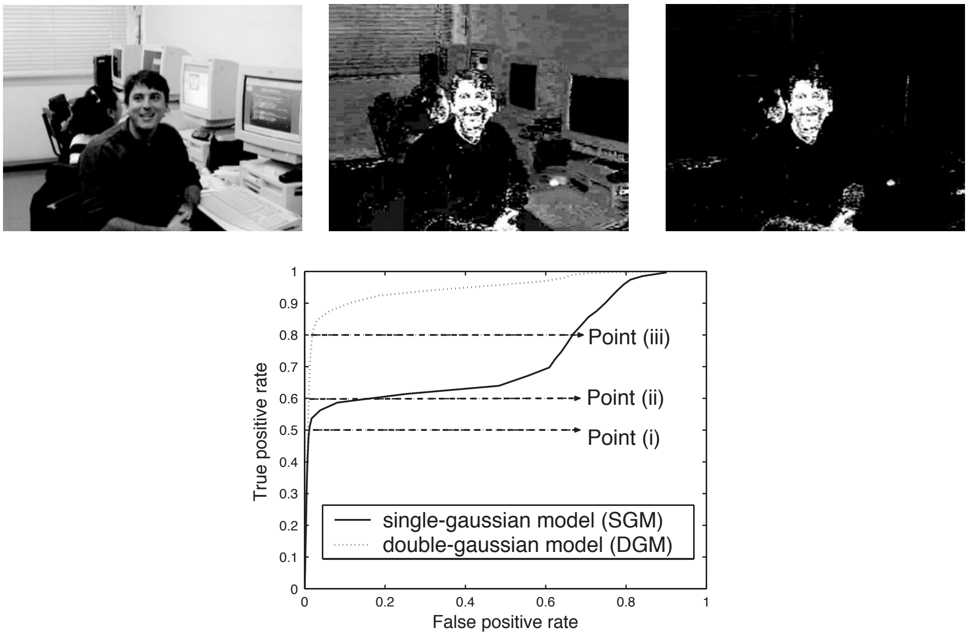

Fig. 1. A sample test image (top left-original colored) and skin-likelihood images for the SGM (top center) and for the DGM (top right). The graphic shows the ROC curves for both SGM and DGM applied to the image: the relative performance depends on the operating point.

Single Gaussian and Gaussian mixture modeling (SGM-GMM): From the training set, a Gaussian model is estimated analytically via the maximum-likelihood (ML) criterion and seven mixture models (from 2 to 8 Gaussian components) are estimated via the expectation-maximization algorithm [3], which also yields ML estimates. Given a pixel with color coordinates (r, g), each model returns an output proportional to the probability of this pixel being skin. This output we call *skin-likelihood*.

ROC curves construction: Each model is applied to the test set as follows. For each pixel in the test set, the corresponding output produced by the model is calculated. If this value is above a given threshold, the pixel is considered to be skin; otherwise, it is regarded as non-skin. This classification result is then compared with the correspondent *label* for this pixel in order to determine how many skin pixels were correctly classified as skin (true positives) and how many non-skin pixels were incorrectly classified as skin (false positives). By setting multiple thresholds in this scheme, we generate receiver operating characteristics (ROC) curves. Fig. 1 shows an example of a test image, the skin-likelihood images for the SGM and the double Gaussian model (DGM), and the correspondent ROC curves.

3. Results

The performance evaluation scheme was applied to the entire test set containing 10,608,076 skin pixels and 440,451,063 non-skin pixels. Eight models were estimated from the training set, with 1–8 Gaussians. For each model, the corresponding ROC curve was generated following the procedure outlined in Section 2. Fig. 2 presents a plot with the ROC curves for the eight models.

Some conclusions are obtained by inspecting the plots. First, the SGM has poorer performance for medium to high true positive rates (TPRs), while for low TPRs its performance is comparable to those of the GMMs. Second, all GMMs display very similar performance over the whole range of possible operating points. Consequently, mixture models are not *necessarily* the best option for skin color modeling in *rg* chromaticity space, but just under the special condition of high TPRs.

4. Conclusions

This note has presented a performance evaluation of the SGM and several GMMs for the human skin color representation. An experimental setup was designed where a data set of skin pixels obtained from Internet images was used to train an SGM and seven versions of GMMs. The eight models so obtained were applied to a test set also containing images from the Internet, but different from those of the training set. By labeling skin and non-skin regions in the test images, ROC curves were computed. The analysis of those curves lead to two main conclusions. First, the SGM exhibits inferior average performance from medium to high

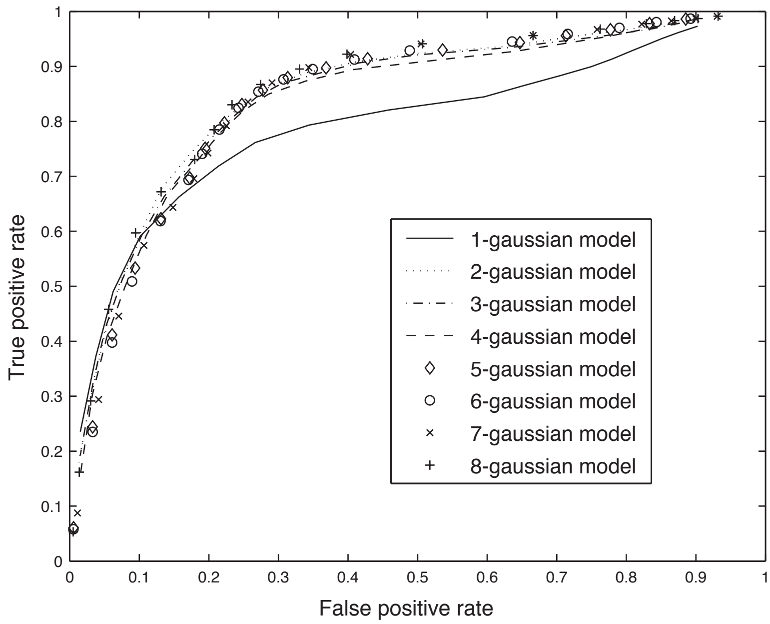

Fig. 2. ROC curves for the SGM and for the several GMMs obtained by applying the models to the entire test set.

TPRs. Second, the GMMs average performances are very similar. These results suggest that skin color mixture models may be more appropriate than the SGM just when high correct detection rates are needed.

We are currently working on evaluating the relative performances of SGM and GMMs for different color spaces in an even larger skin database.

Acknowledgements

T.S.C. thanks Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil. We are grateful to T. Caelli for helpful suggestions. F.B. Rizzato and A. Petry provided critical readings of the manuscript.

References

[1] J. Yang, A. Waibel, A real-time face tracker, Proceedings of the IEEE Workshop Applications of Computer Vision, Princeton, NJ, 1996, pp. 42–147.

[2] S.J. McKenna, S. Gong, Y. Raja, Modelling facial colour and identity with Gaussian mixtures, Pattern Recognition 31 (1998) 1883–1892.

[3] T.S. Caetano, D.A.C. Barone, A probabilistic model for the human skin color, IEEE International Conference on Image Analysis and Processing, Palermo, Italy, 2001, pp. 279–283.

[4] M.-H. Yang, D. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 24 (1) (2002) 236–274.