

# A UNIFIED ADAPTIVE APPROACH TO ACCURATE SKIN DETECTION

Qiang Zhu    Kwang-Ting (Tim) Cheng    Ching-Tung Wu

Dept. of Electrical and Computer Engineering  
University of California at Santa Barbara, CA, 93117, USA

## ABSTRACT

Due to variations of lighting conditions and camera hardware settings and the existence of many ethnic people with a wide range of skin colors, a generic skin model is often inadequate to accurately capture the skin distribution for individual images. In this paper, we propose an adaptive skin detection framework, which allows modeling true skin distribution with significantly higher accuracy and flexibility. First, an adaptive skin model, specific to the image under consideration and refined from the Skin-Similar space, is derived using a Gaussian Mixture Model (GMM) and standard Expectation-Maximization (EM) algorithm. Then, we develop a Support Vector Machine (SVM) classifier to identify the skin Gaussian from the trained GMM (with two Gaussian components) by incorporating spatial and shape information of skin pixels. Extensive experimental results performed on large image databases have demonstrated the effectiveness and benefits of the proposed approach.

## 1. INTRODUCTION

Color usually presents a strong intuitive cue in complex scene images. In recent years, skin detection has emerged as an active research topic in several practical applications, including human body segmentation [1], and objectionable-image filtering [2]. Researchers have been studying various generic skin models in a number of color spaces [3, 4]. However, we can expect variations when images are photographed in various settings, with different kinds of camera hardware, and under a wide range of lighting conditions. Moreover, ethnic groups present a range of skin tones that defy simplistic classification. Therefore, a generic skin model is clearly inadequate to accurately capture the skin-colors in individual images. Specific to the application of face-tracking in videos, an effective idea is to adapt skin models to the current frame by incorporating the information from previous frames [5, 6]. Even though good results have been reported, such techniques can only be applied to video applications, but not for still images.

## 2. OUR APPROACH

In this paper, we develop a unified, adaptive skin-color model that is applicable to arbitrarily still images. In contrast to other techniques, our adaptive model is derived entirely from the information contained in the given image. Specifically, for given images, the adaptive skin model is based on a two-step process. A generic skin model, built from a comprehensive training dataset, is first used to classify the pixels in the given image. The pixel set identified as skin pixels in this step is designated as Skin-Similar space. We have observed several interesting characteristics inherent in this *Skin-Similar* space.

- The true skin color distribution within this space for an image is usually consistent. Thus, a dominant Gaussian can typically be found in this Skin-Similar space, which can be interpreted later as the skin distribution.
- In addition, a large fraction of the false skin pixels in the Skin-Similar space of a given image usually belong to the same object in the image background (e.g. pink wallpaper, curtains with skin-similar colors, etc.). Thus, these pixels often form another weak Gaussian component.
- We observed that for most images in the large collection of test images gathered from the web, these two Gaussians appear to be quite separable.

We note that our proposed method, in principle, divides the hard task (i.e. analyzing the color distribution of an arbitrary image) into two easier sub-tasks. In the first step, a high detection rate, say over 96% is needed to ensure that almost all skin pixels are captured (inevitably a high false-positive rate occurs at this step). The goal of the second step is to reduce the false-positive rate without compromising much on the detection rate. One key observation is already apparent: in this simplified, Skin-Similar space, the GMM with two Gaussian components is generally sufficient to accurately model the real color distribution because the pixels in a noisy background and/or in obvious non-skin regions have been removed from consideration. Therefore, in Step 2, the pixels in the Skin-Similar space are used to train a GMM with two Gaussian kernels (one for modeling the true skin pixels and the other for modeling the false skin pixels) using the

standard EM algorithm. We believe this work is the first attempt to refine the generic skin model using an adaptive, two-step process for arbitrary still images. More details of both steps will be discussed later.

To validate the technique, we built a comprehensive Test Database for Skin Detection (*TDS*), which contains 554 images. Specifically, we chose a collection of images including skin pixels under various lighting conditions and from different ethnic groups. We have manually labeled the skin region for each image as our ground truth. From this collection of images, 24 million skin pixels and 75 million non-skin pixels were identified. It should be mentioned that all experiments in the paper are based on this proposed database, and the popular HSV color space is adopted in our work.

The remainder of our paper is organized as follows. Section 3 details EM-based adaptive modeling. In Section 4, we incorporate the spatial and shape information of skin pixels into the step of identifying the right skin Gaussian from the trained GMM (with two Gaussian components) using SVM classifier. Section 5 presents some experimental results demonstrating the value of the adaptive skin model. In the last Section, we conclude with a short discussion of ideas for future work.

### 3. UNIFIED ADAPTIVE MODELING VIA EM

We now expand upon our earlier comment that the color distribution in the Skin-Similar space can be modeled by a GMM that contains two Gaussian components, one for modeling the true-skin pixels and the other for modeling the false-skin pixels. A GMM is defined as:

$$P_{hs} = \sum_{i=1}^k w_i 2\pi^{-1} |\Sigma_i|^{-1/2} \exp(-\lambda_i (h,s)^2) \quad (1)$$

$$\lambda(h,s) = (1/2) \times [(h,s) - \bar{u}]^T \Sigma^{-1} [(h,s) - \bar{u}] \quad (2)$$

where  $\sum_{i=1}^k w_i = 1$  and  $k$  is equal to two in our case.

Neglecting obscure substantiations, the key updated equations for an EM algorithm are given below. More details of the EM algorithm can be found in [7].

$$w_l^{new} = \frac{1}{N} \sum_{i=1}^N p(l/x_i, \Theta^g) \quad u_l^{new} = \frac{\sum_{i=1}^N x_i p(l/x_i, \Theta^g)}{\sum_{i=1}^N p(l/x_i, \Theta^g)} \quad (3, 4)$$

$$\Sigma_l^{new} = \frac{\sum_{i=1}^N p(l/x_i, \Theta^g) (x_i - u_l^{new})(x_i - u_l^{new})^T}{\sum_{i=1}^N p(l/x_i, \Theta^g)} \quad (5)$$

where  $w, u, \Sigma$  are the weight, mean vector, and

covariance matrix in the Gaussian Mixture Model.  $l$  indicates which component of the GMM and  $x_i$  represents one sample.  $N$  is the number of total training samples. With the current model parameter of  $\Theta^g$ ,  $p(l/x_i, \Theta^g)$  evaluates the probability of sample  $x_i$  belonging to Gaussian kernel  $l$ .

When the number of training samples is small, EM basically performs an unsupervised clustering task in the data space. If the structure, such as the number of components in the mixture models, is known in advance, EM, with a good initial guess, could converge to the true model parameters. Note that both a correct structure assumption and a good initial guess are crucial for producing accurate models using the EM algorithm. The analysis given in the last section indicates that using two Gaussian components to model pixels in the Skin-Similar space should strike the best balance between detection accuracy and cost effectiveness. (For the false skin pixels, multiple Gaussian components might fit the distribution even better; however, for the sake of computation efficiency, one Gaussian for false-skin pixels is adopted and it works well as shown in the experimental results.) Based on this structure assumption, two issues need to be resolved further: how to obtain a good initial guess, and how to distinguish the skin Gaussian from the trained GMM (with two Gaussian components).

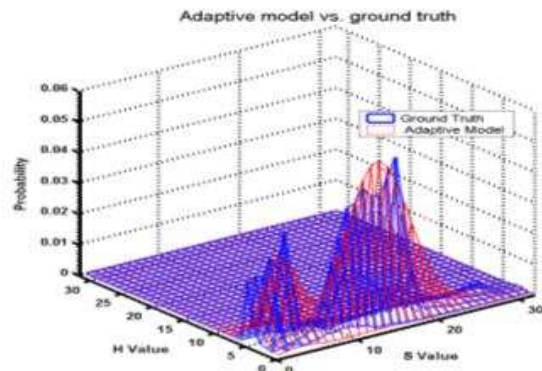


Fig. 1 GMM model trained by EM vs. Ground truth

We have observed that the two assumed Gaussian components (for true- and false-skin pixels) are more separable at the S value. So, we calculate the marginal probability over the H value in the 2-dimensional HS color space. Basically, we choose two separate peaks from the marginalized curve as the initial mean values for the two Gaussian components. For the covariance matrix, empirically we set it to a small initial value. In Fig. 1, we show the visual result of our adaptive model trained by the standard EM algorithm. As you can observe, excellent matches have been observed between the trained model

(red surface) and the ground truth (blue surface).

Now with two Gaussian components built by the EM algorithm, the next step is to identify which of the two components represents the skin Gaussian.

#### 4. IDENTIFICATION OF SKIN GAUSSIAN

To identify the skin Gaussian, some heuristic rules might be applied. For example, the dominant Gaussian, i.e. the component with a larger weight, can be chosen as the skin model or the one with smaller variance (we always expect compact skin distribution in individual images) is preferred. However, such simple strategies will be challenged for a large collection of arbitrary images. As the skin-detection performance could be seriously degraded if the wrong Gaussian is chosen, a more intelligent method is desirable.

The intuitive cues to describe a Gaussian kernel in the trained GMM are its weight, mean and variance. Even though these features are useful, the spatial and shape information for the pixel set corresponding to a Gaussian distribution presents additional distinguishable characteristics. For example, skin pixels often form compact shapes and less likely appear on the borders of the image, while the non-skin pixels may spread over the whole image. Hence, we can define them and incorporate this information into the process of classifying the right skin Gaussian. In our experiment, seven types of features, falling into two different categories, are defined below:

- Group A (color distribution related): weight, Gaussian mean, Gaussian variance.
- Group B (spatial and shape related): spreadness, elongation, X-direction and Y-direction histograms.

Now, with these features, we design an efficient classifier to classify each Gaussian sample (into either skin or non-skin). Our classifier is based on SVM which is a popular and powerful technique for data classification. Basically, SVM projects the feature space into high dimension space to find a hyper-plane, which theoretically best separates samples belong to different classes. In our experiment, we use LIBSVM [8] which is a library for support vector classification (SVM) and regression.

For this experiment, we further manually labeled 1120 images. Then, using a generic skin model, we generate the Skin-Similar map for each image. In the Skin-Similar space, we build two Gaussians for skin pixels and non-skin pixels respectively based on the manually labeled ground truth. In addition, we calculate the other four spatial and shape features (Group B) for each pixel set individually. These 2240 samples (two Gaussians, skin and non-skin, for each image) are then used to train our SVM classifier. In the training process, some common techniques, such as scaling feature values and cross-

validation based model selection, were applied. Following the same feature extraction process, we generated 1108 samples from our *TDS*D database (with 554 images). These samples were used to test the trained classifier. Table 1 shows the experimental results.

Feature space	Training Accuracy	Testing Accuracy
Group A	91.0188%	87.7256%
Group B	88.9634%	90.9747%
Group A+B	96.8275%	96.5704%

Table 1: SVM classification during training and testing

From Table 1, we present three groups of experimental results. The first row gives the training accuracy and testing accuracy using only Group A features for training and classification. The training accuracy is referred to the classification accuracy on the 20% cross-validation set of the training data (which is randomly picked and not used in training). The second row shows the corresponding results using only Group B features. In the last row, features of both groups are used in the classifier. Convincingly higher accuracy is observed for both training set and testing set. This experiment strongly indicates that, by combining the GMM parameters and the spatial and shape features, SVM can accurately identify the right skin Gaussian component in the Skin-Similar space.

#### 5. EXPERIMENTS ON IMAGE DATABASE

For the convenience of comparing the performance between a generic skin model and an adaptive skin model, we constructed various generic skin models (Histogram Model, Single Gaussian Model and Gaussian Mixture Model) from a huge training dataset including 151 million skin pixels and 448 million non-skin pixels. Then, we applied these generic models to the *TDS*D database. Based on the final classification result, we finally chose the GMM model with five Gaussian kernels as the generic skin model for the first-step skin classifier in our two-step adaptive process. Note that all the following experiments and analyses were performed using our *TDS*D database.

In Fig. 2, we compare the final performance of skin detection among the generic skin model (black curve), the upper bounds<sup>1</sup> of the adaptive method (blue curve), and the result of our two-step adaptive process (red curve). These results clearly demonstrate that the proposed adaptive method is superior to the single-step method using the best available generic skin model (the GMM with 5 Gaussian components). For example, at 92% detection rate, we reduce the FP rate by about 7.8% (from

<sup>1</sup> We directly build the Skin / Non-skin Gaussian from the ground truth. Such a skin classification result serves as a theoretical upper bound for our proposed method.

37.8% down to 30%) using the two-step adaptive approach. On the other hand, the generic skin model can achieve a DR of only 86% at a 30% FP rate.

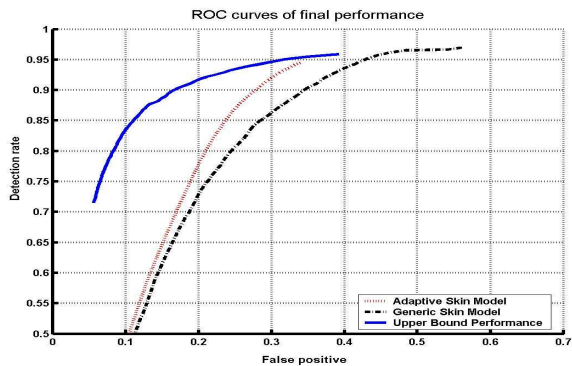
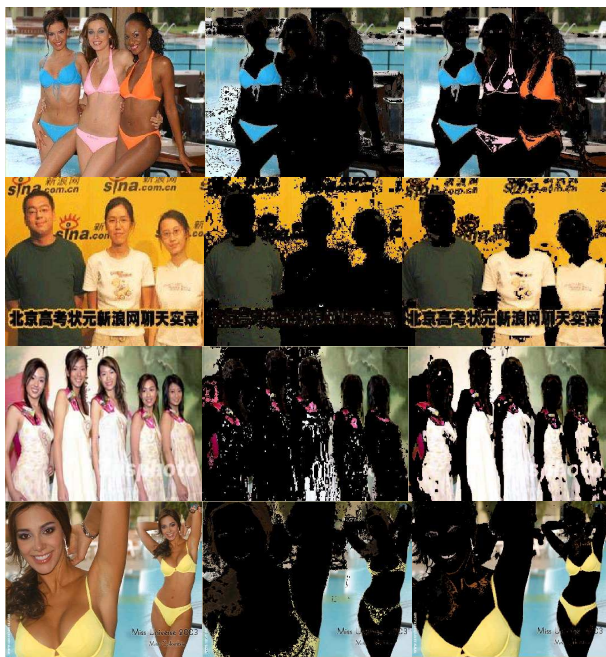


Fig. 2 Two-step adaptive method vs. Generic model

We further demonstrate several typical examples of segmenting the skin regions: some containing people of different races and the others complicated by a confusing background. As you can see, the proposed adaptive approach works much better than the generic skin model.



Original image      Generic model      Adaptive model

Fig. 3 Body segmentation in arbitrary images

## 6. CONCLUSION

This paper presents a novel two-step adaptive framework for accurate skin-color detection. In the first step, we identify the Skin-Similar pixels using a generic

skin model. The standard EM algorithm is then used to train a GMM, with two Gaussian components, in this reduced pixel space. Then an SVM classifier, using the spatial and shape features along with Gaussian parameters, is proposed to identify the right skin Gaussian component from the trained GMM. In comparison with traditional methods which rely on a generic skin model, the experimental results indicate that the new method achieves a significantly lower false-positive rate for skin detection. One of the directions that future research might take is to develop better ways to integrate multi-cues, including color, texture, spatial and shape, into an even more powerful classifier for the skin detection task. Furthermore, we will continue to exploit additional applications, say objectionable image filtering, for the proposed adaptive skin model.

## 7. REFERENCES

- 1) R.L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face detection in Color Images", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696-706, May 2002.
- 2) D.A. Forsyth and M.M. Fleck, "Automatic Detection of Human Nudes", *Int. Jour. of Computer Vision*, Vol. 32, No. 1, pp. 63-77, Aug., 1999.
- 3) Y. Wu and TS Huang, "Nonstationary Color Tracking for Vision-Based Human-Computer Interaction", *IEEE Trans. on Neural Networks*, vol. 13, pp. 948-960, July 2002.
- 4) J. Yang, W. Lu, and A. Waibel, "Skin-Color Modeling and Adaptation", *Proc. of the 3rd Asian Conference on Computer Vision (ACCV98)*, Vol. 2, pp. 687-694, Jan. 1998, Hong Kong.
- 5) M.J. Jones, J.M. Rehg, "Statistical Color Models with Application to Skin Detection", Technical Report, Cambridge Research Lab., 1998.
- 6) J.C. Terrillon, M.N. Shirazi, H. Fukamachi, and S. Akamatsu, "Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images", *Proc. IEEE Intl. Conf. on Face and Gesture Recognition (FG 2000)*, pp. 54-61, Mar. 2000, Grenoble, France
- 7) J. Bilmes, "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models", Technical Report 97-021, Intl. Computer Science Institute, Univ. of California, Berkeley, Apr. 1998.
- 8) Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>