

Adaptive Learning of an Accurate Skin-Color Model

Qiang Zhu Kwang-Ting Cheng Ching-Tung Wu Yi-Leh Wu*

Electrical and Computer Engineering, University of California, Santa Barbara

*VIMA Technologies Inc. Santa Barbara, CA, USA

qzhu@ece.ucsb.edu timcheng@ece.ucsb.edu

Abstract

Due to variations of lighting conditions, camera hardware settings, and the range of skin coloration among human beings, a pre-defined skin-color model cannot accurately capture the wide distribution of skin colors in individual images. In this paper, we propose an adaptive skin-detection method, which allows modeling true skin-color distribution with significantly higher accuracy and flexibility than other methods attain. In principle, the proposed method follows a two-step process. For a given image, we first perform a rough skin classification using a generic skin model which defines the Skin-Similar space. In the second step, a Gaussian Mixture Model (GMM), specific to the image under consideration and refined from the Skin-Similar space, is derived using the standard Expectation-Maximization (EM) algorithm. Then, we use an SVM (Support Vector Machine) classifier to identify the skin Gaussian from the trained GMM (which contains two Gaussian components) by incorporating spatial and shape information of the skin pixels. This adaptive method can be applied to both still images and video applications. Results of extensive experiments performed on live video sequences and large image databases have demonstrated the effectiveness and benefits of the proposed model.

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 detection [1, 2, 3], face tracking [4, 5, 6], and objectionable-image filtering [7, 8]. Researchers have been studying various generic skin models in a number of color spaces [9, 10]. 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. To improve the detection accuracy and reduce the false positive rate, one solution is to adopt an adaptive skin model instead of a static one. 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. An interesting study was

reported in [6] to explain the skin model transformation in consecutive frames with an affine motion. Under this affine motion assumption, the authors propose a second-order Markov model to predict the changes. They further describe a method to learn motion parameters from a sequence. Even though good results have been reported, it is not clear how much of the improvement is due to the sophisticated prediction strategy, and how much to the adaptive skin model itself. It should be noted that existing adaptive approaches attempt to explore the relative similarities and differences between the current and the previous frames. Therefore, such techniques can only be applied to videos, but not for still images.

In [2], a novel adaptive approach is proposed for the application of hand-segmentation for arbitrary colors. A restricted EM algorithm is introduced to train an adaptive GMM for still images, wherein the background is modeled by four Gaussian kernels, and the hand color is modeled by one Gaussian. This modified EM algorithm requires strict prior information, including a good estimation of hand color and a reasonable bound of weighted values, in order to obtain more robust results. To distinguish the skin component from other Gaussian components, the authors heuristically fix the mean of the first Gaussian in estimating the hand color during the training process. Obviously, this significantly degrades the ability of the GMM to model the actual skin color distribution for individual images.

2. Our Approach

In this paper, we develop a unified, adaptive skin-color model that is applicable to arbitrarily selected 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 shown in Fig. 1. 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.

- For individual images, the true skin color distribution within this Skin-Similar space is pretty consistent. A dominant Gaussian can typically be found in the new space, which can be interpreted later as the skin distribution model.

- 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.
- We observed that for most images in the large collection of testbed images gathered from the web, these two Gaussians appear to be quite separable.

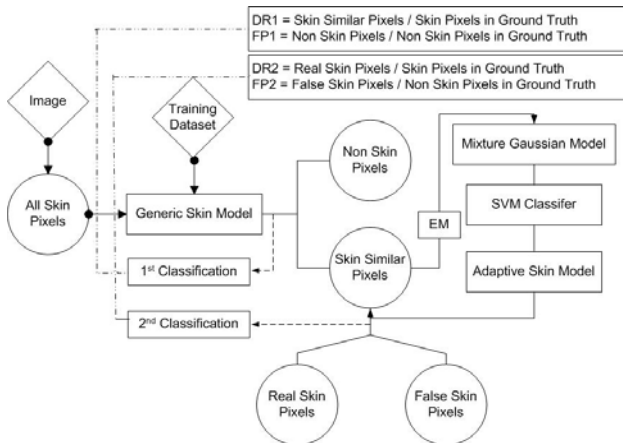


Fig. 1 Two-step adaptive framework for skin detection

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 such an adaptive, two-step process for arbitrary still images. More details of both steps will be discussed later.

The remainder of our paper is organized as follows. Section 3 presents a quantitative analysis to the Skin-Similar space identified by applying a generic skin model to a large image database. The proposed EM-based adaptive modeling is detailed in Section 4. In Section 5, we incorporate the spatial and shape information of the image pixels into the step of identifying the correct skin Gaussian from the trained GMM using an SVM classifier. Section 6 presents some experimental results to demonstrate the value

of the adaptive skin model. In the last Section, we conclude with a short discussion of ideas for future work.

3. Quantitative Analysis of Skin-Similar Space

To validate the technique, we built a comprehensive Test Database for Skin Detection (*TSDS*), 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. Later, we performed two experiments to analyze the color distribution of the Skin-Similar space for each image in our *TSDS* database.

We adopted the popular HSV color space for our experiments. Furthermore, the HSV space was reduced to its HS subspace by ignoring the V component, which contributes very little for the discrimination of skin tone.

3.1 Goodness of Fit of Gaussian

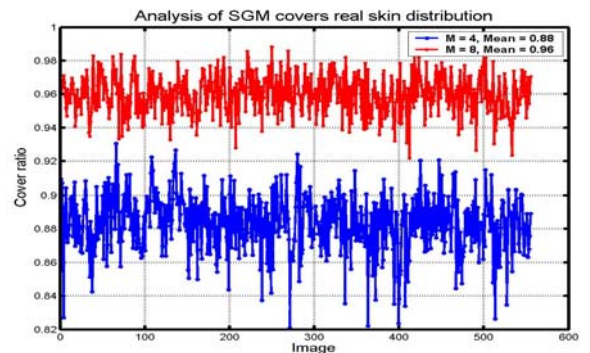


Fig. 2 Skin ratio with a fixed Mahalanobis distance

While there is no ideal metric for evaluating the “goodness of fit” for an assumed Gaussian, we have designed an intuitive metric to measure how well a Gaussian kernel models the true-skin distribution in the Skin-Similar space for a given image. This metric is much simpler than the sophisticated methods proposed in [11]. For each image in our test database, as we have manually labeled the skin region (the “ground truth”), we calculate the mean vector and covariance matrix for the assumed skin Gaussian defined in Equation (1).

$$P_{hs} = 2\pi^{-1} |\Sigma|^{-1/2} \exp(-\lambda(h,s)^2) \quad (1)$$

where $\lambda(h,s) = (1/2) \times [(h,s) - \bar{u}]^T \Sigma^{-1} [(h,s) - \bar{u}]$ is the Mahalanobis distance. Then, for a given Mahalanobis distance, we evaluate how many skin pixels fall within the region bounded by the given distance. In general, if that assumed Gaussian covers all or almost all true skin pixels for a small distance, it indicates a compact skin distribution in the Skin-Similar space. It further implies the suitability of modeling this distribution with a compact Gaussian kernel with respect to each individual image.

Fig. 2 shows the analysis result for the 554 test images in our *TDS*D database using this metric. Each point in the curve represents the ratio of skin pixels covered by the Gaussian kernel with a given Mahalanobis distance for one image. We observe that, for most images, a very high ratio is obtained within a small distance (an average of 88% and 96% for a distance of 4 and 8 respectively). Considering that the color space is quantitated by a 32×32 scale, this result strongly indicates that we can train a compact Gaussian (i.e., one with a small variance matrix) to fit the true-skin distribution in the Skin-Similar space.

3.2 Upper Bound Analysis

Proving a compact skin distribution in the Skin-Similar space shows that a high detection rate can be achieved, but only if we model this distribution properly. In this section, we need to analyze the overlap between the true skin pixels and the false skin pixels in the Skin-Similar space. This analysis will reveal the potential room for reducing the amount of false-skin pixels. Inasmuch as we already have the ground truth for each test image, the Skin-Similar space can be divided into two parts: true skin pixels set and false skin pixels set. A histogram can be derived from each set. The overlap between the two histograms is defined as:

$$\text{overlap} = \sum_{h=1}^{32} \sum_{s=1}^{32} \text{nonskin}_{hs} (\text{skin}_{hs} > T) \quad (2)$$

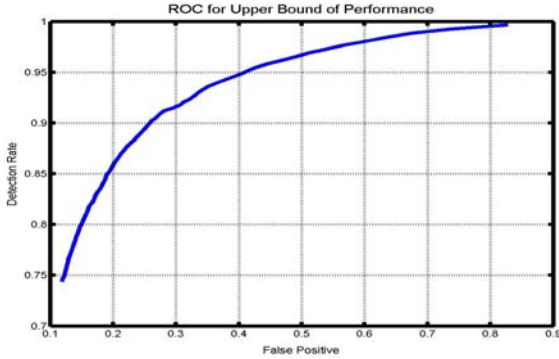


Fig. 3 Upper Bound ROC of Performance

Basically, this equation sums up all false skin pixels misclassified by a skin histogram model with respect to a given threshold T . With a given desirable detection rate of true skin pixels, i.e. fixed T value, the “theoretical lower bound” (It should be noted that these two histogram models are constructed from the ground truth.) of the false-positive rate can be evaluated using the overlap equation (2). By adjusting the threshold T , we can plot the “skin-detection rate vs. false-positive rate” ROC curve in the Skin-Similar space. This ROC curve serves as a theoretical upper bound of the performance improvement achievable by our second skin classification in the two-step process.

In Fig. 3, we plot the ROC curve of classifying the skin pixels in the Skin-Similar space for the 554 test images. For example, suppose we choose the point, corresponding to a 96% detection rate (DR) and 50% false-positive rate (FP)

for the first skin classification. For the second classification in the Skin-Similar space, if we choose the point corresponding to a 95% DR and a 40% FP from the curve in Fig. 3, the overall performance of the two-step classification will result in a 91.2% DR and a 20% FP. That is, the second classification reduces the FP by 30% (from 50% down to 20%) at the cost of 4.8% reduction in DR (from 96% down to 91.2%). This example reveals the potentials of this two-step classification framework of achieving low FP without compromising much on DR.

4. Unified Adaptive Model Learned by 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 Gaussian Mixture Model is defined as:

$$P_{hs} = \sum_{i=1}^k w_i 2\pi^{-1} |\Sigma_i|^{-1/2} \exp\left(-\lambda_i (h,s)^2\right) \quad (3)$$

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 theoretical details of the EM algorithm can be found in [12].

$$w_l^{new} = \frac{1}{N} \sum_{i=1}^N p(l/x_i, \Theta^g) u_i^{new} = \frac{\sum_{i=1}^N x_i p(l/x_i, \Theta^g)}{\sum_{i=1}^N p(l/x_i, \Theta^g)} \quad (4, 5)$$

$$\Sigma_l^{new} = \frac{\sum_{i=1}^N p(l/x_i, \Theta^g) (x_i - u_l^{new})(x_i - u_l^{new})^T}{\sum_{i=1}^N p(l/x_i, \Theta^g)} \quad (6)$$

where w, u, Σ 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 Θ^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 an accurate model 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: (1) how to obtain a good initial guess, and (2) how to distinguish the skin Gaussian from the trained GMM (with two Gaussian components).

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 of S value as the initial mean vector for the two Gaussian components. For the covariance matrix, empirically we set it to a small initial value. In Fig. 4, we show one visual example of the trained GMM using the standard EM algorithm. Excellent matches have been observed between the trained model and the ground truth.

Now with two Gaussian components built by the EM algorithm, the next step is to identify which of the two components in the trained GMM represents the skin Gaussian in the Skin-Similar space.

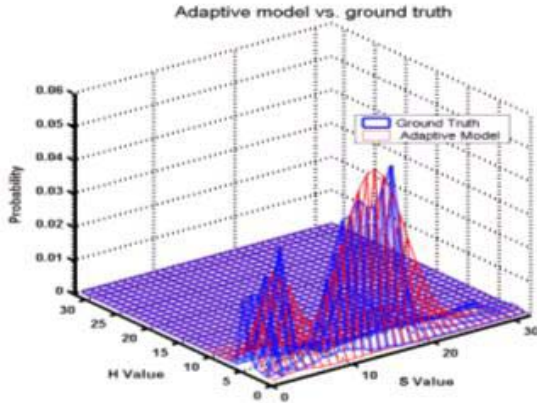


Fig. 4 GMM model trained by EM vs. Ground truth

5. Identification of Skin Gaussian by SVM

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 to 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 as follows:

- Group A (color distribution related): weight, Gaussian mean, Gaussian variance.
- Group B (spatial and shape related): spreadness, elongation [13], X-direction and Y-direction histograms.

Now, with these features, we design an efficient classifier to classify each Gaussian component into either skin or non-skin. Our classifier is based on SVM [14] 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 [15] which is a library for support vector classification (SVM) and regression.

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 labelled 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 *TDSD* 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 Gaussian parameters and the spatial and shape features, SVM can accurately identify the right skin Gaussian from the trained GMM in the Skin-Similar space.

6. Experiments and Applications

6.1 Skin Classification on *TDS* Test Set

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 all these generic models to the *TDS* 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* database.

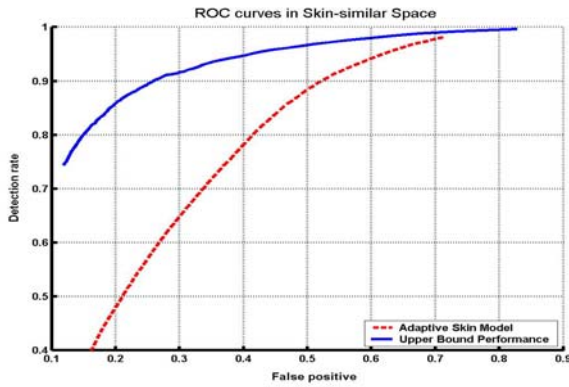


Fig. 5 Performance analysis in Skin-Similar space

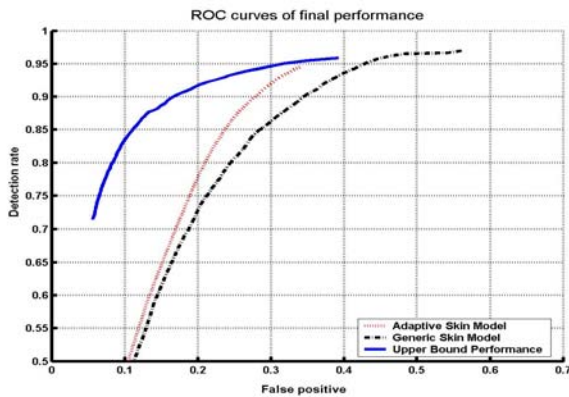


Fig. 6 Two-step adaptive method vs. Generic model

Two ROC curves are plotted in Fig. 5. The blue, solid curve represents the upper bound of Step-2 performance as discussed in Section 3.2. The red, dotted curve indicates the performance of our adaptive GMM model in the Skin-Similar space. We noticed that our adaptive model works quite well at high detection rates, where the curve is very close to the upper bound curve. This implies that our method can effectively separate the separable skin and non-skin pixels in the Skin-Similar space. In Fig. 6, we compare the ultimate performance of skin detection among the

single-step method using a generic skin model (black curve), the upper bounds of the two-step method (blue curve), and our final 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 37.8% down to 30%) using the two-step adaptive approach.

6.2 Application to Human Body Segmentation

With the proposed skin model, we can segment the human body parts, such as face and hand, with surprisingly higher accuracy and flexibility.



Detected skin pixels are replaced by dark points in images

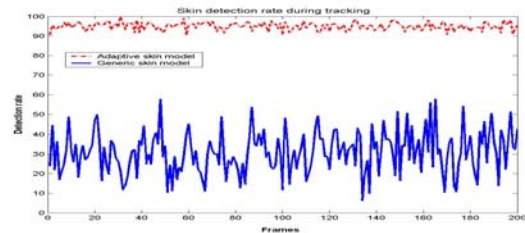


Fig. 7 Face tracking and skin detection rate curve in video

In Fig. 7, we use our adaptive skin model to track the human face in a video clip of 200 frames (downloaded from <http://csr.bu.edu/headtracking/in/>). In the top row, from left to right, we choose one frame to illustrate the original frame, another for the skin detection result by a generic model, and a third by our adaptive model. Furthermore, we manually labeled the skin pixels for the whole sequence to evaluate the skin detection rate for each frame. For the graph depicted below, at the same FP, an average of 96% DR is maintained for the two-step skin model, whereas the DR is only 20%–40% for the one-step generic model.



Original image Generic model Adaptive model

Fig. 8 Body segmentation in arbitrary still images

In Fig. 8, we further demonstrate two typical examples of segmenting the skin regions for arbitrary images: the top one containing people of different races and the bottom complicated by a confusing background. The obviously different results strongly demonstrate the advantage and effectiveness of our proposed two-step adaptive approach.

7. 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 correct 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 and hand gesture recognition, for the proposed adaptive skin model.

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. X. Zhu, J. Yang, and A. Waibel, "Segmenting hands of arbitrary color", Proc. IEEE Intl. Conf. on Automatic Face and Gesture Recognition (FG 2000), pp. 446-453, Mar. 2000, Grenoble, France.
3. C. Garcia and G. Tziritas, "Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis", IEEE Trans. on Multimedia, vol. 1(3): pp. 264-277, Sept. 1999.
4. 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.
5. 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.
6. L. Sigal, S. Sclaroff, and V. Athitsos, "Estimation and Prediction of Evolving Color Distributions for Skin Segmentation under Varying Illumination", Proc. IEEE Computer Vision and Pattern Recognition. (CVPR 2000), pp. 152-159, June 2000, CA, USA.
7. 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.
8. J. Wang, J. Li, G. Wiederhold and O. Firschein, "System for Screening Objectionable Images Using Daubechies' Wavelets and Color Histograms", Proc. Interactive Distributed Multimedia Systems (IDMS'97), Volume 1309, Springer-Verlag LNCS, pp. 20-30, Sept. 1997, Darmstadt, Germany.
9. M.J. Jones, J.M. Rehg, "Statistical Color Models with Application to Skin Detection", Technical Report, Cambridge Research Lab., 1998.
10. 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
11. M. Yang, and N. Ahuja, "Gaussian Mixture Model for Human Skin Color and Its application in Image and Video Database", Proc. of the SPIE: Conf. on Storage and Retrieval for Image and Video Databases (SPIE 99), pp. 458-466, Jan. 1999.
12. 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.
13. B. Li, E. Chang, and C.-S. Li. "Learning image query concepts via intelligent sampling". Proceedings of IEEE Multimedia and Expo, August 2001.
14. Burges, C. J. C. (1998). *A Tutorial on Support Vector Machines for Pattern Recognition*. Kluwer Academic Publishers, Boston.
15. 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>