HUMAN SKIN COLOR DETECTION IN RGB SPACE WITH BAYESIAN ESTIMATION OF BETA MIXTURE MODELS

Zhanyu Ma, Arne Leijon

Sound and Image Processing Lab
KTH - Royal Institute of Technology
Email: {zhanyu,leijon} @kth.se

ABSTRACT

Human skin color detection plays an important role in the applications of skin segmentation, face recognition, and tracking. To build a robust human skin color classifier is an essential step. This paper presents a classifier based on beta mixture models (BMM), which uses the pixel values in RGB space as the features. We propose a Bayesian estimation method based on the variational inference framework to approximate the posterior distribution of the parameters in the BMM and take the posterior mean as a point estimate of the parameters. The well-known Compaq image database is used to evaluate the performance of our BMM based classifier. Compared to some other skin color detection methods, our BMM based classifier shows a better recognition performance.

1. INTRODUCTION

Human skin color detection plays an important and effective role in the applications of skin segmentation, face recognition, and tracking problems [1]. Many heuristic and pattern recognition based methods have been proposed in the past decades. Different color spaces and methods have been evaluated and compared [2, 3, 4]. From the feature point of view, several color spaces could be used for skin detection such as Red-Green-Blue (RGB) [5, 2, 6], CIE-xy, YIQ, YCbCr [7], Tint-Saturation-Luminance (TSL) [8], and Hue-Saturation-Value/Intensity (HSV/HSI) [9]. From the standpoint of classification methods, the explicitly defined threshold (*e.g.* normalized R/G ratio [2]), non-parametric probabilistic model [2, 6], and parametric probabilistic model [6, 7] are the most efficient methods applied in the area of skin color detection.

As the RGB space is the mostly used space, we take the pixel value in RGB space as features. Since it has be shown [10, 11] that beta mixture models (BMM) can model data with compact range better than Gaussian mixture models and the pixel value is in $[0, 255]$, we apply BMM to model the skin pixel distribution in RGB space. With the principles of the variational inference (VI) framework [12, 13, 14, 15], we propose a Bayesian estimation algorithm to estimate the parameter distributions. By applying a set of non-linear approximations, the posterior distribution of the parameters in the BMM is obtained. The posterior mean is considered as the point estimate of the parameters.

The BMM-based classifier is trained and tested with the well-known Compaq image database [6]. The Receiver Operating Characteristic (ROC) [16] is used to evaluate the performance of the BMM classifier. This paper presents a method based on the VI framework for the Bayesian estimate of the parameters in the BMM and shows the skin/non-skin color detection results with the BMM classifier.

2. SKIN COLOR MODEL

In the past decades, different color models were proposed for various properties. The color models used nowadays are mostly oriented to applications. The RGB color space originated from the CRT display and describes the color in terms of three primary color channels: red (R), green (G), and blue (B). This is the mostly used space for image display and storage. Fig. 1 shows a 24-bit RGB color image with a human face and the corresponding (normalized) histogram for each color channel. The pixel values of each color channel are linearly compressed to $[0, 1]$ by $x_c := x_c/255,\ c \in \{r, g, b\}$.

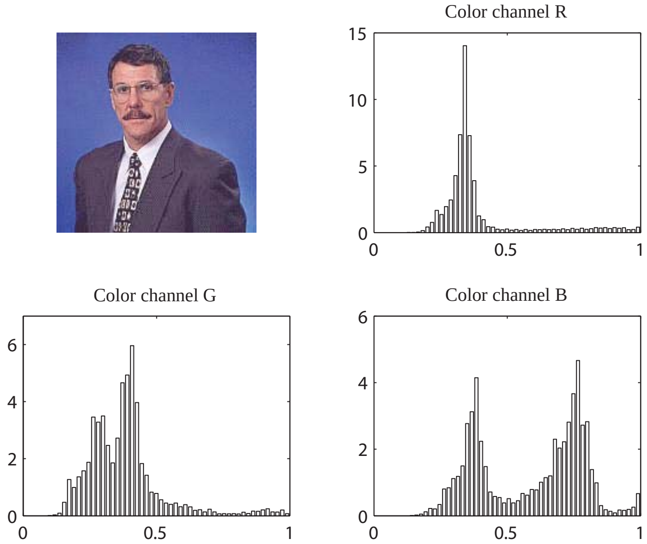

Fig. 1. Human face image and the corresponding skin histograms for RGB channels.

Although the RGB space is sensitive to the luminance, and the images of one object taken in different environments have diverse characteristics on the RGB space [2], the mixture of probabilistic models (*e.g.* skin probability map in [2, 6], Gaussian Mixture Models (GMM) in [6]) can still model the distribution of the pixel values in RGB space efficiently. By building the three-dimensional probabilistic skin and non-skin models in RGB space, the distribution of both skin and non-skin pixels can be represented in a probabilistic way. A pixel could be classified as a skin pixel or a non-skin pixel in an optimized way by utilizing the Bayesian classifier [17]. Several studies used parametric or non-parametric techniques to model the pixel value distribution. However, non-parametric techniques, such as histogram based models, need a large amount of training data and have high computational cost. With the parametric technique, we can obtain the parameters of the model, and this is more convenient in practical problems.

From Fig. 1 we can observe that the pixel values in the RGB space are in a compact range for each color channel (in the 24-bit RGB image, the range is $[0, 255]$ and can be linearly compressed to the range $[0, 1]$). Furthermore, the distribution of the skin color pixel is skewed in each channel. Indeed the conventional GMM can model any distribution shape with a proper number of components. Since it was shown that the BMM outperforms GMM with the same number of components in gray image classification [11], and the color image in RGB space can be considered as the composition of three gray images, one from each color channel, the BMM is a more reasonable choice to model the distribution of skin color. The correlations among the three channels can be represented by mixture models.

3. BETA MIXTURE MODELS

The beta distribution is a family of continuous probability distributions defined on the interval $[0, 1]$ with two positive real parameters. The probability density function of the beta distribution is

$$\mathrm{Beta}(x; u, v) = \frac{1}{\mathrm{beta}(u, v)} x^{u-1}(1-x)^{v-1} , \tag{1}$$

where $\mathrm{beta}(u, v)$ is the beta function

$$\mathrm{beta}(u, v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)} \tag{2}$$

and $\Gamma(\cdot)$ is the gamma function defined as $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$.

Multivariate data are in most cases statistically dependent. However, for any random vector $\mathbf{x}$ consisting of L elements, the dependencies among the elements $x_1, \ldots, x_L$ can be represented by a mixture model, even if each specific mixture component can only generate vectors with statistically independent elements. Therefore, we define the multivariate BMM as

$$\begin{aligned} f(\mathbf{x}; \mathbf{\Pi}, \mathbf{U}, \mathbf{V}) &= \sum_{i=1}^{I} \pi_i \mathbf{Beta}(\mathbf{x}; \mathbf{u}_i, \mathbf{v}_i) \\ \mathbf{Beta}(\mathbf{x}; \mathbf{u}_i, \mathbf{v}_i) &= \prod_{l=1}^{L} \mathrm{Beta}(x_l; u_{li}, v_{li}) , \end{aligned} \tag{3}$$

where $\mathbf{\Pi} = \{\pi_1, \ldots, \pi_I\}$, $\mathbf{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_I\}$, and $\mathbf{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_I\}$. $\{\mathbf{u}_i, \mathbf{v}_i\}$ denote the parameter vectors of the ith mixture component and u_{li}, v_{li} are the (scalar) parameters of the beta distribution for element x_l. For representing the skin color distribution, each observation $\mathbf{x}$ is a three-dimensional vector with $L = 3$ and each element x_l is in the range $[0, 1]$. In the following sections, we will uss observed pixel data $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ approximate the posterior distribution $f(\mathbf{U}, \mathbf{V}, \mathbf{\Pi}|\mathbf{X})$ via the variational inference (VI) framework in section 3.3 and the algorithm for Bayesian estimation will be listed in section 3.4.

3.1 Bayesian estimation and conjugate prior

We make a Bayesian estimate of the parameters in the BMM. An important step in the Bayesian estimation is to find the conjugate prior $f(\mathcal{Z})$ such that the posterior distribution $f(\mathcal{Z}|\mathcal{X})$ has the same form as $f(\mathcal{Z})$. The conjugate prior distribution to the beta distribution in (1) is

$$f(u, v) = \frac{1}{C} \left[\frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)}\right]^{\nu_0} e^{-\alpha_0(u-1)} e^{-\beta_0(v-1)} , \tag{4}$$

where α_0, β_0, ν_0 are free positive parameters and C is a normalization factor (a function of α_0, β_0, ν_0) such that

$$\int_0^\infty \int_0^\infty f(u, v) du dv = 1 .$$

Then we obtain the posterior distribution of u, v as (with N *i.i.d.* scalar observations $\mathbf{x} = \{x_1, \ldots, x_N\}$)

$$\begin{aligned} f(u, v|\mathbf{x}) &= \frac{f(\mathbf{x}|u, v) f(u, v)}{\int_0^\infty \int_0^\infty f(\mathbf{x}|u, v) f(u, v) du dv} \\ &= \frac{\left[\frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)}\right]^{\nu_N} e^{-\alpha_N(u-1)} e^{-\beta_N(v-1)}}{\int_0^\infty \int_0^\infty \left[\frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)}\right]^{\nu_N} e^{-\alpha_N(u-1)} e^{-\beta_N(v-1)} du dv} , \end{aligned} \tag{5}$$

where [1]

$$\begin{aligned} \alpha_N &= \alpha_0 - \sum_{n=1}^{N} \ln x_n , \\ \beta_N &= \beta_0 - \sum_{n=1}^{N} \ln(1 - x_n) , \\ \nu_N &= \nu_0 + N . \end{aligned}$$

3.2 Variational inference and factorized approximation

Analytically, we can not find a closed-form expression for the posterior distribution in (5) due to the computationally intractable integration expression in the denominator. Some stochastic techniques (*e.g.* Gibbs sampling [10]) can be used to calculate the posterior distribution numerically. We propose a method based on the VI framework [12] in this paper. According to the VI framework, the posterior density function of variable $\mathcal{Z}$ given the observation $\mathcal{X}$ (*i.e.* $f(\mathcal{Z}|\mathcal{X})$) is approximated by $g(\mathcal{Z})$. We decompose the log likelihood $\ln f(\mathcal{X})$ as

$$\ln f(\mathcal{X}) = \mathcal{L}(g) + KL(g \parallel f) , \tag{6}$$

where

$$\mathcal{L}(g) = \int g(\mathcal{Z}) \ln \frac{f(\mathcal{X}, \mathcal{Z})}{g(\mathcal{Z})} d\mathcal{Z} \tag{7}$$

and $KL(g \parallel f)$ is the Kullback-Leibler (KL) divergence defined as

$$KL(g \parallel f) = -\int g(\mathcal{Z}) \ln \frac{f(\mathcal{Z}|\mathcal{X})}{g(\mathcal{Z})} d\mathcal{Z} . \tag{8}$$

Since the KL divergence is a non-negative measurement, to maximize the lower bound $\mathcal{L}(g)$ is equivalent to minimize the KL divergence. Especially, when $g(\mathcal{Z})$ is equal to $f(\mathcal{Z}|\mathcal{X})$, the KL divergence vanishes and the lower bound reaches the true log likelihood $\ln f(\mathcal{X})$. If the target distribution is analytically intractable, some approximations can be used to achieve tractability with the factorized approximation (FA) method [12, 18].

The FA method partitions the variable $\mathcal{Z}$ into disjoint parts $\{\mathcal{Z}_m\}$, $m = 1, \ldots, M$ and decomposes the distribution as

$$g(\mathcal{Z}) = \prod_{m=1}^{M} g_m(\mathcal{Z}_m) . \tag{9}$$

Amongst all the distributions having the form in (9), we need to seek a distribution $g(\mathcal{Z})$ that drives the lower bound $\mathcal{L}(g)$ to be largest. By substituting (9) into (7) and denoting $g_m(\mathcal{Z}_m)$ by g_m simply, we obtain

$$\begin{aligned} \mathcal{L}(g) &= \int \prod_{m=1}^{M} g_m \left\{ \ln f(\mathcal{X}, \mathcal{Z}) - \sum_{m=1}^{M} \ln g_m \right\} d\mathcal{Z} \\ &= \int g_n \left\{ \int \ln f(\mathcal{X}, \mathcal{Z}) \prod_{m \neq n} g_m d\mathcal{Z}_m \right\} d\mathcal{Z}_n \\ &\quad - \int g_n \ln g_n d\mathcal{Z}_n + \text{const.} \\ &= \int g_n \ln \widetilde{f}(\mathcal{X}, \mathcal{Z}_n) d\mathcal{Z}_n - \int g_n \ln g_n d\mathcal{Z}_n + \text{const.} , \end{aligned} \tag{10}$$

[1] To prevent the infinity quantity in the practical implementation, we assign ε_1 to x_n when $x_n = 0$ and $1 - \varepsilon_2$ to x_n when $x_n = 1$. Both ε_1 and ε_2 are slightly positive real numbers.

where

$$\ln \widetilde{f}(\mathcal{X}, \mathcal{Z}_n) = \mathbf{E}_{*\neq \mathcal{Z}_n} [\ln f(\mathcal{X}, \mathcal{Z})] + \text{const.} \\ = \int \ln f(\mathcal{X}, \mathcal{Z}) \prod_{m\neq n} g_m d\mathcal{Z}_m + \text{const.} \, . \tag{11}$$

By recognizing that the first two integrals in the final line of (10) is a negative KL divergence between $g_n(\mathcal{Z}_n)$ and $\widetilde{f}(\mathcal{X}, \mathcal{Z}_n)$, we can maximize $\mathcal{L}(g)$ with respect to any possible form of $q_n(\mathcal{Z}_n)$, while keeping $g_{m\neq n}(\mathcal{Z}_m)$ fixed, by minimizing the KL divergence. The optimal value occurs when $g_n(\mathcal{Z}_n) = \widetilde{f}(X, \mathcal{Z}_n)$, which gives us the optimal solution to $g_n(\mathcal{Z}_n)$ as

$$\ln g_n^*(\mathcal{Z}_n) = \mathbf{E}_{*\neq \mathcal{Z}_n} [\ln f(\mathcal{X}, \mathcal{Z})] \; + \text{const.} \, . \tag{12}$$

3.3 Factorized approximation for BMM

The prior of the beta distribution is analytically intractable. With the principles of FA, the conjugate prior in (4) can be approximated as

$$f(u, v) \approx f(u)f(v). \tag{13}$$

Since both u and v are nonnegative variables, we assign the Gamma distribution to u, v respectively as

$$f(u; \mu, \alpha) = \frac{\alpha^\mu}{\Gamma(\mu)} u^{\mu-1} e^{-\alpha u}, \;\; \alpha, \mu \in \mathbf{R}^+ \\ f(v; \nu, \beta) = \frac{\beta^\nu}{\Gamma(\nu)} v^{\nu-1} e^{-\beta v}, \;\; \beta, \nu \in \mathbf{R}^+ \, . \tag{14}$$

Furthermore, by introducing the Dirichlet distribution as the prior distribution of the mixing coefficients, the probability density function of $\mathbf{\Pi}$ can be written as

$$f(\mathbf{\Pi}) = Dir(\mathbf{\Pi}|\mathbf{c}) = C(\mathbf{c}) \prod_{i=1}^{I} \pi_i^{c_i - 1} \, , \tag{15}$$

where $C(\mathbf{c}) = \frac{\Gamma(\widehat{c})}{\Gamma(c_1)\cdots\Gamma(c_I)}, \;\; \widehat{c} = \sum_{i=1}^{I} c_i$.

For each observation $\mathbf{x}_n$, the corresponding $\mathbf{z}_n = (z_{n1}, \ldots, z_{nI})^T$ is an indication vector with one element equals to 1 and the rest equal to 0, where $z_{ni} = 1$ means that the nth observation is generated from the ith component in the BMM. The conditional distribution of $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ and $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ given latent variables $\{\mathbf{U}, \mathbf{V}, \mathbf{\Pi}\}$ is

$$\begin{aligned} & f(\mathbf{X}, \mathbf{Z}|\mathbf{U}, \mathbf{V}, \mathbf{\Pi}) \\ =& f(\mathbf{X}|\mathbf{U}, \mathbf{V}, \mathbf{\Pi}, \mathbf{Z}) f(\mathbf{Z}|\mathbf{\Pi}) \\ =& f(\mathbf{X}|\mathbf{U}, \mathbf{V}, \mathbf{Z}) f(\mathbf{Z}|\mathbf{\Pi}) \\ =& \prod_{n=1}^{N} \prod_{i=1}^{I} [\pi_i \mathbf{Beta}(\mathbf{x}_n|\mathbf{u}_i, \mathbf{v}_i)]^{z_{ni}} \, . \end{aligned} \tag{16}$$

The logarithm of the joint distribution function of $\mathcal{X} = \{\mathbf{X}\}$ and $\mathcal{Z} = \{\mathbf{U}, \mathbf{V}, \mathbf{\Pi}, \mathbf{Z}\}$ is given by

$$\begin{aligned} & \mathcal{L}(\mathcal{X}, \mathcal{Z}) = \ln f(\mathcal{X}, \mathcal{Z}) \\ =& \sum_{n=1}^{N} \sum_{i=1}^{I} z_{ni} \left\{ \ln \pi_i + \sum_{l=1}^{L} \ln \frac{\Gamma(u_{li} + v_{li})}{\Gamma(u_{li})\Gamma(v_{li})} \right. \\ & \left. + \sum_{l=1}^{L} [(u_{li} - 1) \ln x_{ln} + (v_{li} - 1) \ln(1 - x_{ln})] \right\} \\ & + \sum_{l=1}^{L} \sum_{i=1}^{I} [(\mu_{li} - 1) \ln u_{li} - \alpha_{li} u_{li}] \\ & + \sum_{l=1}^{L} \sum_{i=1}^{I} [(\nu_{li} - 1) \ln v_{li} - \beta_{li} v_{li}] + \sum_{i=1}^{I} (c_i - 1) \ln \pi_i + \text{const.} \, . \end{aligned} \tag{17}$$

3.4 Algorithm of Bayesian estimation

The latent variables we have now are $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Pi}$ with the hyper-parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\nu}$, and $\mathbf{c}$. The optimal distribution for $\mathbf{U}$ and $\mathbf{V}$ are obtained by taking the expected value of $\mathcal{L}(\mathcal{X}, \mathcal{Z})$ as (element-wise)

$$\ln f^*(u_{li}; \mu_{li}, \alpha_{li}) = \mathbf{E}_{*\neq u_{li}} [\mathcal{L}(\mathcal{X}, \mathcal{Z})] \\ \ln f^*(v_{li}; \nu_{li}, \beta_{li}) = \mathbf{E}_{*\neq v_{li}} [\mathcal{L}(\mathcal{X}, \mathcal{Z})] \; . \tag{18}$$

Obviously, the expectations in the RHS of (18) could not lead to a closed-form expression. The second-order Taylor expansion of $\ln \frac{\Gamma(u_{li}+v_{li})}{\Gamma(u_{li})\Gamma(v_{li})}$ in terms of $(\ln u_{li}, \ln v_{li})$ can be proven to be a lower bound [19]. The expectation of this lower bound can yield the optimal solution to $\mathbf{U}, \mathbf{V}$ asymptotically. The update equations for the hyper-parameters of $\mathbf{U}, \mathbf{V}$, and $\mathbf{\Pi}$ are listed as follows (element-wise):

$$\begin{aligned} c_i^* =& c_{i_0} + \sum_{n=1}^{N} \mathbf{E}[z_{ni}] \\ \mu_{li}^* =& \mu_{li_0} + \sum_{n=1}^{N} \mathbf{E}[z_{ni}] \overline{u}_{li} \{\psi(\overline{u}_{li} + \overline{v}_{li}) - \psi(\overline{u}_{li}) \\ & + \overline{v}_{li} \cdot \psi^{'}(\overline{u}_{li} + \overline{v}_{li})(\mathbf{E}_v [\ln v_{li}] - \ln \overline{v}_{li}) \Big\} \\ \alpha_{li}^* =& \alpha_{li_0} - \sum_{n=1}^{N} \mathbf{E}[z_{ni}] \ln x_{ln} \\ \nu_{li}^* =& \nu_{li_0} + \sum_{n=1}^{N} \mathbf{E}[z_{ni}] \overline{v}_{li} \{\psi(\overline{u}_{li} + \overline{v}_{li}) - \psi(\overline{v}_{li}) \\ & + \overline{u}_{li} \cdot \psi^{'}(\overline{u}_{li} + \overline{v}_{li})(\mathbf{E}_u [\ln u_{li}] - \ln \overline{u}_{li}) \Big\} \\ \beta_{li}^* =& \beta_{li_0} - \sum_{n=1}^{N} \mathbf{E}[z_{ni}] \ln(1 - x_{ln}) \, , \end{aligned} \tag{19}$$

where (the estimation of ρ_{ni} is in (20) in the next page)

$$\begin{aligned} \overline{u} =& \frac{\mu}{\alpha}, \; \overline{v} = \frac{\nu}{\beta} \\ \mathbf{E}[z_{ni}] =& \frac{\rho_{ni}}{\sum_{k=1}^{I} \rho_{nk}} \\ \mathbf{E}_u[\ln u] =& \psi(\mu) - \ln \alpha \\ \mathbf{E}_v[\ln v] =& \psi(\nu) - \ln \beta \\ \mathbf{E}_u \left[(\ln u - \ln \overline{u})^2\right] =& [\psi(\mu) - (\ln \mu)]^2 + \psi^{'}(\mu) \\ \mathbf{E}_v \left[(\ln v - \ln \overline{v})^2\right] =& [\psi(\nu) - (\ln \nu)]^2 + \psi^{'}(\nu) \, . \end{aligned}$$

To start the iterations, the values of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\nu}$ are chosen such that the prior distributions are assigned with non-informative distributions (flat broad distribution). The parameters for the Dirichlet distribution ($c_i, \; i = 1, \ldots, I$) are assigned with a small value (*i.e.*, 0.001) to ensure the number of mixture components is controlled by the data. By updating the hyper-parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\nu}$, and $\mathbf{c}$ recursively in order, the algorithm will converge so that the KL divergence in (8) is almost equal to 0. Compared to the conventional expectation maximization (EM) based maximum likelihood estimation (MLE), this Bayesian estimation can prevent overfitting and estimate the effective number of mixture components automatically. The latent variables are all unimodally distributed and the posterior distributions are highly peaked. Considering the posterior mean as the point estimate of u_{li}, v_{li}, we take $\widehat{u}_{li} = \overline{u}_{li} = \mu_{li}^*/\alpha_{li}^*$ and $\widehat{v}_{li} = \overline{v}_{li} = \nu_{li}^*/\beta_{li}^*$. More details about the derivations of this algorithm can be found in [19].

$$
\begin{aligned}
\ln\rho_{ni} \approx\ & \mathbf{E}\left[\ln \pi_i\right] + \sum_{l=1}^{L}\left[(\overline{u}_{li}-1)\ln x_{ln} + (\overline{v}_{li}-1)\ln(1-x_{ln})\right] \\
& + \sum_{l=1}^{L}\Bigg\{\ln\frac{\Gamma(\overline{u}_{li}+\overline{v}_{li})}{\Gamma(\overline{u}_{li})\Gamma(\overline{v}_{li})} + \overline{u}_{li}\left[\psi(\overline{u}_{li}+\overline{v}_{li})-\psi(\overline{u}_{li})\right]\left(\mathbf{E}\left[\ln u_{li}\right]-\ln\overline{u}_{li}\right) + \overline{v}_{li}\left[\psi(\overline{u}_{li}+\overline{v}_{li})-\psi(\overline{v}_{li})\right]\left(\mathbf{E}\left[\ln v_{li}\right]-\ln\overline{v}_{li}\right) \\
& + 0.5\cdot\overline{u}_{li}^2\left[\psi^{'}(\overline{u}_{li}+\overline{v}_{li})-\psi^{'}(\overline{u}_{li})\right]\mathbf{E}\left[(\ln u_{li}-\ln\overline{u}_{li})^2\right] + 0.5\cdot\overline{v}_{li}^2\left[\psi^{'}(\overline{u}_{li}+\overline{v}_{li})-\psi^{'}(\overline{v}_{li})\right]\mathbf{E}\left[(\ln v_{li}-\ln\overline{v}_{li})^2\right] \\
& + \overline{u}_{li}\cdot\overline{v}_{li}\cdot\psi^{'}(\overline{u}_{li}+\overline{v}_{li})(\mathbf{E}\left[\ln u_{li}\right]-\ln\overline{u}_{li})(\mathbf{E}\left[\ln v_{li}\right]-\ln\overline{v}_{li})\Bigg\}
\end{aligned}
\tag{20}
$$

Fig. 2. Original images and the corresponding skin detection results.

4. SKIN COLOR/NON-SKIN COLOR DETECTION

The skin/non-skin Bayesian classifier [17] is applied here for detection. With the skin color pixels and non-skin color pixels, we train two BMMs, one for each kind of pixels. Given a new pixel $\mathbf{x}$, the decision rule is

$$
\begin{cases}
\text{If } \frac{f(\mathbf{x}|\mathbf{s})}{f(\mathbf{x}|\sim\mathbf{s})} > \lambda\cdot\frac{f(\sim\mathbf{s})}{f(\mathbf{s})} & \mathbf{x}\in\mathbf{s} \\
\text{else if } \frac{f(\mathbf{x}|\mathbf{s})}{f(\mathbf{x}|\sim\mathbf{s})} < \lambda\cdot\frac{f(\sim\mathbf{s})}{f(\mathbf{s})} & \mathbf{x}\in\sim\mathbf{s} \\
\text{otherwise} & \text{arbitrary decision}
\end{cases}, \tag{21}
$$

where $\mathbf{s}$ and $\sim\mathbf{s}$ denote skin and non-skin respectively. $f(\mathbf{s})$ and $f(\sim\mathbf{s})$ are the prior skin color and non-skin color probabilities. λ is a threshold introduced to adjust the trade-off between the two kinds of decision errors. It is analyzed empirically by experiments. The conditional likelihood of $\mathbf{x}$ is calculated as

$$
f(\mathbf{x}|c) = \sum_{i=1}^{I}\pi_i\prod_{l=1}^{3}\mathrm{Beta}(x_l; u_{li}^c, v_{li}^c),\ c\in\{\mathbf{s},\sim\mathbf{s}\}. \tag{22}
$$

5. EXPERIMENTAL RESULTS AND DISCUSSION

We applied our BMM classifier to the well-known Compaq image database [6]. This database contains around 4600 skin images and around 9000 non-skin images. For each skin color image, a corresponding mask is available in the database and used to separate skin/non-skin areas. All the images in the database are color images obtained from the World Wide Web. Different ethnic people's skin colors are represented in the skin images. Also, the skin images in the database were taken under different angles, positions and brightness conditions.

Table 1. Comparison of different methods for skin color detection

Method	True Positive	False Positive
Implicit mathematical model [5]	83.3% 90.7%	15.6% 13.3%
Thresholding [2]	94.7%	30.2%
Bayes SPM in RGB [2]	93.4%	19.8%
SOM in TS [8]	78%	32%
Maximum entropy in RGB [20]	80%	8%
Bayes SPM in RGB [6]	80% 90%	8.5% 14.2%
GMM in RGB [6]	80% 90%	9.5% 15.5%
Elliptical boundary in CIE-xy [7]	90%	20.9%
Single Gaussian in YCbCr [7]	90%	33.3%
GMM in IQ [7]	90%	30%
Our method with BMM in RGB	**80% 90% 95%**	**7.2% 11.8% 19.7%**

For each evaluation round, the Compaq database was partitioned randomly into a training sub-database and a test sub-database. Each sub-database consists of skin and non-skin images. We randomly selected a training set from the training sub-database with $500,000$ skin color pixels and $1,500,000$ non-skin color pixels. Also, a test set with the same size was randomly drawn from the test sub-database. These pixels were selected randomly and labelled. For the training procedure, the labelled pixels from the training set were used to train the skin color model and non-skin color model respectively. Then the obtained models were applied to classify the pixels in the test set to skin or non-skin categories by the rules in (21). Both the skin color and non-skin color were labelled so that we can calculate the correct decision of classifying the skin pixels into skin (True Positive Rate (TPR)) and the false decision of classifying non-skin pixels into skin (False Positive Rate (FPR)) by comparing the labels and the decisions the model made. Fig. 2 shows some detection results from our model. The overall performance is good and both the skin area of the black lady and the white couples were detected. The missing part on the man's left arm and misclassified part of the hair in a relative brighter background indicates that the illuminance of the image has influence on the detection result. Also, the misclassified part in the background could possibly be adjusted by some texture based methods.

To evaluate the performance of our BMM classifier, we apply the ROC analysis [16]. In the ROC curve, the TPR is plotted on the vertical axis and the FPR is plotted on the horizontal axis. Any point in the ROC curve indicates a better performance if the point is closer to the northwest in the coordinates (the best possible point is TPR = 100% and FPR = 0%). By changing the value of λ in (21), different TPR-FPR pairs are obtained. The ROC curve consists of these TPR-FPR pairs and shows the trade-off between the TPR and

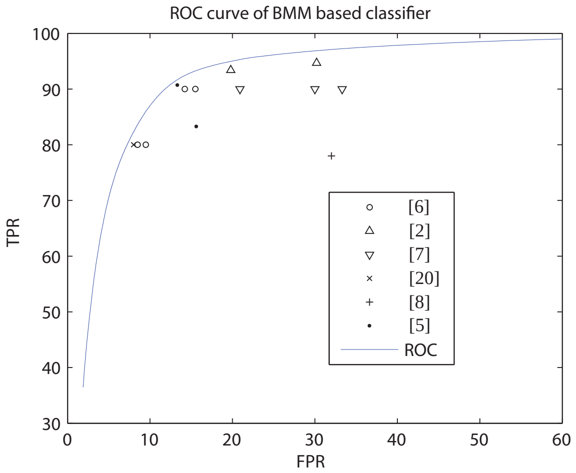

Fig. 3. Recognition score comparison of different methods.

FPR. Our BMM classifier reported 80% TPR with 7.2% FPR and 90% TPR with 11.8% FPR. The accuracy rate (the total number of correct decision out of the total input) of the classifier is 88.9% when the TPR is equal to one minus the FPR. To prevent the effect of randomness, we executed 60 rounds of the train-test procedures mentioned above. The mean values of the TPR, the FPR and the accuracy rate are reported.

Some other classifiers based on the pixel probabilistic model were also analyzed with the Compaq database in the previous literature [5, 2, 8, 20, 6, 7]. They reported different classification scores with different methods and with different color spaces (*e.g.* RGB, YCbCr). The best classification scores from some previous studies and our BMM classifier are listed in table 1. As mentioned in [21], if the transformation from one color space to another is invertible and provided the optimal skin classifier for the color space is used, the differences of the color space does not have influence on the classifier's performance. Although different methods used different separations of the database and employed different learning strategies, it is still interesting to compare our ROC curve with the results in table 1. In Fig. 3, all the results listed in table 1 are in the southeast side of the ROC curve, which means that our BMM classifier outperforms all the other methods. For some other tasks of classification with data in a compact range, the BMM is probably also a promising model.

6. CONCLUSION

This paper presented a BMM-based classifier for the task of human skin/non-skin color detection. A Bayesian estimation algorithm for the parameters was proposed. With the variational inference framework and a set of non-linear approximations, the posterior distributions for the BMM parameters were approximated and the posterior mean was used as the point estimate of the parameters.

The BMM classifier was applied to the well-known Compaq image database, using the pixel values in the RGB color space as the features. The overall detection performance is good. In comparison with other methods based on pixel probabilistic models, our BMM classifier outperforms the previous results.

REFERENCES

[1] M. H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

[2] J. Brand and J. S. Mason, "A comparative assessment of three approaches to pixel-level human skin-detection," in *Proceedings of IEEE International Conference on Pattern Recognition*, 2000, vol. 1, pp. 1056–1059 vol.1.

[3] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," in *Proceedings of Graphicon*, 2003, pp. 85–92.

[4] S.L. Phung, Sr. Bouzerdoum, A., and Sr. Chai, D., "Skin segmentation using color pixel classification: analysis and comparison," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 1, pp. 148 –154, jan. 2005.

[5] M. M. Aznaveh, H. Mirzaei, E. Roshan, and M. Saraee, "A new and improved skin detection method using RGB vector space," in *Proceedings of IEEE International Multi-Conference on Systems, Signals and Devices*, July 2008, pp. 1–5.

[6] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *International Journal of Computer Vision*, vol. 46, no. 1, pp. 81–96, 2002.

[7] J. Y. Lee and S. I. Yoo, "An elliptical boundary model for skin color detection," in *Proceedings of the International Conference on Imaging Science, Systems, and Technology*, 2002.

[8] D. A. Brown, I. Craw, and J. Lewthwaite, "A SOM based approach to skin detection with application in real time systems," in *Proceedings of the British Machine Vision Conference*, 2001.

[9] R. Kjeldsen and J. Kender, "Finding skin in color images," *Automatic Face and Gesture Recognition, IEEE International Conference on*, vol. 0, pp. 312, 1996.

[10] N. Bouguila, D. Ziou, and E. Monga, "Practical bayesian estimation of a finite beta mixture through gibbs sampling and its applications," *Statistics and Computing*, vol. 16, pp. 215–225, 2006.

[11] Z. Ma and A. Leijon, "Beta mixture models and the application to image classification," in *Proceedings of IEEE International Conference on Image Processing*, 2009.

[12] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[13] T. S. Jaakkola, "Tutorial on variational approximation methods," in *Advances in Mean Field Methods.*, M. Opper and D. saad, Eds., pp. 129–159. MIT Press., 2001.

[14] T. S. Jaakkola and M. I. Jordan, "Bayesian parameter estimation via variational methods," *Statistics and Computing*, vol. 10, pp. 25–37, 2000.

[15] N. Ueda and Z. Ghahramani, "Bayesian model search for mixture models based on optimizing variational bounds," *Neural Network*, vol. 15, pp. 1223–1241, 2002.

[16] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letter*, vol. 27, no. 8, pp. 861–874, 2006.

[17] D. Chai, S. L. Phung, and A. Bouzerdoum, "A Bayesian skin/non-skin color classifier using non-parametric density estimation," in *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, May 2003, vol. 2, pp. II–464–II–467 vol.2.

[18] G. Parisi, *Statistical Field Theory*, Addison-Wesley, 1988.

[19] Z. Ma and A. Leijon, "Bayesian estimation of beta mixture models with variational inference," *Submitted.*, 2010, available at `http://www.kth.se/ees/forskning/publikationer/modules/publications_polopoly/reports/2010/IR-EE-SIP_2010_010.pdf?l=en_UK`.

[20] B. Jedynak, H. Zheng, M. Daoudi, and D. Barret, "Maximum entropy models for skin detection," in *Proceedings of Indian Conference on Computer Vision, Graphics and Image Processing*, 2002, pp. 276–281.

[21] A. Albiol, L. Torres, and E. J. Delp, "Optimum color spaces for skin detection," in *Proceedings of International Conference on Image Processing*, 2001.