

# Skin Detection and Segmentation in Color Images

Michał Kawulok, Jakub Nalepa and Jolanta Kawulok

**Abstract** This chapter presents an overview of existing methods for human skin detection and segmentation. First of all, the skin color modeling schemes are outlined, and their limitations are discussed based on the presented experimental study. Then, we explain the techniques which were reported helpful in improving the efficacy of color-based classification, namely (1) textural features extraction, (2) model adaptation schemes, and (3) spatial analysis of the skin blobs. The chapter presents meaningful qualitative and quantitative results obtained during our study, which demonstrate the benefits of exploiting particular techniques for improving the skin detection outcome.

**Keywords** Skin detection · Skin segmentation · Skin color models · Adaptive skin modeling · Face detection and tracking · Hand detection and tracking

## 1 Introduction

Skin detection and segmentation is a challenging problem in color image processing that has been extensively studied over the years. In general, the existing techniques are based on the premise that the skin color can be effectively modeled in various color spaces, which in turn allows for segmenting the skin regions in a given color image. Applications of skin detection are of a wide range and significance, including:

---

M. Kawulok (✉) · J. Nalepa · J. Kawulok  
Institute of Informatics, Silesian University of Technology, Gliwice, Poland  
e-mail: Michal.Kawulok@polsl.pl

J. Nalepa  
e-mail: Jakub.Nalepa@polsl.pl

J. Kawulok  
e-mail: Jolanta.Kawulok@polsl.pl

1. Hand and face detection and tracking for gesture recognition and human-computer interaction [10, 33, 43].
2. Objectionable content filtering for the sake of blocking nude images and videos [59, 101].
3. Feature extraction for content-based image retrieval [56].
4. Image coding using regions of interest [1, 14, 17, 26], and many more.

First of all, we would like to clearly define the terms of *skin detection* and *skin region segmentation*, as they are often used interchangeably which leads to certain confusion. *Skin detection* is a process whose aim is to determine whether a given video sequence, an image, a region or a pixel presents human skin. In most cases, it is performed at a pixel level first—every pixel is classified independently based on its chrominance and other low-level features extracted from its neighborhood. Afterwards, the decision may be verified at a higher level, and for example if it occurs that the pixels classified as skin are sparse in a given image, the detection outcome would be negative for that very image. Skin detection is the preliminary step to *skin region segmentation*, whose aim is to determine the boundaries of skin regions. In the simplest approach, this may not involve any additional processing over the pixel-wise skin detection, based on the assumption that the skin regions are formed by the adjacent pixels classified as skin. However, it has been reported in many works that using more advanced methods the segmentation precision may be definitely improved. Segmenting skin regions is helpful for extracting shapes of hands, faces and other body parts, but it may also be used as a verification step which improves the efficacy of skin detection measured at the pixel level.

This chapter contains a detailed overview of the skin color modeling algorithms, as well as it presents the methods which reduce the skin detection errors. General overview of existing skin detectors is given in Sect. 2, and structure of the chapter is presented there as well. The discussed methods are compared both on the theoretical and experimental basis. Their performance is evaluated quantitatively and qualitatively using images from the benchmark data sets. The evaluation procedure is discussed in Sect. 3.

## 2 General Categories of Skin Detection Methods

A conventional approach towards solving the skin detection problem consists in defining a skin color model, using which every individual pixel can later be classified based on its position in the color space, independently from its neighbors. This research direction has been widely explored and plenty of different pixel-wise methods were proposed that operate in most of existing color spaces. These techniques are discussed in Sect. 4. Skin color can be modeled either as a set of rules and thresholds defined in color spaces based on the experimental study (Sect. 4.2), or it can be trained using machine learning (Sect. 4.3). There were several good reviews published on skin color modeling. In 2003, V. Vezhnevets et al. presented the first thorough survey on skin detection [87]. Later, in 2005, several rule-based methods

were compared by F. Gasparini et al. [29]. In the same year, S.L. Phung et al. presented a comparison, mainly focused on statistical color modeling [64]. The most recent interesting survey on skin detection was published in 2007 by P. Kakumanu et al. [41].

It is worth noting that while color-based methods present different characteristics and therefore they can be found suitable in specific conditions, their effectiveness is limited due to high variance of skin color and its low specificity. Skin color depends on such individual factors like race, age or complexion. Intra-personal differences may also be substantial because of variations in lighting conditions or individual's physical state. Moreover, background objects often have skin-like color, which results in observing false positive errors in the segmentation outcome.

Despite the aforementioned shortcomings, it must be stated that the color is the basic feature for detecting skin regions. However, its discriminating power is highly limited, and by taking advantage of additional data sources, the errors can be significantly reduced. There are some effective approaches towards improving the performance of the pixel-wise color-based detectors, which we have grouped into three main categories, namely:

- A. Texture-based models (Sect. 5). The textural features may improve the stability of the color-based skin models by rejecting the regions which are not smooth. According to our observation, simple textural features are most useful here.
- B. Model adaptation techniques (Sect. 6). Skin color models may be adapted to a particular scene based on a whole-image analysis (Sect. 6.1), tracking (Sect. 6.2), or from a given skin sample (Sect. 6.3). A skin sample, i.e. a region which contains a representative set of skin pixels, can be either acquired automatically using face and hand detectors or be marked manually by an operator. This makes the skin color model more specific, which boosts the detection rate and decreases the false positives.
- C. Spatial analysis (Sect. 7). Spatial analysis consists in segmenting the skin blobs which may help decrease the detection errors by rejecting isolated false positive pixels.

It is worth noting that all of the methods discussed in this chapter require color-based skin modeling, which constitutes the primary source of information for skin detection. While skin color modeling has been widely addressed in the literature, these directions were paid much less attention. This may be somehow surprising, given how much they may contribute to increasing the robustness of skin detectors.

## 3 Evaluation Procedure

### 3.1 Evaluation Metrics

Skin detection performance is measured based on the number of correctly (i.e.  $TP$ —*true positives* and  $TN$ —*true negatives*) and incorrectly classified pixels (i.e.

*FN—false negatives* and *FP—false positives*). Based on these values, the following ratios can be computed:

- A. *False positive rate*:  $\delta_{fp} = FP/(FP + TN)$ , i.e. the percentage of background pixels misclassified as skin [40].
- B. *False negative rate*:  $\delta_{fn} = FN/(FN + TP)$ , i.e. the percentage of skin pixels misclassified as background [40].
- C. *Recall*, also referred to as *correct detection rate* or *true positive rate*:  $\eta_{tp} = TP/(FN + TP) = 1 - \delta_{fn}$ , i.e. the percentage of skin pixels correctly classified as skin [29].
- D. *Precision*:  $\eta_{prec} = TP/(TP + FP)$ , i.e. the percentage of correctly classified pixels out of all the pixels classified as skin [29].
- E. *F-measure*: the harmonic mean of precision and recall [52].

If the classification is non-binary, i.e. for every pixel the skin probability is computed, then the false positive and false negative rates depend on the acceptance threshold. The higher the threshold is, the less false positives are reported, but also the false negatives increase. Mutual relation of these two errors is often presented using *receiver operating characteristics (ROC)* [64], and the *area under curve* can also be used as the effectiveness determinant [69].

In our works we also investigated the false negative rate ( $\delta_{fn}^{(\eta)}$ ) obtained for a fixed false positive error  $\delta_{fp} = \eta$  [46]. Furthermore, we used the *minimal detection error* ( $\delta_{min} = (\delta_{fp} + \delta_{fn})/2$ ), where the threshold is set to a value, for which this sum is the smallest. In this chapter we rely on false positive and false negative rate, and we present their dependence using *ROC* curves, when it is applicable.

### 3.2 Data Sets

In their overview on skin color modeling, S.L. Phung et al. introduced a new benchmark data set, namely *ECU face and skin detection database* [64]. This data set consists of 4000 color images and ground-truth data, in which skin areas are annotated. The images were gathered from the Internet to provide appropriate diversity, and results obtained for this database were often reported in many works on skin detection. Therefore, this set was also used for all the comparisons presented in this chapter. Some examples of images from the data set are shown in Fig. 1. In all experiments reported here, the database was split into two equinumerous parts. The first 2000 images are used for training (*ECU-T*), and the remaining 2000 images are used for validation (*ECU-V*). If an algorithm does not require training, then only the *ECU-V* set is used. The data set contains some images acquired in the same conditions, but as they appear close to each other in the data set, there is no risk that two similar images will repeat in *ECU-T* and *ECU-V*.

Among other data sets which can be used for evaluating skin detection, the following may be mentioned: (1) M.J. Jones and J.M. Rehg introduced the Compaq database [39], (2) S.J. Schmugge et al. composed a data set based on images derived

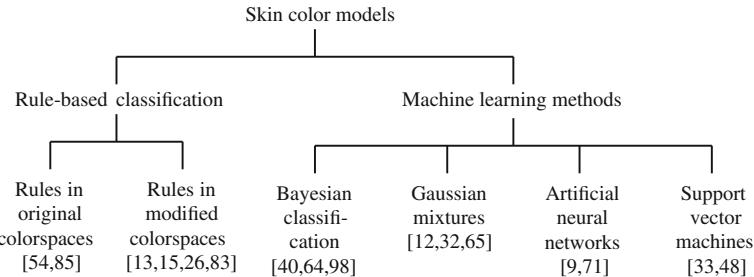


**Fig. 1** Examples of images from *ECU face and skin detection database* [64]

from existing databases, in which they annotated skin regions [69], (3) we have created our hand image database for gesture recognition purposes (available at <http://sun.aei.polsl.pl/~mkawulok/gestures>) [46].

## 4 Skin Color Modeling

Classification of skin color modeling methods, which are given more attention in this section, is presented in Fig. 2, and several most important references are given for every category. In general, the decision rules can be defined explicitly in commonly used color spaces or in a modified color space, in which the skin could be easier separated from the background. Machine learning methods require a training set, from which the decision rules are learned. A number of learning schemes have been used for this purpose, and the most important categories are presented in the figure and outlined in the section.



**Fig. 2** General categories of skin color models

#### 4.1 Color Spaces and Color Normalization

Skin color has been modeled in many color spaces using various techniques, which is summarized in Table 1. In many works it is proposed to reject the luminance component to achieve invariance to illuminance conditions. Hence, those color spaces are then preferred, in which the luminance is separated from the chrominance components, e.g.  $YC_bC_r$ ,  $YUV$ ,  $HSV$ . However, it was reported by many researchers that the illumination plays an important role in modeling the skin color and should not be excluded from the model [65, 72, 90]. The problem of determining an optimal color space for skin detection was addressed by A. Albiol et al., who provided a theoretical

**Table 1** Color spaces used for skin color modeling

Color space	Skin color models
$RGB$	Shin [72], Kovac [54, 76], Brand [11], Jones [40], Choi [17], Bhoyer [9], Seow [71], Taqa [82], Han [33], Ng [62], Jiang [38], Conci [19]
$YC_bC_r$	Hsu [36], Phung [65], Hossain [35], Kawulok [48]
$rg$	Stoerring [79], Greenspan [32], Caetano [12]
$HSI$	Schmugge [69], Jagadesh [37, 67]
$HSV$	Sobottka [74], Tsekeridou [85]
$CIELUV$	Yang [92]
$YIQ$	Duan [23]
$YUV$	Zafarifar [97]
Multiple color spaces	Kukharev [57], Wang [89], Abin [3], Fotouhi [27]

proof [5] that in every color space optimal skin detection rules can be defined. Following their argumentation, small differences in performance are attributed exclusively to the quantization of a color space. This conclusion is certainly correct, however the simplicity of the skin model may depend on the selection of color space. Based on the scatter analysis, as well as 2D and 3D skin-tone histograms, M.C. Shin et al. reported that it is the *RGB* color space which provides the best separability between skin and non-skin color [72]. Furthermore, their study confirmed that the illumination is crucial for increasing the separability between the skin and non-skin pixels. In their later works, they argued that the *HSI* color space should be chosen when the skin color is modeled based on the histogram analysis [69].

Color normalization plays an important role in skin color modeling [8, 53, 86]. M. Stoerring et al. investigated skin color appearance under different lighting conditions [79]. They have observed that location of the skin locus in the normalized *rg* color space, where  $r = R/(R + G + B)$  and  $g = G/(R + G + B)$ , depends on the color temperature of the light source. Negative influence of the changing lighting can be mitigated by appropriate color normalization. There exist many general techniques for color normalization [24, 25, 30, 58, 68], which can be applied prior to skin detection. Among them, the gray world transform is often reported as quite effective, while simple technique. Given an image with sufficient amount of color variations, the mean value of the *R*, *G*, and *B* channels should average to a common gray value, which equals 128 in the case of the *RGB* color space. In order to achieve this goal, each channel is scaled linearly:

$$c_N = c \cdot 128/\mu_c, \quad (1)$$

where  $c$  indicates the channel (i.e. *R*, *G* or *B*),  $\mu_c$  is the mean value in the channel for a given image prior to normalization, and  $c_N$  is the value after normalization. In the modified gray world transform [24], the scale factors are determined in such a way that each color is counted only once for a given image, even if there are multiple pixels having the same position in the color space.

There are also a number of normalization techniques designed for the sake of skin detection. R.-L. Hsu et al. proposed a lighting compensation technique, which operates in *RGB* color space prior to applying an elliptical skin model [36]. The top 5% of the gamma-corrected luminance values in the image define the *reference white* ( $L' = L^\gamma$ , where  $L$  and  $L'$  are input and gamma-corrected luminance values, respectively). After that, the *R*, *G* and *B* components are linearly scaled, so that the average gray value of the reference-white pixels equals 255. This operation normalizes the white balance and makes the skin model applicable to various lighting conditions. P. Kakumanu [42] proposed to use a neural network to determine the normalization coefficients dynamically for every image. J. Yang et al. introduced a modified Gamma correction [91] which was reported to improve the results obtained using statistical methods. U. Yang et al. took into account the physics of the image acquisition process and proposed a learning scheme [93] which constructs an illumination-invariant color space based on a presented training set with annotated skin regions. The method operates in a modified *rg* color space. Skin pixels are

subject to principal components analysis to determine the direction of the smallest variance. Thus, a single-dimensional space is learned, which minimizes the variance of the skin tone for a given training set.

## 4.2 Rule-Based Classification

Rule-based methods operate using a set of thresholds and conditions defined either in one of existing color spaces (e.g. *RGB*) or the image is first transformed into a new color space, in which the skin color can be easier separated from the non-skin color. There have been a number of methods proposed which adopt such an approach and new algorithms are still being proposed. Unfortunately, the recent methods do not contribute much to the state of the art, which is confirmed by the presented experimental results.

One of the first skin color modeling techniques was proposed by K. Sobottka and I. Pitas in 1996. They observed that the skin tone can be defined using two ranges of  $S \in [0.23, 0.68]$  and  $H \in [0, 50]$  values in the *HSV* color model [74]. A modification of this simple technique was later proposed by S. Tsekridou and I. Pitas [85] and it was used for face region segmentation in the image watermarking system [63]. The rule takes the following form in the *HSV* color space:

$$\begin{cases} (0 \leq H \leq 25) \vee (335 \leq H \leq 360) \\ (0.2 \leq S \leq 0.6) \wedge (0.4 \leq V) \end{cases} \quad (2)$$

Projection of these rules onto the *RGB* color space is shown in Fig. 3a using *RG*, *RB*, *GB* and normalized *rg* planes. The darker shade indicates the higher density of skin pixels.

J. Kovac et al. defined fixed rules [54, 76] that determine skin color in two color spaces, namely *RGB* and  $C_bC_r$  (ignoring the luminance channel). The rules in *RGB* are as follows:

$$\begin{cases} (R > 95) \wedge (G > 40) \wedge (B > 20) \\ \max(R, G, B) - \min(R, G, B) > 15 \\ |R - G| > 15 \wedge (R > G) \wedge (R > B) \end{cases} \quad (3)$$

for uniform daylight illumination or

$$\begin{cases} (R > 220) \wedge (G > 210) \wedge (B > 170) \\ |R - G| \leq 15 \wedge (R > G) \wedge (R > B) \end{cases} \quad (4)$$

for flashlight lateral illumination. If the lighting conditions are unknown, then a pixel is considered as skin, if its color meets one of these two conditions. These rules are illustrated in *RG*, *RB*, *GB* and *rg* planes in Fig. 3b.

R.-L. Hsu et al. defined the skin model [36] in the  $YC_bC_r$  color space, which is applied after the normalization procedure outlined earlier in Sect. 4.1. The authors

observed that the skin tone forms an elliptical cluster in the  $C_b C_r$  subspace. However, as the cluster's location depends on the luminance, they proposed to nonlinearly modify the  $C_b$  and  $C_r$  values depending on the luminance  $Y$ , if it is outside the range  $Y \in [125, 188]$ . Afterwards, the skin cluster is modeled with an ellipse in the modified  $C'_b C'_r$  subspace. Skin distribution modeled by these rules is presented in Fig. 3c.

Elliptical model of the skin cluster was also presented by J.-C. Terrillon et al., who argued that the skin color can be effectively modeled using the Mahalanobis distances computed in a modified *STV* color space [83]. This model was further improved by F. Tomaz et al. [84].

G. Kukharev and A. Nowosielski defined the skin detection rules [57] using two color spaces, i.e. *RGB* and  $YC_b C_r$ . Here, a pixel value is regarded as skin:

$$\begin{cases} R > G \wedge R > B \\ (G \geq B \wedge 5R - 12G + 7B \geq 0) \vee (G < B \wedge 5R + 7G - 12B \geq 0) \\ C_r \in (135, 180) \wedge C_b \in (85, 135) \wedge Y > 80. \end{cases} \quad (5)$$

This model is presented in *RGB* and  $(r, g)$  coordinates in Fig. 3d.

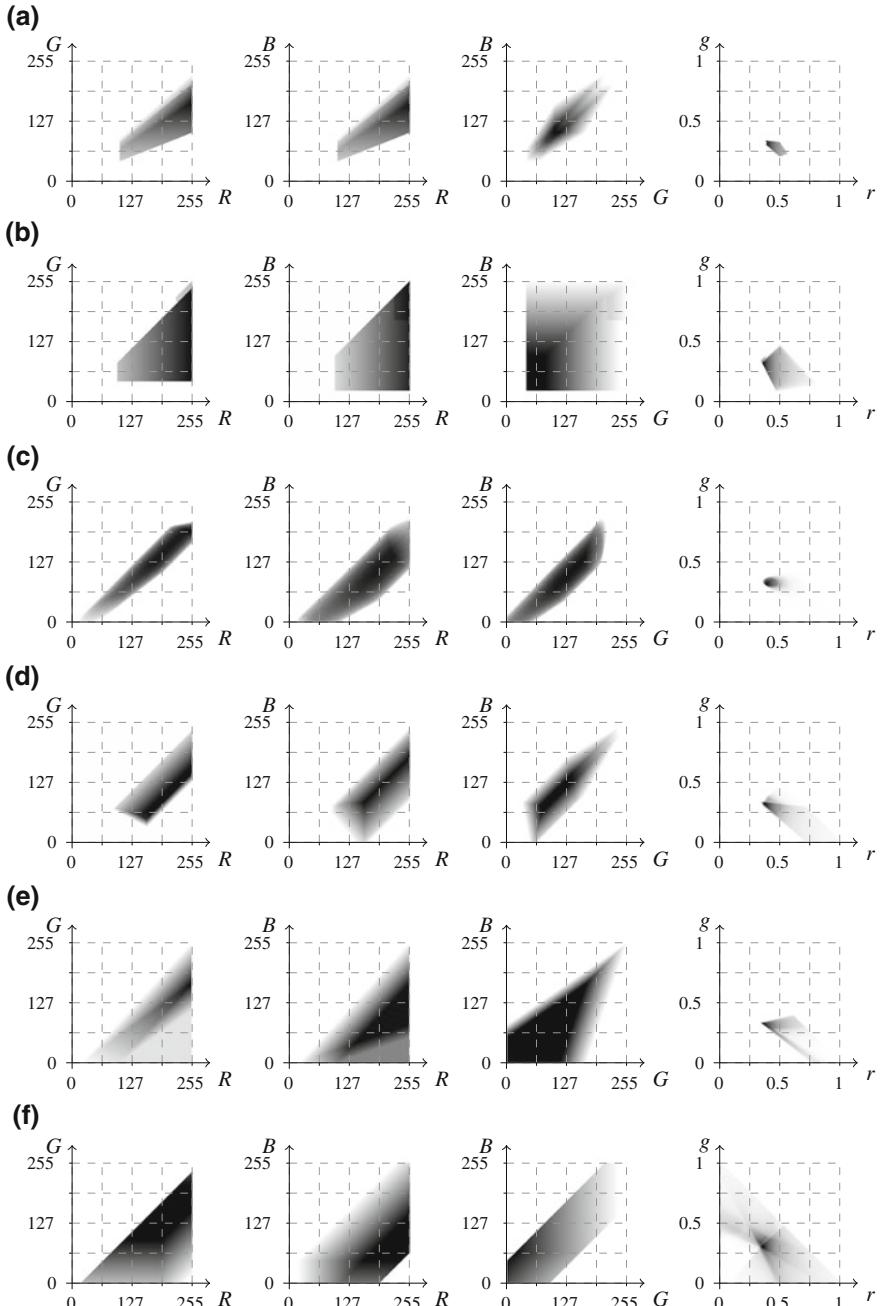
In 2009, A. Cheddad et al. proposed to transform the normalized *RGB* color space into a single-dimensional error signal, in which the skin color can be modeled using a Gaussian curve [13]. After the transformation, a pixel is regarded as skin if it fits between two fixed thresholds determined based on the standard deviation of the curve. The model is also illustrated in the *RGB* color space in Fig. 3e.

Recently, Y.-H. Chen et al. analyzed the distribution of skin color in a color space derived from the *RGB* model [15]. They observed that the skin color is clustered in a three-channel color space obtained by subtracting the *RGB* values:  $sR = R - G$ ,  $sG = G - B$ ,  $sB = R - B$ . After that they set the thresholds in the new color space to classify every pixel. The rules are visualized in Fig. 3f.

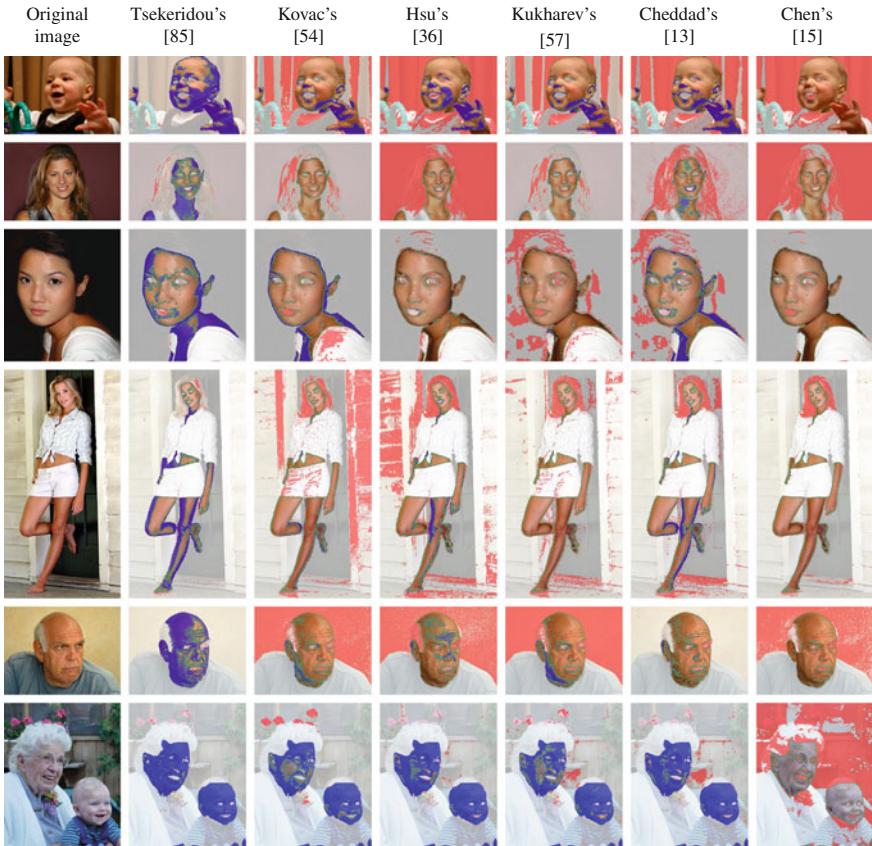
Some skin detection results obtained using six different methods are presented in Fig. 4. False positives are marked with a red shade, while false negatives—with the blue, and true positives are rendered normally. It can be seen from the figure that the detection error is generally quite high, however some methods present better performance in some specific cases. The overall performance scores are compared in Fig. 5. Rule-based models deliver worse results than the Bayesian skin model, however their main advantage lies in their simplicity. If the lighting conditions are controlled and fixed, then using a rule-based model may be a reasonable choice. Nevertheless, in a general case, the machine learning approaches outperform the models based on fixed rules defined in color spaces.

### 4.3 Machine Learning Methods

In contrast to the rule-based methods, the skin model can also be learned from a classified training set of skin and non-skin pixels using machine learning techniques. In most cases, such an approach delivers much better results and it does not require



**Fig. 3** Skin color models presented in  $RG$ ,  $RB$ ,  $GB$  and  $r_g - g$  planes. S. Tsekeridou and I Pitas [85] (a). J. Kovac et al. [54] (b). R.-L. Hsu et al. [36] (c). G. Kukharev and A. Nowosielski [57] (d). A. Cheddad et al. [13] (e). Y.-H. Chen et al. [15] (f)



**Fig. 4** Skin detection results obtained using different rule-based methods

any prior knowledge concerning the camera characteristics or lighting conditions. It is often argued that the main advantage of the rule-based methods is that they do not require a training set. However, the rules are designed based on observed skin-tone distribution, which means that a form of a training set is required as well. J. Brand and J.S. Mason confirmed in their comparative study that the histogram-based approach to skin modeling outperforms the methods which operate using fixed thresholds in the *RGB* color space [11]. Some machine learning techniques require large amount of training data (e.g. Bayesian classifier), while others are capable of learning effectively from small, but representative training sets. In this section the most important machine learning techniques used for skin detection are presented and discussed.

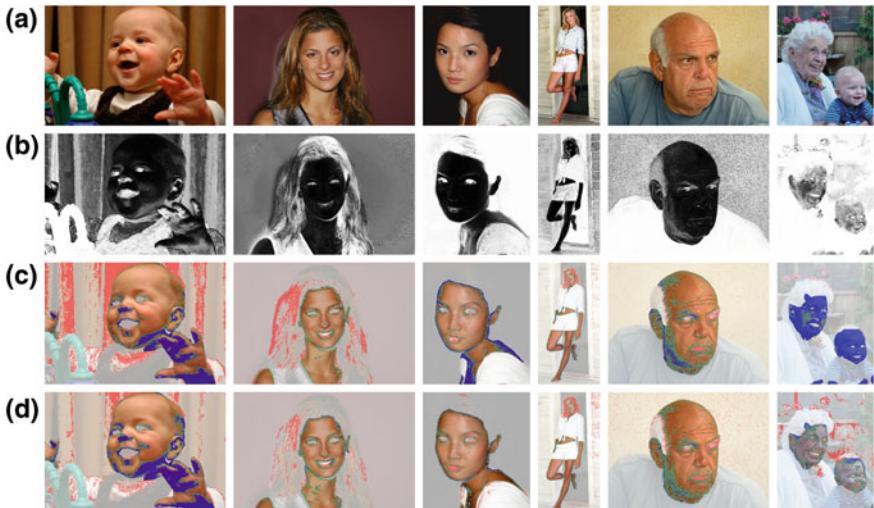
## Bayesian Classifier

Analysis of skin and non-skin color distribution is the basis for many skin detection methods. They may consist in a simple analysis of 2D or 3D histograms of skin color acquired from a training set or may involve the Bayesian theory to determine the probability of observing the skin given a particular color value. Such an approach was adapted by B.D. Zarit et al., whose work [98] was focused on the analysis of the skin and non-skin histograms in two-dimensional color spaces. At the same time, M.J. Jones and J.M. Rehg proposed to train the Bayesian classifier in the *RGB* space using all three components [39, 40]. The main principles of these techniques are as follows.

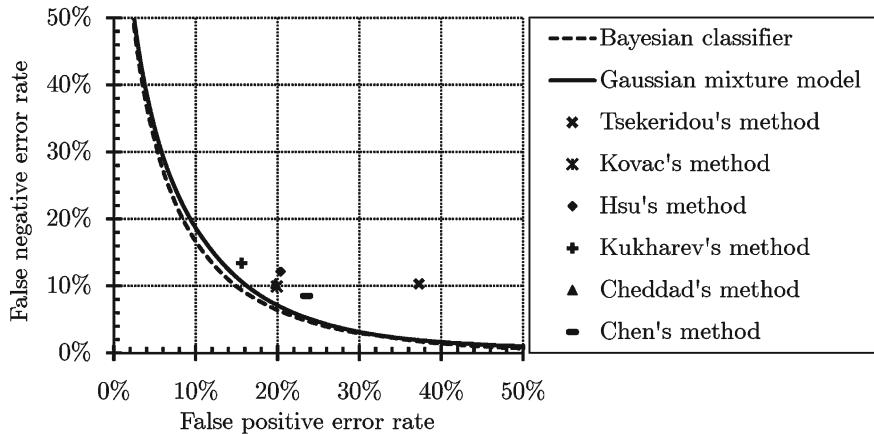
At first, based on a training set, histograms for the skin ( $C_s$ ) and non-skin ( $C_{ns}$ ) classes are built. The probability of observing a given color value ( $v$ ) in the  $C_x$  class can be computed from the histogram:

$$P(v|C_x) = C_x(v)/N_x , \quad (6)$$

where  $C_x(v)$  is the number of  $v$ -colored pixels in the class  $x$  and  $N_x$  is the total number of pixels in that class. Maximal number of histogram bins depends on the pixel bit-depth and for most color spaces it equals  $256 \times 256 \times 256$ . However, it is often reported beneficial to reduce the number of bins per channel. Our experiments, reported later in this section, indicated that the optimal histogram bin number depends on the training set size. Basically, the smaller the training set, the smaller number of bins should be used to achieve higher generalization.



**Fig. 5** Skin detection results obtained using Bayesian classifier: original image (a), skin probability map (b), segmentation using a threshold optimized for the whole *ECU-V* data set (c) and using the best threshold determined for each particular image (d)



**Fig. 6** ROC curves obtained for the Bayesian classifier and the Gaussian mixture model, and errors obtained for the rule-based skin detectors

It may be expected that a pixel presents the skin, if its color value has a high density in the skin histogram. Moreover, the chances for that are larger, if the pixel's color is not very frequent among the non-skin pixels. Taking this into account, the probability that a given pixel value belongs to the skin class is computed using the Bayes rule:

$$P(C_s|v) = \frac{P(v|C_s)P(C_s)}{P(v|C_s)P(C_s) + P(v|C_{ns})P(C_{ns})}, \quad (7)$$

where *a priori* probabilities  $P(C_s)$  and  $P(C_{ns})$  may be estimated based on the number of pixels in both classes, but very often it is assumed that they both equal  $P(C_s) = P(C_{ns}) = 0.5$ . If the training set is large enough, then the probabilities  $P(C_s|v)$  for all possible color values can be determined, and the whole color domain is densely covered. For smaller training sets, the number of histogram bins should be decreased to provide proper representation for every bin. The learning phase consists in creating the *skin color probability look-up table* ( $\mathbf{P}_s$ ), which maps every color value in the color space domain into the skin probability (which is also termed as *skinniness*). After training, using the look-up table, an input image is converted into a skin probability map, in which skin regions may be segmented based on an acceptance threshold ( $P_{th}$ ). The threshold value should be set to provide the best balance between the false positives and false negatives, which may depend on a specific application. This problem is further discussed in Sect. 7.

Examples of the skin segmentation outcome obtained using the Bayesian classifier are presented in Fig. 6. Original images (a) are transformed into skin probability maps (b) which are segmented using two threshold values, namely globally (c) and locally (d) optimized. The former is set, so as to minimize the detection error for the whole data set, while the latter minimizes the error independently for each image. It can be noticed that the detection errors are smaller than in case of using the rule-based

methods, whose results were shown in Fig. 4. The advantage is also illustrated in Fig. 5 in a form of *ROC* curves. Here, the results obtained for the rule-based methods are presented as points, because their performance does not depend on the acceptance threshold. In the case of the Bayesian classifier, as well as the Gaussian mixture model, which is discussed later in this section, skin probability maps are generated, hence the *ROC* curves may be rendered. It can be seen that the Bayesian classifier outperforms the rule-based methods and also it is slightly better than the Gaussian mixture model.

## Gaussian Mixture Models

Using non-parametric techniques, such as those based on the histogram distributions, the skin probability can be effectively estimated from the training data, providing that the representation of skin and non-skin pixels is sufficiently dense in the skin color space. This condition is not necessarily fulfilled in all situations. A technique which may be applied to address this shortcoming, consists in modeling the skin color using a *Gaussian mixture model* (GMM). Basically, if the histogram is approximated with a mixture of Gaussians, then it is smoothed at the same time, which is particularly important in case of sparse representation. GMM has been used for skin detection in various color spaces, in which a single pixel is represented by a vector  $\mathbf{x}$  of the dimensionality  $d$ , whose value depends on a particular color space. Usually  $d = 2$  or  $d = 3$ , but also skin color was modeled using Gaussians in the one-dimensional spaces [13, 94].

In general, using the adaptive Gaussian mixture model, the data are modeled with a mixture of  $\mathcal{K}$  Gaussian distributions. In the majority of approaches, only the skin-colored pixels are modeled with the Gaussian mixtures, nevertheless non-skin color could also be modeled separately. Thus, in such situations, these two models (i.e. skin and non-skin) are created. Each Gaussian distribution function is characterized with a weight  $\alpha_i > 0$  in the model, where  $\sum_{i=1}^{\mathcal{K}} \alpha_i = 1$ . The probability density function of an observation pixel  $\mathbf{x}$  in the mixture model is given as:

$$p(\mathbf{x}|\Theta) = \sum_{i=1}^{\mathcal{K}} \alpha_i p(\mathbf{x}|i; \theta_i), \quad (8)$$

where  $p(\mathbf{x}|i; \theta_i)$  is the probability for a single Gaussian:

$$p(\mathbf{x}|i; \theta_i) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i)\right). \quad (9)$$

Here,  $\alpha_i$  is the  $i$ th Gaussian weight estimation and  $\Theta = (\theta_1, \dots, \theta_{\mathcal{K}})$  is the parameter vector of a mixture composition.  $\theta_i = \{\mu_i, \Sigma_i\}$  consists of  $i$ th Gaussian distribution parameters, that is, the mean value  $\mu_i \in \mathbb{R}^d$  and covariance  $\Sigma_i$ , which is a

$d \times d$  positive definite matrix. The parameters of GMM are estimated based on the expectation-maximization (EM) algorithm.

The EM algorithm is an iterative method for finding the maximum likelihood (ML) function:

$$\mathcal{L}(\mathbf{X}; \Theta) = \prod_{n=1}^N p(\mathbf{x}_n | \Theta). \quad (10)$$

This function estimates the values of the model parameters, so as they best describe the sample data.

The EM algorithm includes two steps, namely:

*Expectation:* Calculate the expected value of the log likelihood function:

$$Q(\Theta | \Theta^{(t)}) = E \left[ \log \mathcal{L}(\mathbf{X}; \Theta) | \mathbf{X}, \Theta^{(t)} \right], \quad (11)$$

where  $\Theta^{(t)}$  is the current set of the parameters.

*Maximization:* Find the parameter that maximizes this quality:

$$\Theta^{(t+1)} = \arg \max_{\Theta} Q(\Theta | \Theta^{(t)}). \quad (12)$$

In this algorithm, the GMM parameters are determined as follows:

$$\hat{\mu}_i^{(t+1)} = \frac{\sum_{n=1}^N p^{(t)}(i | \mathbf{x}_n) \mathbf{x}_n}{\sum_{n=1}^N p^{(t)}(i | \mathbf{x}_n)}, \quad (13)$$

$$\hat{\Sigma}_i^{(t+1)} = \frac{\sum_{n=1}^N p^{(t)}(i | \mathbf{x}_n) (\mathbf{x}_n - \hat{\mu}_i) (\mathbf{x}_n - \hat{\mu}_i)^T}{\sum_{n=1}^N p^{(t)}(i | \mathbf{x}_n)}, \quad (14)$$

$$\hat{\alpha}_i^{(t+1)} = \frac{1}{N} \sum_{n=1}^N p^{(t)}(i | \mathbf{x}_n), \quad (15)$$

$$p^{(t)}(i | \mathbf{x}_n) = \frac{\alpha_i^{(t)} p(\mathbf{x}_n | \theta_i^{(t)})}{p(\mathbf{x}_n | \Theta^{(t)})}. \quad (16)$$

The EM algorithm is initiated with a given number of Gaussian mixtures ( $K$ ) and the model parameters are obtained using the  $k$ -means algorithm.

As mentioned earlier, skin color has been modeled using GMM in various color spaces. Normalized  $rg$  chromaticity space (i.e.  $\mathbf{x} = [r, g]^T$ ) was used by H. Greenspan et al. [32], who divided the color space into  $K$  face color regions and a complementary non-face region (which gives  $K + 1$  decision regions). For each pixel  $\mathbf{x}_n$ , they compute the probability from each Gaussian (8) and the one with the strongest response is selected. When the probability is less than a defined threshold, then the pixel is classified as non-skin, otherwise the pixel is labeled as a member of the selected Gaussian and classified as skin. T.S. Caetano et al. [12] proposed to

compute the probability using the whole mixture model, where the probability for each component is calculated separately. Moreover, they presented and discussed the results if the skin was modeled using different number of Gaussians.

B. Choi et al. presented an adult image detection system which is based on skin color modeling followed by a support vector machines classifier [17]. Here, the Gaussian mixtures operate in the *RGB* color space (i.e.  $\mathbf{x} = [R, G, B]^T$ ). Not only is the probability from the skin model ( $P_s$ ) used here, but also the probability of the non-skin model ( $P_{ns}$ ) is considered for classification. If the ratio of  $P_s$  to  $P_{ns}$  is below a defined threshold, then the pixel is classified as non-skin.

S.L. Phung et al. modeled the skin color using three single Gaussians in a 3D color space ( $\mathbf{x} = [Y, C_b, C_r]^T$ ), instead of using the mixture model [65]. These Gaussians correspond approximately to three levels of luminance: low, medium and high. The pixel is classified as skin if it satisfies the following two tests:

1.  $75 \leq C_b \leq 135$  and  $130 \leq C_r \leq 180$ ;
2. The minimum Mahalanobis distance from  $\mathbf{x}$  to the skin clusters (i.e. to the three Gaussians) is below a certain threshold.

M.F. Hossain et al. also used the  $YC_bC_r$  color space for skin regions extraction [35]. They addressed the problem of varying illumination using two skin models defined for normal and bright lighting conditions. The Gaussian mixture model is used only for chrominance ( $\mathbf{x} = [C_b, C_r]^T$ ), which automatically segments the skin portion. Appropriate Gaussian is selected based on the mean value of the luminance  $Y$ .

M.-H. Yang and N. Ahuja transformed the *RGB* color space to *CIELUV* and discarded the *L* component to make the model independent from lighting conditions [92] ( $\mathbf{x} = [u, v]^T$ ). An image region is regarded as a skin area, if more than 70 % of the pixels in the field are classified to the skin class.

According to the experiments reported by B.N. Jagadish et al. [37, 67], the skin distribution is not symmetric and mesokurtic, which means that it cannot be modeled properly using a mixture of Gaussians. They argue that the Pearson distribution of type-IIb and IVa is most suitable to model skin color distribution in the *HSI* color space.

## Artificial Neural Networks

*Artificial neural networks* (ANN) were considered for skin detection by many researchers, and they were used to classify every individual pixel as skin or not based on its value in the color space [9, 23, 71, 82]. Recently, a thorough survey on these methods was published by H.K. Al-Mohair [4]. Not only were ANN used for modeling the skin color itself in different color spaces, but they were also applied to color normalization [42] and model adaptation [59, 90], which is discussed later in Sect. 6.1.

## Support Vector Machines

*Support vector machines* (SVM) [20] were also applied to skin detection. SVM is a robust and widely adopted binary classifier which has been found highly effective for a variety of pattern recognition problems. Based on a labeled training set, it determines a hyperplane that linearly separates two classes in a higher-dimensional kernel space. The hyperplane is defined by a small subset of the vectors from the entire training set, termed *support vectors* (SV). Afterwards, the hyperplane is used to classify the data of the same dimensionality as the training set data. In the case of skin detection, SVM can be used to classify the pixels, represented as vectors in the input color space. Here, the main problem lies in high, i.e.  $O(n^3)$  time and  $O(n^2)$  memory complexity of SVM training, where  $n$  is the number of samples in the training set. Usually, the number of available skin and non-skin pixels is too large to train SVM, and a sort of training set selection must be proceeded. J. Han et al. proposed to use active learning for this purpose [33] which is one of widely adopted approaches towards dealing with huge training sets for SVM [61, 70]. During our works we have found genetic algorithms quite effective for reducing large training sets for the sake of skin detection using SVM [48].

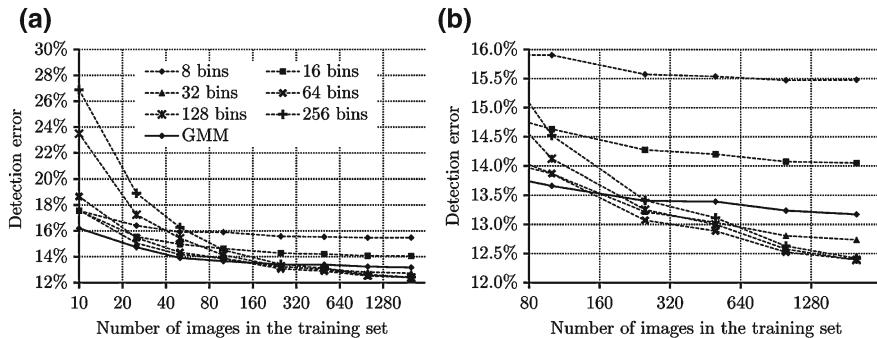
## Conclusions on Using Machine Learning for Skin Detection

The learning schemes discussed earlier in this section are commonly used for skin detection. Apart from them, there were also some attempts to use other popular learning machines. G. Gomez proposed a rule induction method [31] which optimizes the decision rules to minimize the detection error. R. Khan et al. used a random forest [51], trained in the *IHLS* color space [34]. The random forest classifier was claimed to outperform other learning machines, but this may have been attributed to their non-optimal parameters and settings, which are not discussed in that work. Furthermore, size of the training set has not been quoted neither, and it may have fundamental influence on the obtained results. Poor performance of the Bayes classifier quoted in that work may suggest that the training set was small and not representative. Similar objections may be raised in the case of a recent review on using machine learning schemes for skin classification, published by the same authors [52].

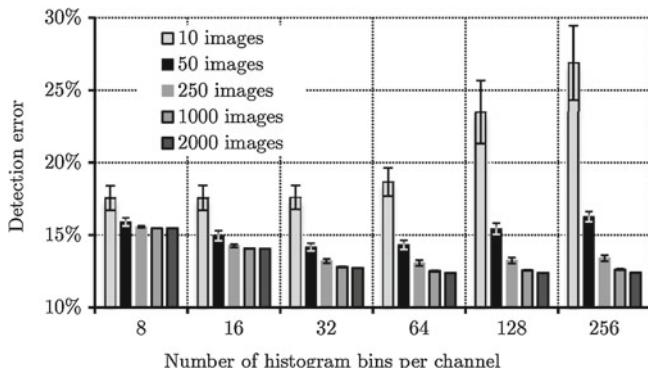
A serious drawback of using advanced machine learning methods (e.g. GMM, ANN, SVM) for skin detection is a long classification time, which usually excludes real-time applications. However, as the dimensionality of the input data is low (i.e. two or three dimensional color space) and the resolution of each dimension is usually limited to 256 values, it is feasible to build a look-up table once the classifier has been trained. This provides an effective optimization of the execution time, however it requires additional computations during the training phase which are usually acceptable.

It was observed that for sufficiently large training sets the effectiveness of skin detection is similar for different classifiers [64]. However, in case of small training sets, some classifiers may better generalize than others, and this was also studied

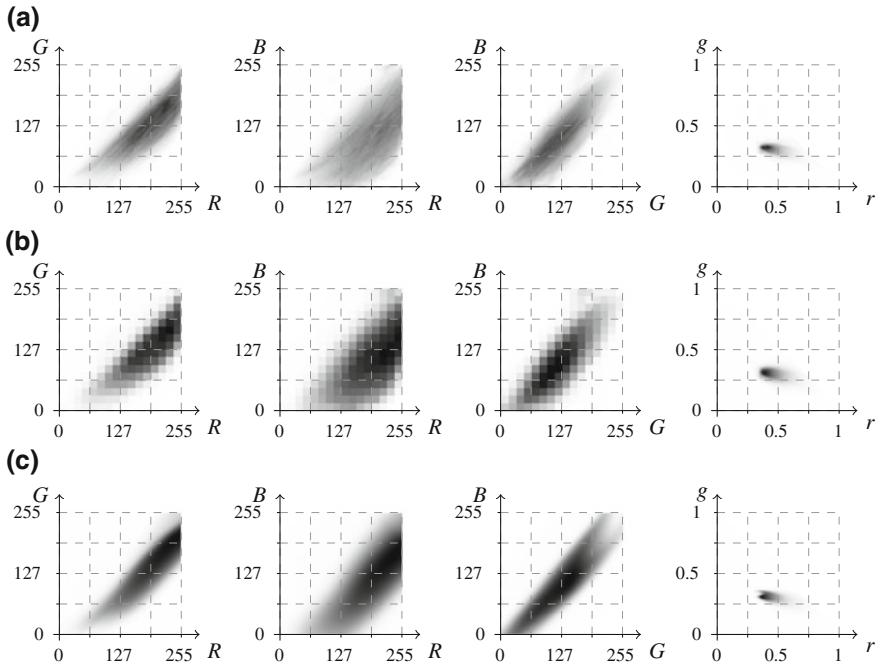
during our research. In our study we investigated the performance of the Bayesian classifier and Gaussian mixture model depending on the number of images in the training set and color space quantization for the Bayesian classification. In the case of GMM, the skin distribution was modeled using two Gaussians, whereas three Gaussians were used to approximate the non-skin distribution. The dependence between the skin detection error ( $\delta_{min}$ ) and the train set size is presented in Fig. 7, where the Bayesian classifier was trained in the *RGB* color space quantized to six different numbers of bins per channel. For the training sets smaller than 2000 images, each presented result was averaged based on the scores obtained using six randomly selected subsets of the *ECU-T* data set. The errors and their standard deviation obtained for five different sizes of the training set are also presented in Fig. 8. Skin distribution learned from the entire available training data is also presented in Fig. 9. It can be observed from Fig. 7 that for small sizes of training sets, it is beneficial to decrease the number of histogram bins per channel, but the improvement is limited using these sets when more training data are available. GMM performs definitely better than the



**Fig. 7** Dependence between the skin detection error ( $\delta_{min}$ ) and the training set size for GMM and Bayesian classifier with a different number of histogram bins. The whole investigated range is presented in (a), and the results for training sets larger than 100 images are shown in (b)



**Fig. 8** Skin detection error ( $\delta_{min}$ ) obtained using the Bayesian classifier depending on the color space quantization and training set size



**Fig. 9** Skin probability distribution presented in  $RG$ ,  $RB_gGB$  and  $r - g$  planes modeled using Bayesian learning with 256 (a) and 32 (b) bins per channel, and GMM with 2 skin and 3 non-skin Gaussians (c). Bayesian classifier (256 bins per channel) (a). Bayesian classifier (32 bins per channel) (b). Gaussian mixture model (2 skin and 3 non-skin Gaussians) (c)

Bayesian classifier for training sets containing less than 100 images, but for large training sets it gets slightly outperformed by the latter. Here, we used two Gaussians to model skin pixels and three Gaussians to model non-skin pixels. Generally, the more Gaussians are used, the better results can be obtained for larger training sets, however we have not observed any advantage over the Bayesian classifier in these cases.

## 5 Textural Features

Analysis of textural features has been investigated in order to improve the performance and to enhance the color-based methods. It is usually applied as the second step in skin segmentation in color images. X. Wang et al. combined the segmentation in the  $RGB$  and  $YC_gC_b$  color spaces with the analysis of the textural information extracted from an image [89]. Firstly, the white balance in the  $YC_bC_r$  color space is performed before the skin modeling to minimize the influence of the external conditions on the color information. The color model in the  $RGB$  color space is defined

by a set of fixed rules. The authors investigated the skin distribution in various color spaces and noticed that the distribution in the  $YC_gC_b$  space is regular and of a circular shape. Thus, the  $YC_gC_b$  space was chosen to define the second skin model. In the proposed method, the detection outcome obtained by adding two resulting binary images for each color space is further improved by incorporating the texture analysis. The grey-level co-occurrence matrix (GLCM) is used to extract the textural features. Given a grey-scale image  $I$  of size  $n \times m$ , the GLCM  $P$  is defined as:

$$P(i, j) = \sum_{x=1}^n \sum_{y=1}^m \begin{cases} 1, & \text{if } I(x, y) = i \wedge I(x + \Delta_x, y + \Delta_y) = j, \\ 0, & \text{otherwise,} \end{cases} \quad (17)$$

where  $(\Delta_x, \Delta_y)$  is the offset between the pixels  $I(x, y)$  and  $I(x + \Delta_x, y + \Delta_y)$ . It is worth noting that the time complexity of determining the GLCM is dependent on the number of grey levels  $g$  and is proportional to  $O(g^2)$  [18], thus  $g$  must be reduced for the real-time processing. The textural features extracted from the GLCM, including contrast, entropy, angular second moment, correlation and homogeneity are finally used for further skin detection. According to the experiments performed for a set of 500 images with the complex background, the method resulted in significantly worse skin detection rate (80.1 %) comparing to the detection in the  $RGB$  (90.3 %) and  $YC_gC_b$  (83.7 %) color spaces only. However, it turned out to be effective in detection of the background and outperformed (88.4 %) the background detection rates obtained for  $RGB$  (71.1 %) and  $YC_gC_b$  (82.4 %).

P. Ng and C.M. Pun proposed a method combining a color-based segmentation with the texture analysis using 2-D Daubechies wavelets [62]. Here, the Gaussian mixture model (GMM) classifier is applied as the initial skin segmentation approach for the original  $RGB$  images. Then, the 2-D Daubechies wavelets are calculated for the sub-images around each pixel classified as a skin pixel. The authors introduced the mask-leakage permit concept allowing for the efficient selection of sub-images. The texture feature of each skin pixel is represented by the wavelet energy vector  $\mathbf{v}_e$  obtained by applying the Shannon entropy on the wavelet coefficients vector  $\mathbf{v}_c$ . The  $\mathbf{v}_e$  vectors are the observation data of the  $k$ -means clustering. Given a set of  $n$  observations  $(\mathbf{v}_e^1, \mathbf{v}_e^2, \dots, \mathbf{v}_e^n)$ , the aim of the  $k$ -means clustering is to divide the observations into  $k$ ,  $k < n$ , clusters minimizing the total intra-cluster variance. Finally, some clusters that are claimed to be the non-skin ones are eliminated based on the properties of the Shannon entropy. According to the experimental results showed for the set of 100 images selected randomly from the Compaq database, the proposed method did not improve the segmentation significantly. The false positives (21.6 %) were decreased by approximately 3 % comparing to the GMM classifier. However, the true positives (86.8 %) dropped by approximately 2 %.

The approach integrating color, texture and space analysis was described by Z. Jiang et al. [38]. Here, the skin probability map (SPM) color filter is used in the  $RGB$  color space. The method does not rely only on the color information, thus the authors lower the threshold  $\Theta$  used as the acceptance threshold for the color filter in order to determine all probable skin pixels, which results in increasing the false positives. A filter utilizing the textural features extracted using the Gabor wavelets is proposed

to reduce the false acceptance rate (FAR). A 2-D Gabor function  $g(x, y)$  is given as:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi j Wx\right), \quad (18)$$

where  $\sigma_x$  and  $\sigma_y$  are the scaling parameters and  $W$  is the central frequency of the filter. Let  $g(x, y)$  be the generating function of the Gabor filter family. The Gabor functions (wavelets)  $g_{m,n}(x, y)$ ,  $m = 0, 1, \dots, M - 1$ ,  $n = 0, 1, \dots, N - 1$ , are created by rotating and scaling of  $g(x, y)$ :

$$g_{m,n}(x, y) = a^{-m} g(x', y'), \quad (19)$$

where  $x' = a^{-m}(x \cos \theta + y \sin \theta)$ ,  $y' = a^{-m}(-x \sin \theta + y \cos \theta)$ ,  $\theta = n\pi/N$ ,  $a > 1$ ,  $M$  is the total number of scales and  $N$  is the total number of orientations. Given an image  $I$ , the texture feature of each pixel is calculated as:

$$T(x, y) = \frac{\sqrt{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} J_{m,n}^2(x, y)}}{MN}. \quad (20)$$

Here,  $J_{m,n}$  is the image Gabor wavelets transform defined as:

$$J_{m,n}(x, y) = \sum_{x_1} \sum_{x_2} I(x_1, x_2) g_{m,n}^*(x - x_1, y - y_2), \quad (21)$$

where  $g_{m,n}^*(x, y)$  is the complex conjugate of  $g_{m,n}(x, y)$ . The grey-scale image  $I$  is obtained by transforming the input *RGB* image:

$$I(x, y) = 0.3R(x, y) + 0.59G(x, y) + 0.11B(x, y). \quad (22)$$

Finally, the texture mask image is defined as:

$$M(x, y) = \begin{cases} 1, & \text{if } T(x, y) \leq \Theta_T, \\ 0, & \text{otherwise,} \end{cases} \quad (23)$$

where  $\Theta_T$  is the texture threshold. On the one hand, applying the texture filter results in decreasing the FAR. However, on the other hand, it can also filter out properly classified skin pixels, which leads to reducing the true acceptance rate (TAR). It is worth noting that the performance of the texture filter is limited in case of textured skin regions (e.g. wrinkles). Additionally, if the image background is as smooth as skin then its pixels are not filtered out, thus the false positives are not reduced (see examples in Fig. 10). The authors proposed to exploit the space information after the color and texture analysis to grow the skin regions by using the watershed segmentation with the region markers. Thus, the markers are selected according to the following rules: (1) if the pixels either do not pass the color filtering or have a



**Fig. 10** Skin detection results obtained using the SPM color filter and refined using the Gabor wavelets: original image (a), skin probability map (b), texture image (c), segmentation using the threshold  $\Theta = 180$  (d), image after applying the texture mask for various  $\Theta_T$  ( $20 \leq \Theta_T \leq 120$ ) (e)

large texture value, then they are set as the non-skin local minima markers, (2) if the pixels are classified by both color and texture filters as the skin pixels, then they are set as the skin local minima markers. The mean and the standard deviation are calculated for the resulting closed regions. Finally, the pixels are classified as the skin pixels if they either pass the initial color filtering or the mean and the standard deviation of the region that encloses the pixel are smaller than the given thresholds. The experimental results obtained for the database containing 600 images collected by the authors proved that the method reduces the FAR significantly (from 20.1 to 4.2 %) with the simultaneous increase in the TAR (from 92.7 to 94.8 %) comparing to the SPM approach. The authors did not provide any sensitivity analysis, thus it is unclear how to set the parameters of the algorithm which turns out to be its significant drawback.

A. Taqa and H.A. Jalab increased performance of the ANN-based skin classifier by including simple textural feature [82]. It is reported in the paper that although the textural features have very low discriminating power, they are capable of improving stability of the color-based models.

A. Conci et al. [19] used the spectral variation coefficient (SVC) to extract textural features relevant to differentiate skin pixels from non-skin background regions for

the *RGB* images. In this approach, the relation between the pixel positions on the texture element (i.e. a *texel*) of size  $M \times M$ ,  $M = 3, 5, \dots, 21$ , is considered. Firstly, the color intensities for each *R*, *G* and *B* channels are blended together according to the defined function and form the new channel values. Nextly, the SVC values are computed to determine the average and the standard deviation of the distances within a texel according to the given metric. Here, the  $D_4$  (Manhattan) metric is used:

$$D_4(p_1, p_2) = |x_1 - x_2| + |y_1 - y_2|, \quad (24)$$

where  $p_1$  and  $p_2$  are the pixels at the positions  $(x_1, y_1)$  and  $(x_2, y_2)$  respectively. It is worth noting that the number of  $D_4$  distance classes to consider is dependent on the texel size  $M \times M$ . The SVC values are finally calculated for each blended channel and for each class of distances:

$$SVC = \arctan\left(\frac{\mu}{s+1}\right)\sqrt{\mu^2 + (s+1)^2}, \quad (25)$$

where  $\mu$  is the average and  $s$  is the standard deviation of the blended channel. They are considered as the coordinates in the Euclidean space and the  $k$ -means clustering is applied for the sample classification. The proposed method was tested using four color images. The authors claim that the algorithm reduces the false negatives, however to obtain meaningful results it would be necessary to perform the validation on a larger data set.

A technique combining a Bayesian skin color classifier based on the non-parametric density estimation of skin and non-skin classes with a multi-scale texture analysis was proposed by B. Zafarifar et al. [97]. Here, the authors point out that the reflections and shadows can cause sudden changes in the luminance channel of a color image. According to this, a new textural feature is defined:

$$T(i, j) = [|Y(i, j-1) - 2Y(i, j) + Y(i, j+1)| + |Y(i-1, j) - 2Y(i, j) + Y(i+1, j)| - \tau]_0^l, \quad (26)$$

where  $Y(i, j)$  is the value of the luminance channel,  $\tau$  is the noise threshold and  $f(x) = [x]_0^l$  clips the value of  $x$  between 0 and  $l$ . Finally, the extracted features are exploited to determine the final segmentation output. The experimental results obtained for 173 annotated natural images showed the decrease in the false positive rates. The disadvantage of this method is its limited performance for heavily textured skin areas (e.g. skin of the elders).

Simple textural features were used by D.A. Forsyth and M.M. Fleck [26] in a system for human nudes detection. Here, the difference between the original and median-filtered intensity image indicates the texture amplitude. Small amplitude increases the skin probability, computed in the log-opponent color space derived from *RGB*. A very similar approach was adopted by A.A. Abin et al. [3], who improved the median-based filtering by measuring the variance in *R*, *G* and *B* channels. Results of the median-based and variance-based filters are combined with each other and used

to refine the color-based skin probability map. The refined probability map serves as an input to a cellular learning automata which is further discussed in Sect. 7. M. Fotouhi et al. proposed to extract textural features using a contourlet transform [27]. The input image is split into small patches, in which the contourlet features are computed in small patches that surround the skin-color pixels. The contourlet feature vectors are later subject to principal components analysis to reduce their dimensionality and classified using a multilayer perceptron with two hidden layers. The main disadvantage of this method is that it is extremely slow compared with alternative approaches (ca. 30 min per  $256 \times 256$  image), while the effectiveness is not competitive, following the results quoted in [27].

In general, simple textural features can be applied to skin detection, taking advantage from the observation that skin regions are usually smooth. As it is the absence of any textural features rather than the presence of their characteristic indicators, it is sufficient to extract relatively simple features. This makes it possible to reject those regions, in which skin colored pixels appear, but their roughness indicates that the particular region does not present human skin. Obviously, such regions would be misclassified using pixel-wise skin color models. The roughness may be manifested in the luminance and color channels, but it can also be most significant only in the skin probability map as observed in our earlier research [45, 47].

## 6 Adaptive Models

As it was explained earlier in Sect. 4, the performance of skin color models is limited because of an overlap between the skin and non-skin pixels in the color space. Basically, the more general the model is supposed to be, the larger the overlap is, and the more it does degrade the detection score. The skin model may be made more specific to a given scene, video sequence, or an image, if they were acquired in fixed lighting conditions and present the same individuals. In such cases, the overlap is usually much smaller and the discriminative power of a skin color model may be definitely higher. However, it is not a trivial task to optimize a color model, and there is no universally effective adaptation procedure. There have been a number of methods proposed which address the problem of adapting a color model, and they can be categorized taking into account two criteria, namely the information source for the adaptation, and the model type which is being adapted. A general classification

**Table 2** Classification of model adaptation methods

Model used for the adaptation		Information source for the adaptation		
		Tracking	Whole-image analysis	Face and hand detection
Threshold-based method	[21]	[59, 66]	[10, 60, 95]	
Histogram analysis	[73, 77]	[6, 99]	[43]	
Gaussian mixture	–	[80, 90, 100, 101]	[28, 78]	

of these methods is presented in Table 2. Later in this section three general groups are described—self-adaptation techniques, which adapt the model to an image using the information obtained by the universal model, and content-based techniques, which rely on other data sources delivered by object detectors.

## 6.1 Global Adaptation Techniques

S.L. Phung et al. observed that an optimal value for the acceptance threshold applied in the probability map depends on the particular image [66]. Also, they argue that a coherent skin region should have homogenous textural features, extracted based on the  $3 \times 3$  Sobel operator. Hence, the acceptance threshold is iteratively adapted, so as to maximize the skin blobs homogeneity.

M.-J. Zhang and W. Gao used ANN for determining an optimal acceptance threshold in a given skin probability map obtained with the Bayes classifier [99]. It is assumed that the threshold should be located in a local minimum of the probability image histogram. Hence, for every minimum, 13 features are extracted, which are regarded as a feature vector that is classified by the neural network. In this way, for every image an optimal threshold is determined which decreases the detection error.

A.A. Argyros and M.I.A. Lourakis [6] proposed to apply a globally-trained Bayesian classifier to extract skin region seeds using a high acceptance threshold. Such classification results in low false positive rate at a possible cost of high false negative error. After that, they apply a threshold hysteresis to expand the skin regions, which are later used to learn the Bayesian classifier. This creates a local skin model, which is combined with the global model, and used to classify the entire image. Finally, the method is applied to track the skin-colored objects, but it is the spatial analysis rather than tracking which delivers the data for the adaptation.

Q. Zhu et al. proposed a double-level approach towards adapting the skin model [100, 101]. First, the skin pixels are detected using a generic model characterized with a very low false negative rate, achieved at a cost of high false positives. Afterwards, the potential skin pixels are modeled using two Gaussians—one is supposed to represent the skin pixels, while the other is expected to be formed by false positives. Color, spatial and shape features are extracted for each Gaussian and the feature vectors are classified using SVM to select the Gaussian which represents the real skin pixels.

J.-S. Lee et al. proposed a method based on a multilayer perceptron which accommodates to specific lighting conditions observed in the input image [59]. At first, for each image in the training set, a chroma histogram of skin pixels is extracted. After that, the chroma histograms for all of the images were merged and grouped, which finally delivered five major chroma clusters. Hence, every cluster represents a number of images from the training set, which, according to the authors, were acquired under similar lighting conditions. Then, a multilayer perceptron was trained with image patches to decide which chroma cluster should be applied to a particular image. Hence, for every input image the patches around the image center are classified by the neural network, and based on their response, the most appropriate color

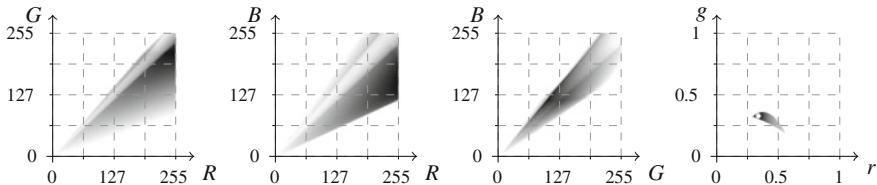
model is applied. In this way, the method adapts to a closed set of predefined lighting conditions.

ANN were also used for the adaptation by G. Yang et al. [90]. They observed that skin color distribution in the  $C_r C_b$  plane depends on the luminance ( $Y$ ), and it is more compact if the luminance value range is smaller. Hence, they used a neural network to determine the parameters of a Gaussian (i.e. the mean and standard deviation) based on the image histogram in  $Y$ . The Gaussian is subsequently used for skin detection. Unfortunately, the authors did not present any comparison with non-adaptive modeling in the three-dimensional  $YC_bC_r$  space.

An interesting adaptation technique was proposed by H.-M. Sun [80]. First of all, a histogram-based global skin model is trained, and it is applied to every input image to detect skin pixels. Then, the pixels classified as skin are subject to clustering in the color space and their distribution is modeled using a Gaussian mixture. The number of clusters and GMM's parameters are determined based on the distribution of the pixels classified as skin using the global model. Finally, the entire image is classified again based on a linear combination of the dynamically learned GMM and the global model. This combination stabilizes the result and is more effective than relying exclusively on the dynamic local model. The local adaptation makes it possible to decrease the skin probability of the pixels which do not form clusters in the color space. Usually, the groups of skin pixels in the input image are characterized with similar chrominance, which results in observing clusters in the color space. Contrary to that, the false positives are often isolated and will not be detected as a cluster during the adaptation phase.

## 6.2 Tracking-Based Adaptation

M. Soriano et al. proposed to adapt the model dynamically based on observed changes in a histogram extracted from the tracked skin region [77]. Initially, the skin is modeled in the  $rg$  color space based on the image histograms. The *skin histogram* values are divided by the corresponding values from the *whole-image histogram*. This produces a *ratio histogram* which is further used for classification. The ratio histogram can be defined only for non-zero whole-image histogram values, and the remaining values are assigned with zero skin probability. The detected skin regions are tracked in video sequences and the ratio histogram is adjusted to compensate the changes in lighting conditions that influence the skin appearance. The tracking procedure consists in determining the largest connected component of the pixels classified as skin, which lies inside an expanded bounding box of a facial region detected in the previous frame. From this region, the skin histogram is generated, and the ratio histogram is computed using that histogram as well as the histogram of the entire image. Using the updated ratio histogram, the final position of the facial region is determined. However, there is a risk that if the tracking works incorrectly, then the model gets destabilized at some point. In order to prevent such a situation, it is assumed that a pixel can be classified as skin only if it falls inside a certain skin



**Fig. 11** Skin locus which determines the adaptation boundary in the Soriano's method [77]

locus, limited by two quadratic functions defined in  $rg$  color space, as presented in Fig. 11. Most of the skin pixels fall inside the locus indeed, however it cannot be used directly as a decision surface for the classification. Otherwise, the false positives are very high.

L. Sigal et al. addressed the problem of varying lighting conditions which may severely affect the skin detection efficacy [73]. They used the expectation maximization algorithm to adapt the histogram-based classifier based on tracked skin regions.

F. Dadgostar proposed to take advantage of the motion detectors to adapt the skin color model [21]. Here, the skin color is defined by the lower and upper Hue thresholds in the  $HSV$  color model. If the pixel's Hue value is positioned between these thresholds, then it is considered as skin. Hue histogram of the in-motion skin pixels is extracted and analyzed to determine the optimal values of the thresholds. In this way the skin model is adapted to a scene (i.e. the lighting conditions and an individual, whose skin is to be detected) and skin regions can be segmented with higher effectiveness.

### 6.3 Face or Hand Region-Based Adaptation

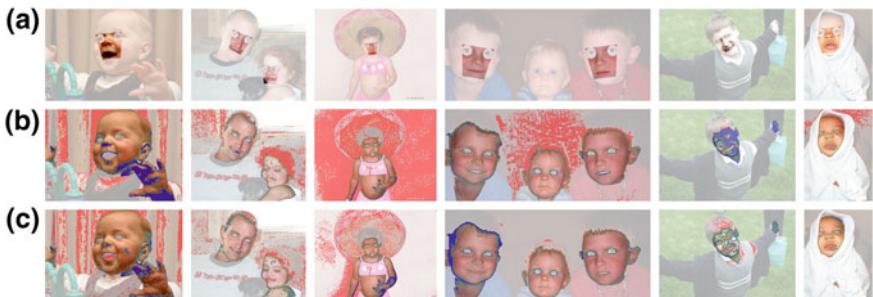
There are methods which effectively adapt the skin model to local conditions based on the detected faces. The first such an approach was proposed in 2002 by J. Fritsch et al. [28]. The adaptation scheme is similar to the Soriano's method [77], discussed earlier in the chapter. Here, the skin pixels are modeled using a unimodal Gaussian defined in the  $rg$  color space. The Gaussian's parameters are updated based on a facial region obtained using a face detector, which operates in the luminance channel. Similarly to [77], the adaptation is stabilized using a skin locus. The pixels from the facial region which do not lie inside the locus are ignored for the adaptation. Also, the adaptation works with a certain inertia, and the updated parameters are determined as a weighted mean of the current values and those extracted from the facial region.

H. Stern and B. Efros proposed to dynamically select the best color space, in which the skin model can be learned based on a detected facial region [78]. First, the color space is selected ( $RG$ ,  $rg$ ,  $HS$  or  $C_bC_r$ ), in which the skin pixels from the facial region can be best separated from the non-skin pixels around the region. After choosing the color space, the skin color is modeled using a single Gaussian.

R. Khan et al. combined the adaptive modeling based on face detection with spatial analysis using graph cuts [50]. Here, the faces are used as foreground seeds during the segmentation procedure. The main drawback, however, is the processing time of 1.5 seconds for small  $100 \times 100$  images. S. Bilal et al. used both faces and hands detected using Haar features [88] to define the skin color ranges in  $C_b$  and  $C_r$  channels of the  $YC_bC_r$  color space [10]. Such a local model was later applied to all of the subsequent images registered in a video sequence, which improved the hand tracking algorithm. J.F. Lichtenauer et al. used the positive skin samples acquired from detected faces to adapt the skin model in an introduced adaptive chrominance space [60].

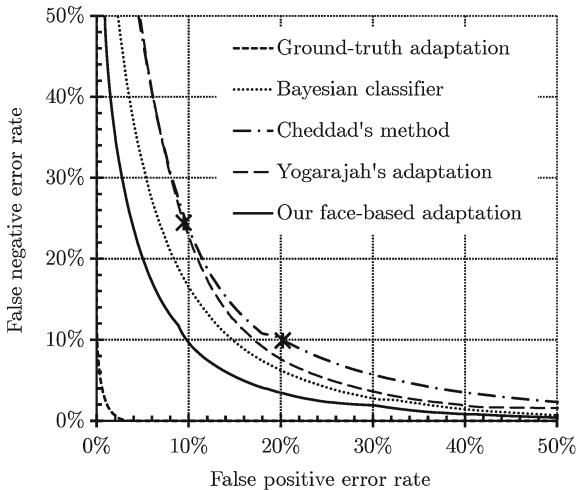
In our earlier works we also proposed to adapt the Bayesian skin classifier based on detected facial region [43]. The faces were detected using our face detector [49], which is capable of detecting the eyes with higher precision than the Viola-Jones method [88]. Based on the positions of the eyes, the face region is created as presented in Fig. 12a. The pixels which fall inside this region are regarded as skin, hence the skin histogram is generated. We reported that while the global model keeps 64 bins per channel, the skin is best modeled locally using only 16 bins, and then it is extrapolated to the resolution of 64 bins per channel. The probability look-up tables for the local and global model are afterwards combined with each other and applied to the whole image. Examples of segmentation results are presented in Fig. 12. Although the detection errors have not been eliminated, and in some cases they are still high, the improvement is significant, compared with the results obtained using the global model (Fig. 12b).

P. Yogarajah et al. presented a method for dynamic adaptation of the thresholds in the method proposed by A. Cheddad [13]. The thresholds in the single-dimensional error space are set dynamically based on the detected facial region [94, 95]. First, the error signal value distribution in the facial region is modeled using a single Gaussian. Then, the thresholds are set in the same way as proposed in [13]. Performance of this method is presented and compared with our earlier approach in Fig. 13 in the form of *ROC* curves. Here, the results are presented for 1375 images from *ECU-V* set, in which the faces were detected. Although Cheddad's and Yogarajah's methods offer binary classification, the probability maps can be easily generated by scaling



**Fig. 12** Detected facial regions (a) and skin segmentation result obtained using the Bayesian classifier (b) and adaptive facial region-based method [43] (c)

**Fig. 13** ROC curves obtained for two global models and two face-based adaptation techniques. Ground-truth adaptation is presented as well. Binary decision result for Cheddad's and Yogarajah's methods are marked with asterisks on the curves



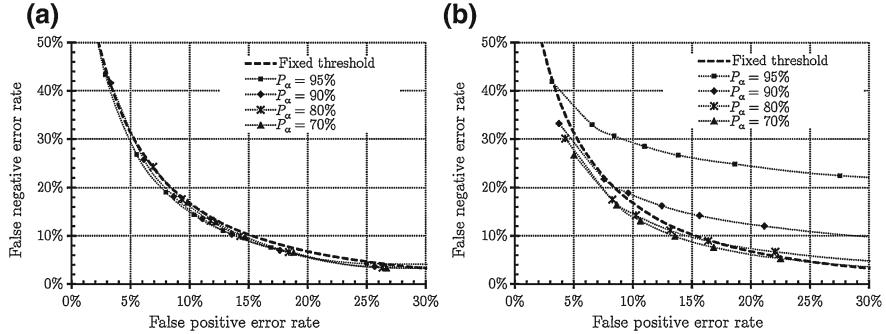
the distance from the approximated Gaussian mean. In this way, the *ROC* curves can be rendered, and the binary-decision results (which are obviously positioned on the curve) are marked in the figure with asterisks. It may be seen that both adaptation techniques offer some improvement, however it is much more significant using our approach [43]. Yogarajah's method was recently extended by W.R. Tan et al. [81], who combined the threshold-based adaptation with GMM modeling.

## 7 Spatial Analysis Methods

The limitations of the color-based skin models were investigated and reported by Q. Zhu et al. [100, 101]. They have shown that even if the color model is perfectly adapted to every individual image, the skin pixels usually cannot be separated from the non-skin pixels. This boundary is illustrated in Fig. 13 in a form of the ground-truth adaptation. Here, the Bayesian classifier was trained based on every individual image using the ground-truth information, and after that the learned rules were applied to the very same image. The detection errors are very small, but nevertheless they illustrate the limitations of the pixel-wise detectors. Naturally, the errors are much higher when the used skin color model is not perfectly adapted to the image.

The limits of pixel-wise detectors can be quite effectively addressed by analyzing the spatial alignment of the pixels classified as skin. It may be observed that skin pixels usually form consistent blobs in the images, while many false positives are scattered in the spatial domain. This general observation underpins several skin segmentation methods outlined in this section.

H. Kruppa et al. assumed that in most cases the skin regions are of an elliptical shape. Hence, they proposed a skin model [56] which maximizes the mutual infor-



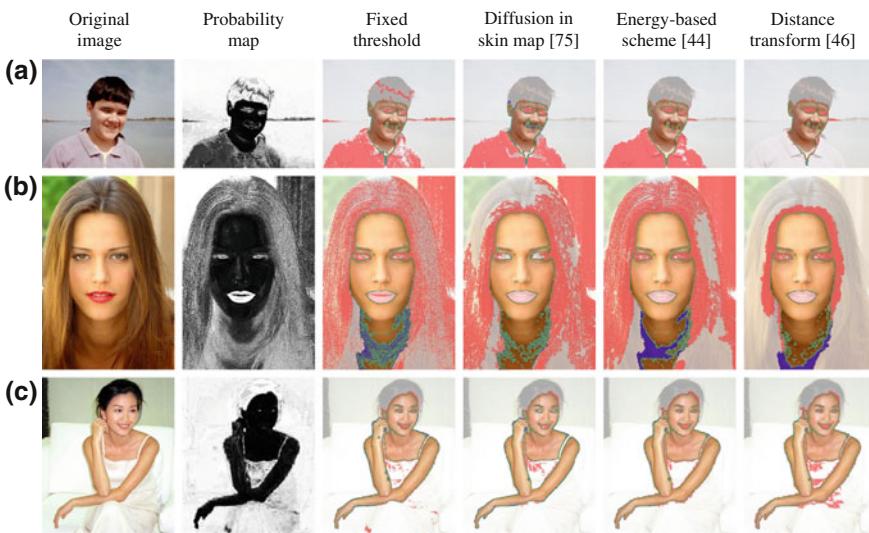
**Fig. 14** ROC curves obtained using a threshold hysteresis without (a) and with (b) size-based verification

mation between color-based and shape-based probability. The model is iteratively adapted for every image using a gradient descent optimization.

As it was mentioned in Sect. 6.1, A.A. Argyros and M.I.A. Lourakis [6] used a threshold hysteresis to perform skin regions segmentation. A similar technique has also been applied by H. Baltzakis et al. for tracking skin regions in video sequences [7]. This operation offers a certain reduction of the detection errors. First, the pixels which exceed a high probability threshold ( $P_\alpha$ ) are considered as seeds for the region growth procedure. Next, the adjacent pixels, whose skin probability is over the second, lower threshold ( $P_\beta$ ), are iteratively adjoined to the region. *ROC* curves obtained for different  $P_\alpha$  are presented in Fig. 14a. Each curve was rendered based on the errors obtained using different values of the lower acceptance threshold  $P_\beta$ . It can be observed that the higher  $P_\alpha$  is, the lower false positive error ( $\delta_{fp}$ ) is obtained in the seeds (i.e. the initial point of each curve with the smallest false positive value). The curves obtained using the threshold hysteresis are positioned slightly below the fixed-threshold curve, and for  $\delta_{fp} < 15\%$ ,  $P_\alpha = 95\%$  offers the largest improvement. During our research [44] we proposed to verify the skin seeds based on their relative size. The verification consists in rejecting those seeds, whose area is smaller than 10 % of the largest skin seed. *ROC* curves obtained following this procedure are presented in Fig. 14b. Here, the initial points are located below the fixed-threshold curve, hence it may be concluded that the small rejected seeds contain a significant amount of false positives. However, it can be noticed that if  $P_\alpha$  is high, then the false negatives decrease very slowly compared with the false positives increase, which positions the results over the fixed-threshold curve. The reason is that some skin regions contain only a small number of pixels with probability exceeding  $P_\alpha$ , and they are rejected during the verification. This points out that if  $P_\alpha$  is too large, then many skin regions will be excluded, and if it is too small, then the false positive error may be large already in the seeds, which by definition cannot be decreased using any region-growth algorithm. From the presented curves it may be concluded that the verification step is beneficial, because the error reduction is larger, but  $P_\alpha$  should not exceed 80 % in that case.

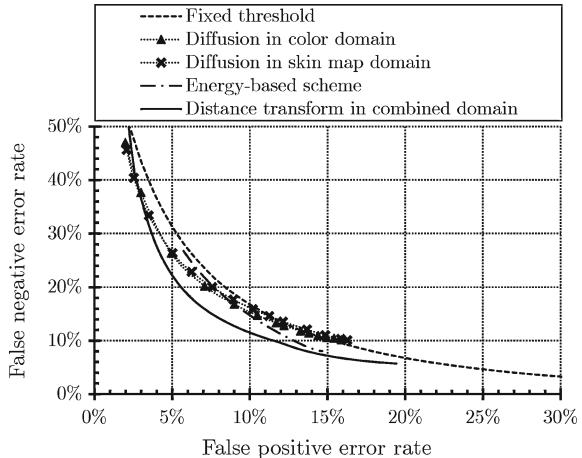
J. Ruiz-del-Solar and R. Verschae proposed to perform a controlled diffusion in color or skin probability domain [75]. This procedure consists of two general steps: (1) diffusion seeds extraction, and (2) the proper diffusion process. The seeds are extracted using pixel-wise skin probability maps, and they are formed by those pixels, whose skin probability exceeds a certain high threshold ( $P_\alpha$ ). During the second step, the skin regions are built from the seeds by adjoining the neighboring pixels which meet the diffusion criteria, defined either in the probability map or color space domain. These criteria are as follows: (1) distance between a source pixel  $x$  and a pixel  $y$  (that is to be adjoined) in the diffusion domain  $D_d$  is below a given threshold:  $|D_d(x) - D_d(y)| < \Delta_{max}$ , and (2) skin probability for the pixel which is to be adjoined must be over a certain threshold:  $P(y) > P_\beta$ . It is worth to note that this is the threshold hysteresis with an additional constraint imposed on a maximal difference between the neighboring pixels (either in terms of their probability or color). Hence, this works well if the region boundaries are sharp (diffusion stops due to high local differences), but the method is identical to the threshold hysteresis when there exists a smooth transition between the pixel values that leads from one region to another (the diffusion will then “leak” outside the region).

Spatial analysis of skin pixels was also the subject of our research. At first, we proposed an energy-based scheme for skin blob analysis [44]. Skin seeds are formed by high-valued pixels in the skin probability maps, similarly as in the diffusion method. In addition, we perform the size-based verification as explained earlier in this section. It is assumed that the seed pixels receive a maximal energy amount equal to 1, which is spread over the image. Amount of the energy that is passed depends on the probability value of the target pixel. If there is no energy to be passed, then the pixel is not adjoined to the skin region. Although this method implements a cumu-



**Fig. 15** Examples of skin detection results obtained using different spatial analysis methods

**Fig. 16** ROC curves obtained using spatial analysis methods



lative propagation (which helps reduce the “leakages” identified in the controlled diffusion), only skin probability is taken into account and local differences between the pixels are ignored. Recently, we have proposed to rely on the distance transform proceeded in a combined domain of hue, luminance and skin probability [46]. This approach is more effective than the energy-based scheme as the latter uses only skin probability during the propagation. Moreover, it has also a serious advantage over the Solar’s method, because the cumulative character of the distance transform perfectly addresses the main shortcoming of the diffusion, i.e. the vulnerability to smooth transitions between skin and non-skin regions. ROC curves obtained for different spatial analysis methods are presented in Fig. 15 and some qualitative results are shown in Fig. 16. For images (a) and (b) the distance transform in a combined domain reduces the errors significantly, while the alternative spatial analysis techniques do not offer much improvement. In the case of image (c), the energy-based scheme delivers the best result, however the errors are small for all of the tested methods.

There were also many other approaches towards using spatial analysis for improving skin detection. A.A. Abin et al. used the cellular automata to determine skin regions [3], but this process requires many iterations to achieve satisfactory results and cannot be applied for real-time processing. Also, K. Chenaoua and A. Bouridane used conditional random fields [55] to take advantage from spatial properties of skin regions [16]. However, this may be a time-consuming procedure as it involves simulated annealing for every analyzed image. M. Abdullah-Al-Wadud [2] proposed to transform an image into a single-dimensional color distance map, in which a water-flow procedure is carried out to segment skin regions. A.Y. Dawod used basic edge detectors to determine the boundaries of skin regions [22]. Z. Yong-jia et al. investigated the possibilities of using level sets for determining the boundaries of skin regions in a given skin probability map [96]. Unfortunately, it is difficult to conclude from the paper whether the method offers any advantage over other techniques.

## 8 Summary

There are plenty of approaches towards skin detection and segmentation, and the most relevant out of the existing techniques were outlined in this chapter. For virtually all of the methods, skin color modeling is the primary source of information here, because it is the color which constitutes the principal discriminative feature of human skin. Although skin detection is still an open problem and no satisfactory solution has been developed so far, in our opinion there is not much to be gained in the skin color modeling itself. This is caused by the overlap between the skin and non-skin pixels that can be observed in all of the popularly used color spaces. This overlap determines an effectiveness limit which cannot be eliminated globally relying on the pixel-wise classification.

A great potential is still hidden in the adaptive skin color modeling, which was well demonstrated by the ground-truth adaptation *ROC* curve in Fig. 13. Another group of powerful techniques for improving skin detection is underpinned by the spatial analysis and segmentation-based verification of the detection outcome. It may be beneficial to combine these two approaches by iteratively adapting the model, taking advantage of the spatial analysis. A similar direction was explored by A.A. Argyros and M.I.A. Lourakis in 2004 [6], but many advances have been made since then, both concerning the model adaptation as well as the spatial analysis. In our opinion, this is the most promising, while still little explored possibility for increasing the efficacy of skin detection that is worth being investigated in the future.

**Acknowledgments** This work has been supported by the Polish Ministry of Science and Higher Education under research grant no. IP2011 023071 from the Science Budget 2012–2013 and the European Union from the European Social Fund (grant agreement number: UDA-POKL.04.01.01-00-106/09).

## References

1. Abdullah-Al-Wadud M, Chae O (2007) Region-of-interest selection for skin detection based applications. In: International conference on convergence information technology, pp 1999–2004
2. Abdullah-Al-Wadud M, Chae O (2008) Skin segmentation using color distance map and water-flow property. In: Proceedings of the information assurance and security (ISIAS '08), pp 83–88
3. Abin AA, Fotouhi M, Kasaei S (2009) A new dynamic cellular learning automata-based skin detector. *Multimedia Syst* 15(5):309–323
4. Al-Mohair HK, Mohamad-Saleh J, Suandi SA (2012) Human skin color detection: a review on neural network perspective. *Int J Innovative Comput Inf Control (IJICIC)* 8(12):8115–8131
5. Albiol A, Torres L, Delp E (2001) Optimum color spaces for skin detection. In: Proceedings of the IEEE international conference on image processing, pp 122–124
6. Argyros AA, Lourakis MIA (2004) Real-time tracking of multiple skin-colored objects with a possibly moving camera. In: Proceedings of the ECCV, LNCS, vol 3023. Springer, pp 368–379

7. Baltzakis H, Pateraki M, Trahanias P (2012) Visual tracking of hands, faces and facial features of multiple persons. *Mach Vis Appl* 23:1141–1157
8. Berbar MA (2011) Novel colors correction approaches for natural scenes and skin detection techniques. *Int J Video Image Process Netw Secur* 11(2):1–10
9. Bhoyer KK, Kakde OG (2010) Skin color detection model using neural networks and its performance evaluation. *J Comput Sci* 6(9):963–968
10. Bilal S, Akmeliawati R, Salami MJE Shafie AA (2012) Dynamic approach for real-time skin detection. *J Real-Time Image Process*
11. Brand J, Mason J (2000) A comparative assessment of three approaches to pixel-level human skin-detection. In: Proceedings of the 15th international conference on pattern recognition vol 1, pp 1056–1059
12. Caetano TS, Olabarriaga SD, Barone DAC (2003) Do mixture models in chromaticity space improve skin detection? *Pattern Recogn* 36:3019–3021
13. Cheddad A, Condell J, Curran K, Mc Kevitt P (2009) A skin tone detection algorithm for an adaptive approach to steganography. *Signal Process* 89(12):2465–2478
14. Chen MJ, Chi MC, Hsu CT, Chen JW (2003) ROI video coding based on H.263+ with robust skin-color detection technique. In: IEEE international conference on consumer electronics, pp 44–45
15. Chen YH, Hu KT, Ruan SJ (2012) Statistical skin color detection method without color transformation for real-time surveillance systems. *Eng Appl Artif Intell* 25(7):1331–1337
16. Chenaoua K, Bouridane A (2006) Skin detection using a Markov random field and a new color space. In: Proceedings of the IEEE international conference on image processing, pp 2673–2676
17. Choi B, Chung B, Ryoo J (2009) Adult image detection using Bayesian decision rule weighted by SVM probability. In: Proceedings of the 4th international conference on computer sciences and convergence information technology (ICCIT '09), pp 659–662
18. Clausi D, Jernigan M (1998) A fast method to determine co-occurrence texture features. *IEEE Trans Geosci Remote Sensing* 36(1):298–300
19. Conci A, Nunes E, Pantrigo JJ, Sánchez Á (2008) Comparing color and texture-based algorithms for human skin detection. In: Proceedings of the ICEIS, pp 166–173
20. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
21. Dadgostar F, Sarrafzadeh A (2006) An adaptive real-time skin detector based on hue thresholding: a comparison on two motion tracking methods. *Pattern Recogn Lett* 27(12):1342–1352
22. Dawod A, Abdullah J, Alam M (2010) Adaptive skin color model for hand segmentation. In: Proceedings of the international conference on computer applications and industrial electronics (ICCAIE), pp 486–489
23. Duan L, Lin Z, Miao J, Qiao Y (2009) A method of human skin region detection based on PCNN. In: Proceedings of the international symposium on neural networks: advances in neural networks, ISNN, Part III, LNCS, vol 5553. Springer, Berlin, pp 486–493
24. Finlayson G, Hordley S, Hubel P (2001) Color by correlation: a simple, unifying framework for color constancy. *IEEE Trans Pattern Anal Mach Intell* 23(11):1209–1221
25. Finlayson GD, Schiele B, Crowley JL (1998) Comprehensive colour image normalization. In: Proceedings of the european conference on computer vison (ECCV), vol 1, Freiburg, Germany, pp 475–490
26. Forsyth DA, Fleck MM (1999) Automatic detection of human nudes. *Int J Comput Vis* 32:63–77
27. Fotouhi M, Rohban M, Kasaei S (2009) Skin detection using contourlet-based texture analysis. In: Proceedings of the 4th international conference on digital telecomm (ICDT'09), pp 59–64
28. Fritsch J, Lang S, Kleinehagenbrock M, Fink G, Sagerer G (2002) Improving adaptive skin color segmentation by incorporating results from face detection. In: Proceedings of the IEEE international workshop on robot and human interactive, communication, pp 337–343
29. Gasparini F, Corchs S, Schettini R (2005) Pixel based skin colour classification exploiting explicit skin cluster definition methods. In: Proceedings of the 10th congress of the international colour association, vol 1, pp 543–546

30. Gatta C, Rizzi A, Marini D (2000) Ace: an automatic color equalization algorithm. In: Proceedings of the first european conference on color in graphics image and vision (CGIV02)
31. Gomez G, Morales EF (2002) Automatic feature construction and a simple rule induction algorithm for skin detection. In: Proceedings of the ICML workshop on machine learning in computer vision, pp 31–38
32. Greenspan H, Goldberger J, Eshet I (2001) Mixture model for face-color modeling and segmentation. *Pattern Recogn Lett* 22:1525–1536
33. Han J, Awad G, Sutherland A, Wu H (2006) Automatic skin segmentation for gesture recognition combining region and support vector machine active learning. In: Proceedings of the IEEE international conference on automatic face and gesture recognition. IEEE Computer Society, Washington DC, USA, pp 237–242
34. Hanbury A (2003) A 3D-polar coordinate colour representation well adapted to image analysis. In: Proceedings of the Scandinavian conf on image analysis (SCIA). Springer, Berlin, pp 804–811
35. Hossain MF, Shamsi M, Alsharif MR, Zoroofi RA, Yamashit K (2012) Automatic facial skin detection using Gaussian mixture model under varying illumination. *Int J Innovative Comput Inf Control* 8(2):1135–1144
36. Hsu RL, Abdel-Mottaleb M, Jain A (2002) Face detection in color images. *IEEE Trans Pattern Anal Mach Intell* 24(5):696–706
37. Jagadesh BN, Rao K, Satyanarayana C, RajKumar GVS (2012) Skin colour segmentation using finite bivariate pearsonian type-IIb mixture model and k-means. *Signal Image Process Int J* 3(4):37–49
38. Jiang Z, Yao M, Jiang W (2007) Skin detection using color, texture and space information. *Proc Int Conf Fuzzy Syst Knowl Discov* 3:366–370
39. Jones M, Rehg J (1999) Statistical color models with application to skin detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), vol 1, pp 637–663
40. Jones M, Rehg J (2002) Statistical color models with application to skin detection. *Int J Comput Vis* 46:81–96
41. Kakumanu P, Makrogiannis S, Bourbakis NG (2007) A survey of skin-color modeling and detection methods. *Pattern Recogn* 40(3):1106–1122
42. Kakumanu P, Makrogiannis S, Bryll R, Panchanathan S, Bourbakis N (2004) Image chromatic adaptation using ANNs for skin color adaptation. In: Proceedings of the IEEE international conference on tools with artificial intelligence (ICTAI), pp 478–485
43. Kawulok M (2008) Dynamic skin detection in color images for sign language recognition. In: Proceedings of the ICISP, LNCS, vol 5099. Springer, pp 112–119
44. Kawulok M (2010) Energy-based blob analysis for improving precision of skin segmentation. *Multimedia Tools Appl* 49(3):463–481
45. Kawulok M (2012) Texture analysis for skin probability maps refinement. In: Proceedings of the MCPR, LNCS, vol 7329. Springer, pp 75–84
46. Kawulok M (2013) Fast propagation-based skin regions segmentation in color images. In: Proceedings of the IEEE international conference on automatic face and gesture recognition, FG, pp 1–7
47. Kawulok M, Kawulok J, Smolka B (2012) Discriminative textural features for image and video colorization. *IEICE Trans Inf Syst* 95-D(7):1722–1730
48. Kawulok M, Nalepa J (2012) Support vector machines training data selection using a genetic algorithm. In: Statistical techniques in pattern recognition, S+SSPR 2012, LNCS, vol 7626. Springer, pp 557–565
49. Kawulok M, Szymanek J (2012) Precise multi-level face detector for advanced analysis of facial images. *IET Image Process* 6(2):95–103
50. Khan R, Hanbury A, Sablatnig R, Stöttinger J, Khan F, Khan F (2012) Systematic skin segmentation: merging spatial and non-spatial data. In: Multimedia tools and applications, pp 1–25

51. Khan R, Hanbury A, Stöttinger J (2010) Skin detection: a random forest approach. In: Proceedings of the 17th IEEE international image processing (ICIP) conference, pp 4613–4616
52. Khan R, Hanbury A, Stöttinger J, Bais A (2012) Color based skin classification. *Pattern Recogn Lett* 33(2):157–163
53. Kovac J, Peer P, Solina F (2002) Eliminating the influence of non-standard illumination from images. Technical report
54. Kovac J, Peer P, Solina F (2003) Human skin color clustering for face detection. In: EUROCON 2003 computer as a tool, vol 2, pp 144–148
55. Krahenbuhl P, Koltun V (2011) Efficient inference in fully connected CRFs with Gaussian edge potentials. In: Proceedings of the neural information processing systems (NIPS)
56. Kruppa H, Bauer MA, Schiele B (2002) Skin patch detection in real-world images. In: Proceedings of the DAGM symposium on pattern recognition, LNCS, vol 2449. Springer, pp 109–117
57. Kukharev G, Nowosielski A (2004) Fast and efficient algorithm for face detection in colour images. *Mach Graph Vis* 13:377–399
58. Lam HK, Au O, Wong CW (2004) Automatic white balancing using standard deviation of RGB components. In: Proceedings of the international symposium on circuits and systems (ISCAS) vol 3, pp 921–924
59. Lee JS, Kuo YM, Chung PC, Chen EL (2007) Naked image detection based on adaptive and extensible skin color model. *Pattern Recogn* 40:2261–2270
60. Lichtenauer J, Reinders MJT, Hendriks EA (2007) A self-calibrating chrominance model applied to skin color detection. In: Proceedings of the VISAPP, vol 1, pp 115–120
61. Musicant DR, Feinberg A (2004) Active set support vector regression. *IEEE Trans Neural Netw* 15(2):268–275
62. Ng P, Pun CM (2011) Skin color segmentation by texture feature extraction and k-mean clustering. In: Proceedings of the 2011 3rd international conference on computational intelligence, communication systems and networks (CICSyN), pp 213–218
63. Nikolaidis A, Pitas I (2000) Robust watermarking of facial images based on salient geometric pattern matching. *IEEE Trans Multimedia* 2(3):172–184
64. Phung S, Bouzerdoum A, Chai D (2005) Skin segmentation using color pixel classification: analysis and comparison. *IEEE Trans Pattern Anal Mach Intell* 27(1):148–154
65. Phung SL, Bouzerdoum A, Chai D (2002) A novel skin color model in YCbCr color space and its application to human face detection. In: Proceedings of the international conference on image processing, vol 1, pp I-289–I-292
66. Phung SL, Chai D, Bouzerdoum A (2003) Adaptive skin segmentation in color images. In: Proceedings of the IEEE international conference on acoustics, speech and signal proceedings, pp 353–356
67. Rao K, Jagadesh BN, Satyanarayana C (2012) Skin colour segmentation using finite bivariate pearsonian type-IVa mixture model. *Comput Eng Intell Syst* 3(5):45–56
68. Ratnasingam S, McGinnity T (2012) Chromaticity space for illuminant invariant recognition. *IEEE Trans Image Process* 21(8):3612–3623
69. Schmugge SJ, Jayaram S, Shin MC, Tsap LV (2007) Objective evaluation of approaches of skin detection using roc analysis. *Comput Vis Image Underst* 108(1–2):41–51
70. Schohn G, Cohn D (2000) Less is more: active learning with support vector machines. In: Proceedings of the 17th international conference on machine learning, pp 839–846. Morgan Kaufmann Publishers Inc, USA
71. Seow MJ, Valaparla D, Asari V (2003) Neural network based skin color model for face detection. In: Proceedings of the applied imagery pattern recognition workshop, pp 141–145
72. Shin M, Chang K, Tsap L (2002) Does colorspace transformation make any difference on skin detection? In: Proceedings of the IEEE workshop on applications of computer vision (WACV), pp 275–279
73. Sigal L, Sclaroff S, Athitsos V (2003) Skin color-based video segmentation under time-varying illumination. *IEEE Trans Pattern Anal Machine Intell* 26:862–877

74. Sobottka K, Pitas I (1996) Face localization and facial feature extraction based on shape and color information. In: Proceedings of the IEEE international conference on image processing (ICIP), vol 3, pp 483–486
75. del Solar JR, Verschae R (2004) Skin detection using neighborhood information. In: Proceedings of the IEEE international conference on automatic face and gesture recognition, pp 463–468
76. Solina F, Peer P, Batagelj B, Juvan S (2002) 15 seconds of fame: an interactive, computer-vision based art installation. In: Proceedings of the international conference on control, automation, robotics and vision (ICARCV), vol 1, pp 198–204
77. Soriano M, Martinkauppi B, Huovinen S, Laaksonen M (2000) Skin detection in video under changing illumination conditions. In: Proceedings of the international conference on pattern recognition (ICPR), vol 1, pp 839–842
78. Stern H, Efros B (2002) Adaptive color space switching for face tracking in multi-colored lighting environments. In: Proceedings of the IEEE international conference on automatic face and gesture recognition (FG 2002). IEEE Computer Society, Washington DC, USA, pp 249–254
79. Stoerring M, Andersen HJ, Granum E, Granum E (1999) Skin colour detection under changing lighting conditions. In: Proceedings of the 7th symposium on intelligent robotics systems, pp 187–195
80. Sun HM (2010) Skin detection for single images using dynamic skin color modeling. *Pattern Recogn* 43(4):1413–1420
81. Tan WR, Chan CS, Yogarajah P, Condell J (2012) A fusion approach for efficient human skin detection. *IEEE Trans Ind Inf* 8(1):138–147
82. Taqa A, Jalab H (2010) Increasing the reliability of skin detectors. *Sci Res Essays* 5(17):2480–2490
83. Terrillon J-C, David M, Akamatsu S (1998) Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments. In: Proceedings of the 3rd international conference on automatic face and gesture recognition, pp 112–117, Nara, Japan
84. Tomaz F, Candeias T, Shahbazkia H (2003) Improved automatic skin detection in color images. In: Proceedings of the 7th digital computing: techniques and applications, pp 419–427
85. Tsekeridou S, Pitas I (1998) Facial feature extraction in frontal views using biometric analogies. In: Proceedings of the EUSIPCO '98, pp 315–318
86. Tu Y, Yi F, Chen G, Jiang S, Huang Z (2010) Skin color detection by illumination estimation and normalization in shadow regions. In: Proceedings of the IEEE international conference on information and automation (ICIA), pp 1082–1085
87. Vezhnevets V, Sazonov V, Andreeva A (2003) A survey on pixel-based skin color detection techniques. In: IN Proceedings of the GRAPHICON-2003, pp 85–92
88. Viola P, Jones M (2004) Robust real-time face detection. *Int J Comput Vis* 57(2):137–154
89. Wang X, Zhang X, Yao J (2011) Skin color detection under complex background. In: Proceedings of the international conference on mechatronic science, electric engineering and computer, pp 1985–1988
90. Yang G, Li H, Zhang L, Cao Y (2010) Research on a skin color detection algorithm based on self-adaptive skin color model. In: Proceedings of the international conference on communications and intelligence information security (ICCIIS), pp 266–270
91. Yang J, Fu Z, Tan T, Hu W (2004) Skin color detection using multiple cues. In: Proceedings of the international conference on image processing (ICPR), vol 1, pp 632–635
92. Yang MH, Ahuja N (1999) Gaussian mixture model for human skin color and its applications in image and video databases. In: ProcSPIE 99, CA, San Jose, pp 458–466
93. Yang U, Kang M, Toh KA, Sohn K (2010) An illumination invariant skin-color model for face detection. In: Proceedings of the IEEE international conference on biometrics: theory applications and systems (BTAS), pp 1–6
94. Yogarajah P, Condell J, Curran K, Cheddad A, McKevitt P (2010) A dynamic threshold approach for skin segmentation in color images. In: Proceedings of the IEEE international conference on image processing (ICIP), pp 2225–2228

95. Yogarajah P, Condell J, Curran K, McKevitt P, Cheddad A (2012) A dynamic threshold approach for skin segmentation in color images. *Int J Biometrics* 4(1):38–55
96. Yong-jia Z, Shu-ling D, Xiao X (2008) A Mumford-Shah level-set approach for skin segmentation using a new color space. In: Proceedings of the international conference on system simulation and scientific computing (ICSC), pp 307–310
97. Zafarifar B, Martiniere A, de With P (2010) Improved skin segmentation for TV image enhancement, using color and texture features. In: Proceedings of the international conference on consumer electronics (ICCE), pp 373–374
98. Zarit BD, Super BJ, Quek FKH (1999) Comparison of five color models in skin pixel classification. In: Proceedings of the international workshop on recognition, analysis, and tracking of faces and gestures in, real-time systems, pp 58–63
99. Zhang MJ, Gao W (2005) An adaptive skin color detection algorithm with confusing backgrounds elimination. In: Proceedings of the international conference on image processing (ICIP), vol 2, pp 390–393
100. Zhu Q, Cheng KT, Wu CT, Wu YL (2004) Adaptive learning of an accurate skin-color model. In: Proceedings of the IEEE international conference on automatic face and gesture recognition (FG 2004). IEEE Computer Society, Washington DC, USA, pp 37–42
101. Zhu Q, Wu CT, Cheng KT, Wu YL (2004) An adaptive skin model and its application to objectionable image filtering. In: Proceedings of the ACM international conference on multimedia (MULTIMEDIA '04). ACM, New York, USA, pp 56–63