Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 1106–1122
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

A survey of skin-color modeling and detection methods

P. Kakumanu, S. Makrogiannis, N. Bourbakis*

ITRI/Department of Computer Science and Engineering, Wright State University, Dayton OH 45435, USA

Received 2 February 2005; received in revised form 3 February 2006; accepted 5 June 2006

Abstract

Skin detection plays an important role in a wide range of image processing applications ranging from face detection, face tracking, gesture analysis, content-based image retrieval systems and to various human computer interaction domains. Recently, skin detection methodologies based on skin-color information as a cue has gained much attention as skin-color provides computationally effective yet, robust information against rotations, scaling and partial occlusions. Skin detection using color information can be a challenging task as the skin appearance in images is affected by various factors such as illumination, background, camera characteristics, and ethnicity. Numerous techniques are presented in literature for skin detection using color. In this paper, we provide a critical up-to-date review of the various skin modeling and classification strategies based on color information in the visual spectrum. The review is divided into three different categories: first, we present the various color spaces used for skin modeling and detection. Second, we present different skin modeling and classification approaches. However, many of these works are limited in performance due to real-world conditions such as illumination and viewing conditions. To cope up with the rapidly changing illumination conditions, illumination adaptation techniques are applied along with skin-color detection. Third, we present various approaches that use skin-color constancy and dynamic adaptation techniques to improve the skin detection performance in dynamically changing illumination and environmental conditions. Wherever available, we also indicate the various factors under which the skin detection techniques perform well.

Keywords: Skin-color modeling; Skin detection; Color spaces and color constancy

1. Introduction

Skin detection plays an important role in a wide range of image processing applications ranging from face detection, face tracking, gesture analysis, content-based image retrieval (CBIR) systems and to various human computer interaction domains. Recently, skin detection methodologies based on skin-color information as a cue has gained much attention as skin color provides computationally effective yet, robust information against rotations, scaling and partial occlusions. Skin color can also be used as complimentary information to other features such as shape and geometry and can be used to build accurate face detection systems [1–4]. Skin-color detection is often used as a preliminary step in face recognition, face tracking and CBIR systems. Skin-color information can be considered a very effective tool for identifying/classifying facial areas provided that the underlying skin-color pixels can be represented, modeled and classified accurately.

Most of the research efforts on skin detection have focused on visible spectrum imaging. Skin-color detection in visible spectrum can be a very challenging task as the skin color in an image is sensitive to various factors such as:

- *Illumination*: A change in the light source distribution and in the illumination level (indoor, outdoor, highlights, shadows, non-white lights) produces a change in the color of the skin in the image (color constancy problem). The illumination variation is the most important problem among current skin detection systems that seriously degrades the performance.
- *Camera characteristics*: Even under the same illumination, the skin-color distribution for the same person differs

* Corresponding author.
E-mail addresses: kpraveen@cs.wright.edu (P. Kakumanu), bourbaki@cs.wright.edu,nbourbakis@woh.rr.com, nikolaos.bourbakis@wright.edu (N. Bourbakis).

0031-3203/$30.00 © 2006 Published by Elsevier Ltd on behalf of Pattern Recognition Society.
doi:10.1016/j.patcog.2006.06.010

from one camera to another depending on the camera sensor characteristics. The color reproduced by a CCD camera is dependent on the spectral reflectance, the prevailing illumination conditions and the camera sensor sensitivities.
- *Ethnicity*: Skin color also varies from person to person belonging to different ethnic groups and from persons across different regions. For example, the skin color of people belonging to Asian, African, Caucasian and Hispanic groups is different from one another and ranges from white, yellow to dark.
- *Individual characteristics*: Individual characteristics such as age, sex and body parts also affects the skin-color appearance.
- *Other factors*: Different factors such as subject appearances (makeup, hairstyle and glasses), background colors, shadows and motion also influence skin-color appearance.

Many of the problems encountered in visual spectrum can be overcome by using non-visual spectrum such as infrared (IR) [5,6] and spectral imaging [7–9]. Skin-color in non-visual spectrum methods is invariant to changes in illumination conditions, ethnicity, shadows and makeup. However, the expensive equipment necessary for these methods combined with tedious setup procedures have limited their use to specific application areas such as biomedical applications. In this paper, we concentrate on visual spectrum-based skin detection techniques that are applicable for 2D images or single frames of video.

From a classification point of view, skin detection can be viewed as a two class problem: skin-pixel vs. non-skin-pixel classification. The primary steps for skin detection in an image using color information are (1) to represent the image pixels in a suitable color space, (2) to model the skin and non-skin pixels using a suitable distribution and (3) to classify the modeled distributions. Several color spaces have been proposed and used for skin detection in the literature. The choice of color space also determines how effectively we can model the skin-color distribution. Skin-color distribution is modeled primarily either by histogram models or by single/Gaussian mixture models. Several techniques on skin-color model classification, ranging from simple look-up table approaches to complex pattern recognition approaches have been published. A survey of different color spaces for skin-color representation and skin-pixel detection methods is given by Vezhnevets et al. [10]. However, a comprehensive survey of the up-to-date techniques on skin-color modeling and classification is still missing. The goal of this paper is to provide a critical review of different skin-color modeling and detection methods. Many of the existing skin detection strategies are not effective when the illumination conditions vary rapidly. To cope with changes in illumination conditions and viewing environment only few skin detection strategies use color constancy and dynamic adaptation techniques. In color constancy approaches, the images are first color corrected based on an estimate the illuminant color. The skin-color modeling and detection are then applied on these preprocessed images. In dynamic adaptation techniques, the existing skin-color model is transformed to the changing illumination conditions. An up-to-date review of such illumination adaptation approaches for skin detection is also presented. Majority of the reported works on skin detection perform well only for a limited set of illumination conditions and skin types. Wherever available, we indicate the different factors under which the reported strategies perform well. We also report the skin detection performances in terms of true and false positives, if available. True positive represents the number of skin pixels correctly classified as skin pixels while false positive represents the number of non-skin pixels classified as skin pixels.

The remainder of the paper is organized as follows: Section 2.1 gives a brief description of the popular color spaces used in skin-color detection. Skin modeling and classification techniques are described in Section 2.2. Skin detection strategies that use color constancy and dynamic adaptation techniques are reviewed in Sections 3. Section 4 provides summary and conclusions.

2. Skin-color modeling and classification

2.1. Color spaces

The choice of color space can be considered as the primary step in skin-color classification. The RGB color space is the default color space for most available image formats. Any other color space can be obtained from a linear or non-linear transformation from RGB. The color space transformation is assumed to decrease the overlap between skin and non-skin pixels thereby aiding skin-pixel classification and to provide robust parameters against varying illumination conditions. It has been observed that skin colors differ more in intensity than in chrominance [11]. Hence, it has been a common practice to drop the luminance component for skin classification. Several color spaces have been proposed and used for skin detection. In this section, we review the most widely used color spaces for skin detection and their properties.

2.1.1. Basic Color Spaces (RGB, normalized RGB, CIE–XYZ)

RGB is the most commonly used color space for storing and representing digital images, since the data captured by a camera is normally provided as RGB. RGB correspond to the three primary colors: red, green and blue, respectively. To reduce the dependence on lighting, the RGB color components are normalized so that sum of the normalized components is unity ($r + g + b = 1$). Since the sum of these components is 1, the third component does not hold any significant information and is normally dropped so as to obtain a reduction in dimensionality. It has been observed that under certain assumptions, the differences in skin-color pixels due to lighting conditions and due to ethnicity can be greatly reduced in normalized RGB (*rgb*) space. Also, the skin-color clusters in *rgb* space have relatively lower variance than the corresponding clusters in RGB and hence are shown to be good for skin-color modeling and detection [11,12]. Due to the above advantages, *rgb* has been a popular choice for skin-detection and has been used by Bergasa et al. [13], Brown et al. [14], Caetano and Barone [15], Oliver et al. [16], Kim et al. [17], Schwerdt and Crowley [18], Sebe et al. [19], Soriano et al. [20], Störring et al. [21], Wang and Sung [22], Yang and Ahuja [12], Yang et al. [11]. The CIE (*Commission Internationale de l'Eclairage*) system describes color as a luminance component *Y*, and two additional components *X* and *Z*. CIE–*XYZ* values were constructed from psychophysical experiments and correspond to the color matching characteristics of human visual system [23]. This color space has been used by Brown et al. [14], Chen and Chiang [24], Wu et al. [25].

2.1.2. Perceptual color spaces (HSI, HSV, HSL, TSL)

The perceptual features of color such as hue (H), saturation (S) and intensity (I) cannot be described directly by RGB. Many non-linear transformations are proposed to map RGB on to perceptual features. The HSV space defines color as *Hue*—the property of a color that varies in passing from red to green, *Saturation*—the property of a color that varies in passing from red to pink, *Brightness* (also called *Intensity* or *Lightness* or *Value*)—the property that varies in passing from black to white. The transformation of RGB to HSV is invariant to high intensity at white lights, ambient light and surface orientations relative to the light source and hence, can form a very good choice for skin detection methods. The HSV color space has been used by Brown et al. [14], Garcia and Tziritas [26], McKenna et al. [27], Saxe and Foulds, [28], Sobottka and Pitas [29], Thu et al. [30], Wang and Yuan [31], Zhu et al. [32]. A variation of the HS components using logarithmic transformation, called Fleck HS was introduced by Fleck and has been used by Zarit et al. [33]. Another similar color space is TSL color space which defines color as *Tint*—hue with white added, *Saturation* and *Lightness*. The TSL color space has been used by Terillon et al. [34], Brown et al. [14].

2.1.3. Orthogonal color spaces (YCbCr, YIQ, YUV, YES)

The orthogonal color spaces reduce the redundancy present in RGB color channels and represent the color with statistically independent components (as independent as possible). As the luminance and chrominance components are explicitly separated, these spaces are a favorable choice for skin detection. The YCbCr space represents color as luminance (*Y*) computed as a weighted sum of RGB values, and chrominance (Cb and Cr) computed by subtracting the luminance component from B and R values. The YCbCr space is one of the most popular choices for skin detection and has been used by Hsu et al. [35], Chai and Bouzerdoum [36], Chai and Ngan [37], Wong et al. [38]. A variant of YCbCr called YCgCr was used by deDios and Garcia [39]. This new color space differs from YCbCr in the usage of Cg color component instead of the Cb component and was reported to be better than YCbCr. Other similar color spaces in this category include YIQ, YUV and YES, and represent color as luminance (*Y*) and chrominance. For skin detection, these color spaces has been used by YIQ—Dai and Nakano [40], YUV—Marques and Vilaplana [41], YES—Saber and Tekalp [42], Gomez et al. [43].

2.1.4. Perceptually uniform color spaces (CIE-Lab and CIE-Luv)

Perceptual uniformity represents how two colors differ in appearance to a human observer and hence uniform color spaces (UCS) were defined such that all the colors are arranged by the perceptual difference of the colors. However, the perceptual uniformity in these color spaces is obtained at the expense of heavy computational transformations. In these color spaces, the computation of the luminance (L) and the chroma (*ab* or uv) is obtained through a non-linear mapping of the *XYZ* coordinates. For skin detection, the CIE-Lab space has been used by Cai and Goshtasby [44], Kawato and Ohya [45]. The CIE-Luv space has been used by Yang and Ahuja [12]. Farsnworth [46] has proposed a more perceptually uniform system than Lab or Luv, and it has been used by Wu et al. [25].

2.1.5. Other color spaces

It has been observed that skin contains a significant level of red. Hence, some researchers [43,47] have used color ratios (e.g., R/G) to detect skin. Gomez et al. [43] have used an attribute selection approach to select different complimentary color components from various color spaces. They showed that the mixture of color components (E of YES, the ratio R/G and H from HSV) performed better than the existing color spaces for indoor and outdoor scene images. Also, the authors argue that this new mixture space is not-sensitive to noise from a wide range of unconstrained sources and illumination conditions. Brand and Mason [47] have evaluated the performance of color ratios with other algorithms on the Compaq data set [48]. They concluded that the combination of color ratios ($R/G + R/B + G/B$) provided a better response than the individual ratios.

2.2. Skin-color classification

From a classification point of view, skin-color detection can be viewed as a two class problem: skin-pixel vs. non-skin-pixel classification. Different researchers have used different techniques to approach this problem. The following section gives a brief description of the most common methods used.

2.2.1. Explicit skin-color space thresholding

The human skin colors of different individuals cluster in a small region in color space provided that the images are taken under illumination controlled environments [11]. Hence, one of the easiest and often used methods is to define skin-color cluster decision boundaries for different color space components. Single or multiple ranges of threshold values for each color space component are defined and the image pixel values that fall within these predefined range(s) for all the chosen color components are defined as skin pixels.

Dai and Nakano [40] used a fixed range on I component in YIQ space for detecting skin pixels from images containing mostly people with yellow skin. The I component includes colors from orange to cyan. All the pixel values in the range, $R_I = [0, 50]$ are described as skin pixels in this approach. Sobottka and Pitas [29,49] used fixed range values on the HS color space. The pixel values in the range $R_{\mathrm{H}} = [0, 50]$ and $R_{\mathrm{S}} = [0.23, 0.68]$ are defined as skin pixels. These values have been determined to be well suited for discriminating skin pixels from non-skin pixels on the M2VTS database, containing images of yellow and white skin people. Chai and Ngan [50] proposed a face segmentation algorithm in which they used a fixed range skin-color map in the CbCr plane. The pixel values in the range $R_{\mathrm{Cb}} = [77, 127]$, and $R_{\mathrm{Cr}} = [133, 173]$ are defined as skin pixels on the ECU face and skin database. Garcia and Tziritas [26] segmented skin by using eight planes in YCbCr space or by using six planes in HSV space. Wang and Yuan [31] have used threshold values in rg space and HSV space. The threshold values in the range $R_r = [0.36, 0.465]$, $R_g = [0.28, 0.363]$, $R_{\mathrm{H}} = [0, 50]$, $R_{\mathrm{S}} = [0.20, 0.68]$ and $R_{\mathrm{V}} = [0.35, 1.0]$ are used for discriminating skin and non-skin pixels. Yao and Gao [101] first transformed YUV space to skin chrominance and lip chrominance spaces and then used fixed range values on these spaces to detect skin and lip pixels, respectively. Wong et al. [38] used thresholding on the luminance level, Y in YCbCr space. Tomaz et al. [51] also used thresholding in TS space. In Gomez and Morales [52], the authors start with rgb and the constant $\frac{1}{3}$. A constructive induction algorithm is used to construct a number of three component decision rules from these four components through simple arithmetic operations. The new color representation spaces thus found had better performance in precision and success rate than skin probability maps (SPM) (Section 2.2.2) though it is computationally slower than the SPM method.

2.2.2. Histogram model with naı¨ve bayes classifiers

In this method, a 2D or 3D color histogram is used to represent the distribution of skin tones in color space. Color histograms are stable object representations unaffected by occlusion, changes in the view, and can be used to differentiate a large number of objects [11]. The color space is quantized into a number of histogram bins. Each histogram bin (also defined as look-up table cell) stores the count associated with the occurrence of the bin color in the training data set. The histogram bin counts are converted into probability distribution, $P(c)$ as follows:

$$P(c) = \frac{count(c)}{T}, \tag{1}$$

where $count(c)$ gives the count in the histogram bin associated with color c and T is the total count obtained by summing the counts in all the histogram bins. These values correspond to the likelihood that a given color belongs to the skin. All the pixel values for which the corresponding color likelihood is greater than a predefined threshold are defined as skin pixels. Zarit et al. [33], Yoo and Oh [53] used a histogram-based approach to classify skin pixels.

Jones and Rehg [54], built a 3D RGB histogram model with two billion pixels collected from 18,696 web images. They reported that 77% of the possible RGB colors are not encountered and most of the histogram is empty. There is a marked skew in the distribution towards the red corner of the color cube due to the presence of skin in web images, though only 10% of the total pixels are skin pixels. This suggested that skin colors occur more frequently than other object colors. Jones and Rehg also computed two different histograms, skin and non-skin histograms. Given skin and non-skin histograms, the probability that a given color belongs to skin and non-skin class (also called class conditional probabilities) is defined as

$$P(c/skin) = \frac{s(c)}{T_s}, \quad P(c/non\text{-}skin) = \frac{n(c)}{T_n}, \tag{2}$$

where $s(c)$ is the pixel count in the color c-bin of the skin histogram, $n(c)$ is the pixel count in the color c-bin of the non-skin histogram. T_s and T_n represents the total counts in the skin and non-skin histogram bins. From the generic skin and non-skin histograms, Jones and Rehg demonstrated that there is a reasonable separation between skin and non-skin classes. This fact can be used to build fast and accurate skin classifiers even for images collected from unconstrained imaging environments such as web images, given that the training dataset is sufficiently huge. (A larger training set can lead to better probability density function estimations.) Given the class conditional probabilities of skin and non-skin-color models, a skin classifier can be built using Bayes maximum likelihood (ML) approach [55]. Using this, a given image pixel can be classified as skin, if

$$\frac{P(c/skin)}{P(c/non\text{-}skin)} \geqslant \Theta, \tag{3}$$

where $0 \leqslant \Theta \leqslant 1$ is a threshold value which can be adjusted to trade-off between true positives and false positives. This threshold value is normally determined from the ROC (receiver operating characteristics) curve calculated from the training data set. The ROC curve shows the relationship between the true positives and false positives as function of the detection threshold Θ. The histogram-based Bayes classifier (also called as skin probability map, SPM in short) has been widely used for skin segmentation. The method is simple and computationally fast as we need only two table look-ups to compute the probability of the skin. This has been used by Brand and Mason [47], Chai and Bouzerdoum [36], Fleck et al. [56], Gomez and Morales [52], Gomez et al. [43], Jones and Rehg [54], Marcel and Benigo [57], Schwerdt and Crowley [18], Sigal et al. [95], Srisuk and Kuritach [58], Zarit et al. [33].

2.2.3. Gaussian classifiers

Many of the representative works on skin-color distribution modeling have used Gaussian mixtures. The advantage of these parametric models is that they can generalize well with less training data and also have very less storage requirements.

2.2.3.1. Single Gaussian models (SGM) Under controlled illuminating conditions, skin colors of different individuals cluster in a small region in the color space. Hence, under certain lighting conditions, the skin-color distribution of different individuals can be modeled by a multivariate normal (Gaussian) distribution in normalized color space [11,59]. Skin-color distribution is modeled through elliptical Gaussian joint probability distribution function (pdf), defined as

$$p(c) = \frac{1}{(2\pi)^{1/2}|\sum|^{1/2}} \exp\left[-\frac{1}{2}(c-\mu)^{\mathrm{T}}\sum\nolimits^{-1}(c-\mu)\right], \tag{4}$$

where c is a color vector, μ and $\sum$ are the mean vector and the diagonal covariance matrix, respectively.

$$\mu = \frac{1}{n}\sum_{j=1}^{n} c_j, \quad \sum = \frac{1}{n-1}\sum_{j=1}^{n}(c_j-\mu)(c_j-\mu)^{\mathrm{T}}. \tag{5}$$

The parameters, μ and $\sum$ are estimated over all the color samples (c_j) from the training data using ML estimation approach. The probability $p(c)$ can be used directly as a measure of skin-color likeliness and the classification is normally obtained by comparing it to a certain threshold value estimated empirically from the training data [17,60,61]. Alternatively, we can also compare Mahalanobis distance λ, from the image pixel color c [35,62–64] to a certain threshold. This threshold value is determined from the ROC curve (Section 2.2.2) calculated from the training data.

$$\lambda = (c-\mu)^{\mathrm{T}}\sum\nolimits^{-1}(c-\mu). \tag{6}$$

2.2.3.2. Gaussian mixture models (GMM) Though the human skin-color samples for people of different races cluster in a small region in the color space, it has been shown that different modes co-exist within this cluster and hence it cannot be modeled effectively by a single Gaussian distribution [12]. Also, under varying illuminating conditions, the single mode assumption does not hold. Many researchers, therefore, have used Gaussian mixtures—a more capable model to describe complex shaped distributions. A Gaussian mixture density function is the sum of individual Gaussians, expressed as

$$p(c) = \sum_{i=1}^{N} w_i \frac{1}{(2\pi)^{1/2}|\sum_i|^{1/2}} \times \exp\left[-\frac{1}{2}(c-\mu_i)^{\mathrm{T}}\sum_i^{-1}(c-\mu_i)\right], \tag{7}$$

where c is a color vector and μ_i and $\sum_i$ are the mean and the diagonal covariance matrix. N is the number of Gaussians and the weight factor, w_i is the contribution of the ith Gaussian. The parameters of a GMM (μ_i, $\sum_i$, and w_i) are approximated from the training data through the iterative expectation-maximization (EM) technique. To converge well, the EM technique needs a good initial guess of the parameters. These initial parameters can be obtained by *k-means* clustering of the training data. A detailed description of the EM technique applied to skin GMM is described in Yang and Ahuja [12]. The skin classification is performed using the same methodology as described in the SGM case.

The choice of the number of Gaussian components N is very critical to the training data and the choice of color space. Different researchers have used different values of N ranging from 2 to 16. Yang and Ahuja [12] used two Gaussians in Luv color space on a Michigan face database. They provided statistical tests to justify their hypothesis that a SGM is not sufficient to model the skin distribution for the dataset considered. Greenspan et al. [65] also provided statistical tests to show that a GMMs is a robust representation that can accommodate large variations in color space, highlights and shadows. In their trained two component GMM, one of the components captured the distribution of the normal light skin color while the other captured the distribution of the more highlighted regions of the skin. Caetano et al. [66] used 2.8 Gaussians in *rg* color space on images containing both black and white people. Lee and Yoo [63] used six Gaussians for skin classification on a compaq dataset whereas Thu et al. [30] used four Gaussian components. The key idea behind using multiple components is that different parts of the face illuminated in a different manner can be detected by different components. Jones and Rehg [54] trained two separate models for the skin and non-skin classes. They used 16 Gaussians in each of these models. Cai and Goshtasby [44], Jebara and Pentland (1998), McKenna et al. [27], Oliver et al. [16] also used GMMs for skin classification.

2.2.4. Elliptical boundary model

As an alternative to the computationally intensive GMMs, Lee and Yoo [63] proposed an "elliptical boundary model", whose performance is comparable to that of GMM and yet the computational complexity is as simple as training a SGM. The elliptical boundary model is defined as

$$\Phi(c) = [c-\Psi]^{\mathrm{T}}\Lambda^{-1}[c-\Psi], \tag{8}$$

where c is the color vector, Ψ and Λ are the model parameters defined as

$$\Psi = \frac{1}{n}\sum_{i=1}^{n} c_i, \quad \Lambda = \frac{1}{N}\sum_{i=1}^{n} f_i(c_i-\mu)(c_i-\mu)^{\mathrm{T}}, \tag{9}$$

where N is the total number of samples in the training data set, f_i is the number of samples with chrominance c_i and μ is the mean of the chrominance vectors in the training data set. The pixel with chrominance c is classified as a skin pixel, if $\Phi(c) < \phi$, where ϕ is a threshold value chosen empirically as a trade-off between the true and false positives. The authors compared the performance of the proposed model with SGM and GMM with six components on Compaq database. The elliptical boundary model performed slightly better than that of GMM. However, the drawback of this model is that its usage is limited to binary classification only.

2.2.5. Multi layer perceptron (MLP) classifier

MLPs have been successfully applied in many pattern recognition problems due to their ability to learn complex non-linear input–output relationships and ability to generalize any given data. MLPs, the most popular type of artificial neural networks (ANN), are feed forward networks of simple processing elements or neurons. The weights in the network are updated iteratively through a gradient descent technique known as back-propagation algorithm. At each iteration in the training process, the MLP processes all of its inputs in a feed-forward fashion, compares the resulting outputs with the expected ones and back-propagates these errors to adjust each weight in the network according to its contribution to the overall error. However, one disadvantage of NN approach is that the performance of the network is dependent on several factors such as the number of hidden layers, the number of hidden nodes and learning rates and to get an optimal performance the network has to be extensively tuned.

In MLP-based skin classification, a NN is trained to learn the complex class conditional distributions of the skin and non-skin pixels. Chen and Chiang [24] used a three layered feed forward NN in CIE-*xy* space for skin classification. Karlekar and Desai [67], Phung et al. [68] used MLP in CbCr space for skin classification. Sahbi and Boujemaa [69] trained a three layered NN in RGB space on a dataset of web images. The NN classifies the image regions as a collection of either skin or non-skin regions. The skin classification obtained from NN is subjected to further fine tuning using a Gaussian model. The parameters of these models are evaluated using a fuzzy clustering approach. One of the major issues with training pattern recognition systems for skin and non-skin classification is to collect good representable samples for both skin and non-skin. Collecting skin samples can be easy; however collecting non-skin samples is extremely harder. To overcome this problem, Seow et al. [70] trained a three layered NN in RGB space not only to extract the skin regions but also to interpolate the skin regions in 3D color cube. The NN interpolated area of the color cube is considered as skin region and the rest as non-skin region. Anagnostopoulos et al. [71] used fuzzy logic (FL) chromatic rules in RGB space on frontal face images to detect skin regions. The output from FL rules is fed to a probabilistic neural network (PNN) to classify the detected regions as faces or non-faces.

2.2.6. Self organizing map (SOM) classifier

The self-organizing map (SOM) (also known as Kohonen feature map) is one of the most popular ANN models. The SOM algorithm is based on unsupervised, competitive learning. The SOM constructs a topology preserving mapping from the high-dimensional space onto map units in such a way that relative distances between data points are preserved. The map units, or neurons, usually form a two-dimensional regular lattice where the location of a map unit carries semantic information. The SOM can thus serve as a clustering tool of high-dimensional data, although it works equally as well with low-dimensional data. Another important feature of the SOM is its capability to generalize.

Brown et al. [14] trained two separate SOMs to learn skin-color and non-skin-color pixel distributions on a dataset of over 500 images. They also compared the performance of SOMs against GMMs and on various color spaces (HS, XY, TS and *rg*). The SOMs consistently performed better than GMMs. The results showed that the choice of color space is more critical to the effective performance of the GMM than that of SOM. However, the performance of the SOMs on Compaq database (only 0.25% of the total skin samples were used for training) is comparatively lower than that of the results reported in Jones and Rehg [54]. The authors claim that using a larger training data and larger number of neurons might improve the performance of the SOM.

2.2.7. Maximum entropy (MaxEnt) classifier

Maximum entropy (MaxEnt) modeling is a statistical method for estimating probability distributions from data. The basic principle is that when the knowledge about the distribution of the data is minimum, the distribution should be as uniform as possible, that is, have maximal entropy. Labeled training data is used to derive a set of constraints for the model that characterize the class-specific expectations for the distribution. Constraints are represented as expected values of features. The parameters of the model are estimated using a ML approach. Max Ent models have been successfully applied for many applications related to speech recognition, audio-visual speech processing and natural language processing applications.

Jedynak et al. [72] used a Max Ent model for skin-pixel classification. The color and skinness for one pixel and two adjacent pixels are chosen as relevant features. The histogram of these features is computed on the Compaq dataset. The parameters of MaxEnt model are then updated using Beth Tree approximation, which consists of approximating the pixel lattice by a tree locally. The parameters of the

MaxEnt model are then expressed analytically as functions of histograms of the features. Finally, Gibbs sampler algorithm was used for inferring the probability of skin for each pixel. The details of the procedure can be obtained from Jedynak et al. [72]. However, the number of parameters in this model is very large. Assuming that a color in RGB space can take 256 values, the total number of parameters in this model is $256^3 * 256^3 * 2 * 2$. Hence the training of the model is very time consuming. To reduce the training time, the authors propose the use of belief propagation algorithm [73]. On Compaq database, the detection rate of this model is 82.9%, at a false positive rate of 10%.

2.2.8. *Bayesian network (BN) classifier*

Bayesian networks are directed acyclic graphs that allow efficient and effective representation of the joint probability density functions. Each vertex in the graph represents a random variable, and edges represent direct correlations between the variables [74]. Two examples of popular Bayesian Network classifiers are the Naı¨ve Bayes (NB) classifier and the Tree-Augmented Naı¨ve Bayes (TAN) classifier. The NB classifier assumes that all the features are conditionally independent given the class label, though this is not always true. To enhance the performance over the NB classifier, Friedman et al. [74] proposed the use of TAN classifier which also considers the correlations between the variables. For learning the TAN classifier, we do not fix the structure of the Bayesian network, but we try to find the TAN structure that maximizes the likelihood function given the training data out of all the possible TAN structures. In general, searching for the best tree structure has no efficient solution. However, Friedman et al. [74] showed that searching for the best TAN structure can be computed in polynomial time.

Sebe et al. [19] used a BN for skin modeling and classification. One of the problems with pattern recognition approaches is the availability of labeled training data. The authors propose a new method for learning the structure of the BN with labeled and unlabeled data. They used a stochastic structure search (SSS) algorithm for learning the structure of the BN, which improves the performance over NB and TAN classifiers. The details of the procedure can be obtained from Sebe et al. [19]. With a training data of only 60,000 samples (600 labeled + 54,000 unlabeled) from Compaq dataset, the proposed approach has detection rates of 95.82%, 98.32% with only 5% and 10% false positives, respectively.

2.3. *Comparison of skin-color classifiers*

A good skin classifier should be able to detect different skin types (white, pink, yellow, brown and dark) under a wide variety of illumination conditions (white, non-white, shadows, indoor and outdoor) and different backgrounds. Many of the skin detection techniques consider only a few skin types and a few possible illumination conditions. Table 1 lists the summary of the skin detection strategies discussed in Section 2.2. It also lists the various illumination conditions and the different skin types reported for evaluating the performances of these classifiers. The performances of the skin classifiers in terms of true positive rate (TPR) and false positive rate (FPR) are also reported. Many of the reported performances were on different datasets. Hence, it is not possible to obtain a fair evaluation of all these methods, as they are not evaluated on common representative train and test datasets.

Some of the works have reported the performances on the Compaq and ECU skin/non-skin datasets. The Compaq dataset [54] consists of 13,640 color images collected from World Wide Web. Out of these, 4675 images contain skin pixels, while the remaining 8965 images do not contain skin pixels. Since the images were collected from web, these images contain skin pixels belonging to persons of different origins and with un-constrained illumination and background conditions. The ECU database [75] consists of 4000 color images: about 1% of these images were taken with digital camera and the reset were collected manually from the Web. The lighting conditions in these images include indoor and outdoor lighting with varied backgrounds, and the skin types include whitish, brownish, yellowish and darkish skins.

The decision boundaries in the explicit thresholding method are fixed and are determined empirically from the skin-color images. On the Compaq database, the thresholding of *I*-axis in YIQ has a detection rate of 94.7 (FPR—30.2) [14]. On the ECU database, thresholding in YCbCr space has a detection rate of 82.0 (FPR—18.7) [75]. It should be noted that this method has good skin detection rate at the expense of high false positives. The explicit thresholding method has the advantage of being simple and fast. However, it has many limitations. The fixed threshold values differ from one color space to another and differ from one illumination to another. It is very difficult to find a range of threshold values that covers all the subjects of different skin color. This technique is less accurate in case of shadows and situations where the skin color is not distinguishable from background. In most of the situations, the explicit thresholding technique for skin segmentation is less precise and hence normally followed by a dynamic adaptation approach (Section 3.2).

The performance in the histogram technique is affected by the degree of overlap between the skin and non-skin classes in a given color space and the choice of the detection threshold. The histogram technique has detection rates of 90.0 (FPR—14.2) and 88.9 (FPR—10) on Compaq and ECU databases, respectively. As we can observe from the table, these performance rates are slightly higher than that of the GMM or the MLP techniques on the corresponding databases. However, due to its inability to interpolate and generalize the data, the histogram method needs a very large training dataset to get a good classification rate. Also, this method has higher storage requirements. For example, a 3D RGB histogram with 256 bins per channel requires 2^{48} bytes of storage, assuming one 4 byte integer per bin. One of

Table 1
Performance of different skin detection strategies

Authors	Color space	Intensity comp.	Skin detection method	Pre-train	Test database	Diff. skin types	Diff. illum.	Skin types	Illum. types	True and false positives	
Jones and Rehg, 02	RGB	Yes	Bayes	No	Compaq	Yes	Yes	n/a	Un-constrained	**90**	**14.2**
	RGB	Yes	GMM (16)	Yes	(13640 web					**90**	**15.5**
Brown et al., 01	TSL	No	SOM	Yes	images, 4675					78	32
Jedynak et al., 03	RGB	Yes	MaxEnt. model	Yes	skin + 8965					82.9	10
Lee and Yoo, 02	Xyz	No	Ellip. model	Yes	non-skin					90	20.9
	YCbCr	No	SGM	Yes	images)					90	33.3
	YIQ	No	GMM (6)	Yes						90	30
Brand et al., 00	RGB	Yes	Bayes	No						93.4	19.8
	YIQ	Yes	I-axis Thresh.	No						94.7	30.2
	RGB	Yes	Thresh. ratios	No						94.7	32.3
Sebe and Huang, 04	Rgb	No	BN	Yes						**99.4**	**10**
Fu and Yang, 04	HSV	Yes	GMM(14) + Hist. Merging	Yes						n/a	n/a
Phung et al., 05	YCbCr	Yes	Thresholding	No	ECU (4000	Yes	Yes	White	Indoor,	82.0	18.7
	RGB	Yes	Bayes	No	images, 1%			pink	outdoor,	**88.9**	**10**
	RGB	Yes	MLP	Yes	from digital			yellow	diff.	**88.5**	**10**
	YCbCr	Yes	SGM	Yes	cam. + rest			brown	backgrounds	88.0	10
	YCbCr	Yes	GMM	Yes	from web)			dark		85.2	10
Anagnostopoulos et al., 03	RGB	Yes	Fuzzy rules + PNN	Yes	317 images	n/a	n/a	n/a	n/a	82.4	n/a
Yang and Ahuja, 99	Luv	No	GMM (2)	Yes	Michigan Face Database	Yes	n/a	n/a	n/a	n/a	n/a
Caetano et al., 02	Rgb	No	SGM	Yes	800 images	Yes	Yes	White	Un-constrained	n/a	n/a
Caetano et al., 02	Rgb	No	GMM (2)	Yes	(web + Stirling Face Data)			yellow brown dark		87	30
Seow et al., 03	RGB	Yes	NN	Yes	ODU	Yes	Yes	n/a	n/a	n/a	n/a
Greenspan and Goldberger, 01	Rgb	No	GMM (2)	Yes	682 images (AR + AH + Video)	Yes	Yes	White yellow pink	Controlled	n/a	n/a
Thu and Meguro, 02	HSV	Yes	GMM (4) + Multi-Thresh.	Yes	n/a	Yes	Yes	White yellow dark	Indoor, outdoor, shadows	n/a	n/a
Jayaram et al., 04	SCT	Yes	Bayes	No	805 images	Yes	Yes	White	Indoor	98.2	n/a
	SCT	Yes	SGM	Yes	(AR + UOPB + UW CBIR Database)			yellow brown dark	white, non-white	94.4	n/a
Marcel and Benigo, 02	RGB	Yes	Histogram + MLP	Yes	XM2VTS	Yes	Yes	n/a	Controlled	n/a	n/a

the other important factors to consider in this method number of histogram bins. A 3D RGB histogram with 256 bins per channel has 256^3 bins. Due to the sparse distribution of skin points in RGB space, we reduce the color cube size, thereby reducing the number of bins in the histogram. This helps in creating a more compact histogram. Jones and Rehg [54] found that a 32-bin histogram performed better than a 256-bin RGB histogram on the Compaq dataset. However, the number of bins that gives the best performance varies with the color space representation and the size of the training dataset. Phung et al. [75] found that on ECU database, a 256-bin histogram is more sensitive to the size of the training data than a 32-bin histogram. The 256-bin histogram performed better for large datasets. However, they also reported that the performance of 64, 128 and 256-bin histograms are almost similar.

Caetano et al. [62] compared the performance of SGM and GMMs (2–8 components) on a dataset of 800 images containing people from a large spectrum of ethnic groups. The skin models were trained using 550 images collected from web and Stirling Face Database. For the dataset considered, the performance of GMMs with different components was similar. The performance of SGM is similar to those of the GMMs for low FPR. However, for medium and high TPR, the GMMs performed better. From these results, they suggested that the mixture models may be more appropriate than single Gaussian model when high correct detection rates are required. Similar results were obtained by Lee and Yoo [63] on Compaq dataset. However, as we can observe form the table, on ECU dataset, the performance of Gaussian mixture is lower than that of unimodal Gaussian. In this case, the Gaussian mixture was not trained on ECU dataset, but the parameters were taken from that of Jones and Rehg [54]. It should be noted that the histogram models were found to be slightly superior to GMMs in terms of skin-pixel classification performance [54,75,76]. However, GMMs have been a popular choice for skin-color segmentation as they can generalize very well with less training data. The mixture models though approximate the skin-color distribution effectively, the initialization and iterative training of the GMM are computationally expensive especially with large data sets. Also, the mixture model is slower to use during classification since all the Gaussians must be evaluated in computing the probability of a single color value. To reduce the train and recall costs of the GMMs, Fu et al. [77] proposed the use of multidimensional histograms in the EM framework to group the neighboring data points and reduce the size of the data set. The conventional EM algorithm is computationally expensive as it considers individual data points one by one exhaustively. However, since overlapping or closely distributed models data points have equal probabilities, we can group all these points and treat the neighboring points as a single data point with multiple occurrences. To represent the data set the authors use multidimensional histograms and the EM algorithm is modified accordingly. This method reduces the training time drastically. For example, Jones and Rehg [54] spent about 24 hours to train the skin and non-skin GMMs with 16 components on Compaq database using 10 work stations in parallel. With the multidimensional histogram technique of Fu et al. [77], a 15-component skin GMM took only 250 s. on Pentium IV 3.2 GHz workstation to train on the same Compaq database.

In Phung et al. [75], the authors compared the performance of MLP network with Bayesian, Gaussian and Explicit threshold skin classifiers on ECU database. As we can observe from Table 1, the Bayesian and MLP classifiers have similar performance and outperform the other classifiers. However, when compared with Bayesian techniques, the MLP has very low storage requirements and hence is a better candidate when storage is also a consideration. On the Compaq dataset, the Bayesian network of Sebe et al. [19] have superior performance. This method also has the advantage of using very low-labeled data.

To characterize one of the methods described in Table 1 as a winner, we need to consider many other factors such as the sizes of the train and test sets used. The size of the training set and the variety of samples used in the training set have a direct impact on the classifier performance. Also, some skin classification methods require extensive training and fine tuning of the parameters to achieve an optimal performance. The training time is often ignored. However, it may be important for real-time applications that require online training on different data sets. It should be noted that the evaluation criteria are dependent on the purpose of the skin classifier. For example, if the skin classifier is used as pre processing step in face detection, then the system may prefer achieving high true positives at the expense of high false positives. In such systems, the false positives can be reduced as described in Section 3 or by using multiple features such as texture, shape and motion along with color information. However, if the skin classifier is alone used to identify faces, then attaining both high TPR and low FPR is very important.

2.4. Comparison of color space representations

Several color spaces have been proposed and used for skin detection. Table 1 lists the various color spaces used for different skin classifiers. As indicated in Table 1, some methods drop the intensity component so as to provide robust parameters against various illumination conditions. Given different color spaces, to choose the most appropriate color space for skin classification is a very difficult task.

To the best of our knowledge, the first comparison of color spaces was made by Littman and Ritter [78]. They compared a neural approach based on linear maps for skin color with normal distributions using three different colors: RGB, YIQ and YUV on a small dataset containing hand images collected indoor. They reported that the performance is predominantly independent of the color space representation and the neural approach gives a better performance. Zarit et al. [33]

compared five different color spaces (CIE Lab, HSV, Fleck HS, *rgb*, YCbCr) and two skin detection methods (look-up table and Bayes classifier) on a very limited data set of 48 training and 64 test images. The performance of the color spaces with Bayes classifier is similar. The look-up table method performed best when used with HSV or Fleck-Hs. They concluded that HS-based color spaces provided better results. Terillion et al. [102] have evaluated nine different chrominance spaces (normalized *rg*, CIE-*xy*, CIE-DSH, TSL, HSV, YES, YIQ, CIE-Lab, CIE-Luv) for skin segmentation with SGMs and GMMs on a data set of 100 images with 144 faces and 65 subjects (30 Asians, 34 Caucasians, 1 African). They reported that normalized color spaces produced better discrimination between skin and non-skin distributions with SGMs. However, for un-normalized color spaces the comparable performance is achieved only with GMMs. They concluded that the TSL space provided better results.

Albiol et al. [79] provided a *theoretical proof* that for every color space there is an optimum skin detector with comparable performance provided that there is an invertible transformation between the color spaces. Skin detection using Bayes classifier on three color spaces (RGB, YCbCr, HSV) were compared on a database of 200 images and their performance was found to be similar. They also showed that the performance of a 3D color space (YCbCr) color space is better than the corresponding 2D color space (CbCr) as the transformation form 3D–2D space is not invertible.

Many of the above comparisons were performed on very limited data sets. However, to rank the color spaces based on their separability between skin and non-skin pixels, the datasets to be considered must include very large number of skin and non-skin samples. The skin pixels must include a large number of subjects from different races with fewer restrictions on subject appearances and must cover a wide range of illumination conditions both indoor and outdoor. Shin et al. [80] evaluated the separability of skin and non-skin clusters in nine different color spaces (CIE-Lab, CIE-XYZ, HIS, normalized RGB, RGB, SCT, YCbCr, YIQ, YUV) using metrics derived from scatter matrices and skin, non-skin histograms. They used a database of 805 images with different skin tones and illumination. Out of these, 507 skin images were collected from AR face database [81] and University of Oulu Physics database (UOPB) [82] and 298 non-skin images were collected from University of Washington CBIR database [83]. They found that the separability between skin and non-skin clusters is the highest in RGB color space (i.e., without color transformation). Also, dropping the luminance component and transforming 3D color space to 2D significantly decreases the separability, thereby reducing the skin segmentation results. Phung et al. [75] evaluated the same on ECU face and skin database with histogram technique and concluded the same results as that of Shin et al. [80]. In Jayaram et al. [76], the same group [80] compared different color spaces but with two different skin classification approaches: histogram and Gaussian approaches. In this work, they found that the choice of color modeling technique makes a significant difference and the choice of color transformation improves the performance but not consistently. Fu et al. [77]compared the performance of four color spaces (RGB, HSV, YCbCr and rg) with GMMs. They found that *rg* performed the worst due to its transformation from 3D to 2D while the HSV color space in which the chrominance and luminance information de-correlated performs the best. Their results also suggest that the choice of skin modeling technique plays an important role in the performance for a particular color space.

The most important criteria for the performance of any given skin classifier is the degree of overlap between the skin and non-skin clusters in a color space. From the above discussion, we can conclude that non-parametric models such as histogram-based Bayes classifier are not affected by the color transformation as the degree of overlap is unaffected by mapping from one color space to another lossless transformation. However, parametric modeling techniques such as Gaussian modeling are affected by choice of color space. It should be noted that the parametric models are also affected by the amount and quality of the training data available. The choice of appropriate color space should also depend on the available image format and the necessity of a particular color space in post-processing steps. For example, some non-linear color space transformations are too computational expensive to be used in real-time. Many of the existing techniques drop the luminance components to reduce the illumination effects in skin detection without any evidence. As the results suggest, ignoring the luminance component degrades the skin detection performance.

3. Illumination adaptation approaches

One of the most important factors for the success of any skin-color model is its ability to be able to adapt to the changes in the lighting and the viewing environment. The skin-color distribution of the same person under different lighting conditions differs. Even under the same lighting conditions, background colors and shadows may influence skin-color appearance. Furthermore, if a person is moving, the apparent skin colors change as the person's position relative to the camera or light changes. Human visual system can dynamically adapt to the varying lighting conditions and can approximately preserve the actual color of the object. This ability of the humans to reduce the effect of light on the color of the object and to retain a stable perceptual representation of the surface color is referred to as color constancy or chromatic adaptation. However, unlike humans, image capturing devices such as digital cameras are not capable of adapting to the rapidly varying illuminations across scenes. Most of the skin-color detection strategies described in Section 3 are immune only to slight variations in lighting and shading conditions. To handle the rapidly changing illumination conditions for skin detection, there are primarily two different classes of approaches: color constancy and dynamic adaptation which are described briefly in the following section.

3.1. Skin-color constancy approaches

Color constancy approaches transform the image contents to a known canonical illuminant that reflects precisely the physical contents of the scene. This consists of estimating the illuminant color and then correcting the image pixel-wise based on the estimate of the illuminant. Estimating the illuminant is critical in solving the image chromatic adaptation problem. A number of strategies have been proposed to estimate the image illuminant. All these algorithms are based on the assumptions of either the existing camera characteristics or the illuminant properties or the distribution of the color values. Gray World algorithms [84] assume that the average reflectance of the image(s) is gray and the illuminant is estimated as the color shift from the average gray value of the image. Retinex algorithms [85] try to model the human color perception system and estimate the illuminant by comparing the average color value at each pixel to the maximum value found by looking at a larger area in the image. The gamut mapping algorithms [86,99] consider all the possible mappings between the set of color values under the known and unknown illuminants. The set of all possible color values under any illuminant is referred to as gamut. A unique solution to this mapping is then obtained. In Bayesian color constancy [87], a ML approach is used to determine the illuminant estimate so to maximize the likelihood of the observed data. In neural network-based method [88], a NN is used to estimate the chromaticity of the illuminant. The input to the network is a binary value indicating the presence of sampled chromaticity. The output of the network is the expected chromaticity. For a comparison of different approaches to color constancy, readers are referred to Barnard et al. [89,90]. In skin-color constancy approaches, the color constancy algorithm is applied as a preprocessing step. Based on the estimate of the illuminant, we first color correct the image using a diagonal transformation. Skin-color modeling and detection is then applied on these color corrected images.

3.1.1. Gray World and white patch approaches

The Gray World algorithm is one of the simplest and widely used algorithms for estimating the illuminant. It assumes that given an image with sufficiently varied colors, the average reflectance of the surfaces in the image is gray. Hence, any shift from gray of the measured averages on the sensor responses correspond to the color of the illuminant. Kovac et al. [91] used the Gray World method to color correct the images before applying skin detection. They concluded on a dataset of 40 images that apply color correction improves the performance of the skin classifier.

The white patch algorithm searches for a white patch in the image under the assumption that the brightest patch in the image is white. The chromaticity of the illuminant is the chromaticity of the white patch. Hsu et al. [35] used a version of white patch algorithm for color correction. The pixels with the top 5% of luma are taken as a reference white patch. The R, G and B components of the image are then adjusted such that the average color value of this reference white patch is grey. These color corrected image pixels are transformed non-linearly into YCbCr space. The skin-pixels are detected using an elliptical model with Mahalanobis distance. Good skin detection results (around 96% detection rate) were obtained on HH1 MPEG7 video and Champion databases, which contain images with both frontal and non-frontal faces under varying backgrounds and illumination conditions. However the number of false positives is high. A subsequent facial feature detection procedure is also proposed which reduces dramatically the false positives.

3.1.2. Neural network approaches

In NN based approaches, a NN is trained to learn the relationships between the colors in the image and the expected illuminant. The advantage of using NNs is that there are no explicit assumptions regarding the image content as in Gray World or white patch methods. Nayak and Chaudari [92] used a NN for color constancy in tracking human palm. In their approach, a NN ($2 * 20 * 3$) is trained using a back propagation to directly learn the illuminant parameters. The inputs to the NN are RGB components of the skin pixels, while the output of the network is the expected canonical RGB components. The NN is trained on a set of images containing human palm under varying illumination conditions. The results reported suggest that NN adapts to the illuminant parameters and NN adapted palm images can be tracked precisely in a variety of cluttered backgrounds and varying illuminations.

Kakumanu et al. [93,103] also used NN for color constancy. The proposed three layered NN ($1600 * 48 * 8 * 2$) directly estimates the illuminants so as to bring the skin color to gray. The input to the NN is an *rg* histogram and the output of the network is the expected illuminant of the skin in *rg* space. The NN is trained on a dataset of 255 images and tested on 71 images, the images representing a wide range of illuminations both indoor and outdoor, different backgrounds and non-white light sources. A simple thresholding technique is used to detect skin from these NN color corrected images.

3.1.3. Skin locus approach

Störring et al. [21] have used a physics-based model to solve the problem of color constancy for skin detection. The physics-based model describes an expected area for the skin chromaticities under certain illuminations. The knowledge of skin pixels defines an area in chromaticity space which is called *skin locus*. A thresholding can be applied on this skin locus to classify skin from non-skin pixels for a range of illumination and white balancing conditions. However, the skin locus region is strictly dependent on the camera sensor responses. Therefore, Soriano et al. [20] made the locus directly from images taken under four representative light sources, Horizon (2300 K), Incandescent A (2856 K), fluorescent TL84, and daylight 6500 K. The obtained locus was shown to be useful for selecting pixels in probability based face tracking.

3.2. Dynamic adaptation approaches

Dynamic color modeling approaches adapt by transforming the previously developed skin-color models to the changing environment. Histograms and GMMs are the popular choices for modeling skin-color distribution in dynamic approaches. The advantage of the histogram approach is that the probability density function can be computed trivially. The advantage of the GMMs is that the parameters can be updated in real-time, given that the number of Gaussian components is known apriori.

3.2.1. Explicit threshold adaptation

Cho et al. [94] proposed an adaptive thresholding technique in HSV color space. They defined a thresholding box in HSV color space to separate skin from non-skin pixels. Initially, the dimensions of this box are obtained by observing the skin-color distributions of several color sample images. When a new image is tested, in order to get robustness against illumination changes, the threshold values of the S and V components are updated based on a color histogram built in SV space. The thresholding box is updated by considering the center of gravity of the color histogram calculated from the color values over 10% of the maximum color value in the box. However, this box is not enough to separate skin-color regions from background regions of slightly different colors. Cluster analysis is then performed to determine if dominant background color vectors exist in the box and if present, they are separated from the skin-color vectors by a linear classifier. The results were reported on a data set of 379 web images containing yellow people. However, the method has a disadvantage as the threshold values are bound to change if the people from different races and broader range of varying illuminations are considered.

3.2.2. GMM adaptation

Yang and Waibel [11,59] were one of the first to propose an adaptive approach for skin-based face tracking. They used a GMM for skin modeling in *rg* space. The mean and the covariance of the model as the illumination changes dynamically are approximated using a linear combination of the previous parameters based on ML approach. This adaptive model is applied for a real-time face tracker and works fine for slightly varying indoor illuminations. A similar approach was used by Zue and Yang [100] for tracking hands. In this work, a restricted EM algorithm is used to train the adaptive GMM, where the background is modeled by four Gaussian components and the hand color is modeled by one Gaussian component. Oliver et al. [16] used an adaptive GMM in *rgb* space to cope up with the changing viewing and illumination conditions. An initial mixture model is obtained off-line with EM algorithm. To update the GMM parameters online, they used an incremental EM technique. McKenna et al. [27] also used an adaptive GMM in HS space. The number of components in this model was fixed while the parameters were updated online using stochastic equations. Due to the lack of ground truth, the adaptive model might adapt to non-skin image regions (for example during occlusions). To reduce this problem, such outliers were detected using log-likelihood measurements and the corresponding frame color data is not used for subsequent frames adaptation. Bergasa et al. [13] used an unsupervised and adaptive Gaussian skin-color model (UAGM) in *rg* space for skin segmentation. The GMM is initialized with clusters computed by a competitive vector quantization (VQ) learning algorithm. The model parameters are updated in time using a linear combination of the previous parameters. Schwerdt and Crowley [18] used the first-order moments of the skin-color histogram in *rg* space to represent the position and the spatial extent of the skin-colored region. To reduce the effect of non-skin pixels during tracking, the skin histogram is weighted by a Gaussian function. The mean and covariance of this Gaussian function for a new image frame are updated form the values of the previous frame.

3.2.3. Histogram adaptation

Soriano et al. [20,21] developed an adaptive skin-color modeling technique using skin locus. Skin locus (Section 3.1.3) is the range of skin colors in the chromaticity space. Initially in an image frame, skin pixels are extracted from the tracked bounding box and within the skin locus defined in *rg* chromaticity space. The ratio histogram is updated using the histogram of these skin pixels and the histogram of the whole image. The updated histogram is back projected to define the search space (the bounding box) for the next frame and the process is repeated. The rationale behind using the skin locus approach is that, the skin chromaticities in an image occupies only a small portion the skin locus and it is easy to track this small region than the entire skin locus.

3.2.4. HMM adaptation

Sigal et al. [95] used a second-order Markov model to predict the evolution of skin color (HSV) histogram over time. To get an initial estimate of the skin-color distribution to be tracked, a Bayes classifier constructed on Compaq Dataset is considered. Once the initialization is done, the Learning stage performs the EM step over the first few video frames. For each frame, the EM algorithm's E step is histogram-based segmentation, and the M step is histogram adaptation. The evolution of the skin-color distribution at each frame is

Table 2
Performance of different skin detection strategies with illumination adaptation approaches

Authors	Color space	Intensity comp.	Skin detection method	Illum. adapt.	Pre-train	Test database	Diff. skin types	Diff. Illum	Skin types	Illum. types	True positives	
Skin-color constancy methods												
Hsu et al. 02	YCbCr	No	SGM	WP	Yes	HH1	Yes	Yes	n/a	n/a	96.6	n/a
						Champion	Yes	Yes	n/a	n/a	99.1	n/a
Kovac and Peer, 03	YUV	Yes	Elliptic model	GW	Yes	60 images	n/a	Yes	n/a	White, non-white	n/a	n/a
Nayak et al. 03	RGB	Yes	n/a	NN	Yes	n/a	No	Yes	n/a	Cluttered-background low-lights	n/a	n/a
Kakumanu et al. 04	RGB	Yes	Thresh.	NN	Yes	326 images	No	Yes	n/a	Indoor, outdoor, incandescent fluorescent shadows	n/a	n/a
Störring et al. 03	Rgb	No	Thresh.	Skin locus	No	UOPB	Yes	Yes	White yellow dark	Horizon incandescent fluorescent daylight	n/a	n/a
Dynamic adaptation methods												
Cho and Jang, 01	HSV	No	Thresh.	Adapt. tresh.	Yes	379 web images	Yes	Yes	White red bright dark	n/a	n/a	n/a
Yang and Waibel, 98	Rgb	No	SGM	SGM adapt.	Yes	n/a	Yes	Yes	White yellow brown dark	n/a	n/a	n/a
Oliver et al. 97	Rgb	No	GMM	GMM adapt.	Yes	n/a	Yes	Yes	n/a	n/a	n/a	n/a
Bergasa et al. 00	Rgb	No	GMM	GMM adapt.	Yes	n/a	Yes	Yes	White yellow dark	n/a	n/a	n/a
Soriano et al. 03	Rgb	No	Skin locus thresh.	Hist. adapt.	No	UOPB	Yes	Yes	White yellow dark	Horizon incandescent fluorescent daylight	n/a	n/a
Sigal and Sclaroff, 04	HSV	Yes	Bayes	HMM adapt.	Yes	Movies	Yes	Yes	n/a	Cluttered-background indoor, outdoor, incandescent shadows	86.8	n/a

parameterized by translation, scaling, and rotation in color space. During prediction and tracking stage, histograms are dynamically updated based on feedback from the current segmentation and predictions of the Markov model and this updated histogram is used for skin segmentation in the next frame. The method is shown to be robust for a wide range of illumination conditions including non-white and multiple light sources.

3.3. Comparison of illumination adaptation approaches

Table 2 lists the summary of the illumination adaptation approaches to skin-color modeling and detection methods as discussed in Sections 3.1 and 3.2.2. The explicit skin-color constancy or the dynamic adaptation approaches to skin detection are proved to provide better results than the static skin detection approaches discussed in Section 2.2, especially in a wide range of illumination conditions and cluttered backgrounds.

Kovac et al. [91] reported on a dataset of 40 images that applying color correction improves the performance of skin classifier. Kakumanu et al. [93] trained a NN for skin-color adaptation on a dataset of 255 images and tested on 71 images, the images representing a wide range of illuminations both indoor and outdoor, different backgrounds and non-white light sources. They reported that applying color constancy technique improved the skin-color stabilization in images. Also, the proposed NN trained on skin patch performs favorably against Gray World, White Patch and NN trained on white patch methods. The advantage of NN approach to color constancy is that there are no explicit assumptions regarding the image content as in Gray World or White Patch algorithms. MartinKauppi et al. [96] compared skin locus approach with three different algorithms: adaptive skin filter approach of Cho et al. (Section 3.2.1), statistical approach of Jones and Rehg (Section 2.2.3) and color correction approach of Hsu et al. (Section 3.1.2) on a video taken from the Face Video Database [82]. The video contains a person walking in a corridor. The illumination over the face varies from the light of fluorescent lamp to daylight from the windows and to mixtures of these both. For the video considered, the performance of the skin-locus approach is better. However, the skin locus approach can be applied only if the camera parameters are known. Sigal et al. [95] tested the HMM Illumination adaptation method on 21 video sequences from nine popular movies. These videos contain illumination conditions ranging from white to non-white light sources, cluttered backgrounds and shadows. The average classification for skin is 86.84% and for the background is 93.55%, a 24% performance improvement when compared with the static histogram technique of Jones and Rehg [54].

The rapidly varying illumination conditions, shadows and cluttered background pose a major problem to the performance of the existing skin-color classifiers. As indicated by the above reported comparisons, it is clear that using an illumination adaptation approach, either skin-color constancy or the dynamic adaptation approach, improves the performance of the skin-color classifier in these situations. Hence, it is suggested to use one of these methods to improve the skin classification performance. An in-depth comparison of these methods is not possible, as the above methods are not evaluated on large and common datasets. The particular method to use is dependent on the application constraints such as real-time issues and the nature of the dataset at hand.

4. Summary and conclusions

In this paper, we have presented an extensive survey of the up-to-date techniques for skin detection using color information in the visual spectrum for 2D images. A good skin classifier must be able to discriminate between skin and non-skin pixels for a wide range of people with different skin types such as white, yellow, brown and dark and be able to perform well under different illumination conditions such as indoor, outdoor and with white and non-white illumination sources. We reviewed the following critical issues regarding skin detection:

- *Choice of appropriate color space*: The color space representation is often lead by the skin detection methodology and the application. Non-parametric methods such as histogram-based methods are not affected by the color space representation. However, parametric modeling techniques such as Gaussian modeling are affected by the choice of color space. It should be noted that the parametric models are also affected by the amount and the quality of the training data available. The choice of color space should also depend on the available image format and the necessity of a particular color space in post-processing steps. For example, some non-linear color space transformations are too computational expensive to be used in real-time. Dropping the intensity component so as to obtain robust parameters against illumination conditions as approached in many of the works actually reduces the skin detection performance.
- *Skin-color modeling and classification techniques*: The histogram-based Bayes classifier is feasible for skin detection only with large datasets. Also, this method has very high storage requirements when compared to mixture or NN models. The performance of mixture and NN models is comparable to that of histogram methods even when small datasets are available.
- *Color constancy and dynamic adaptation approaches*: Obtaining robust color representations against varied illumination conditions is still a major problem. However, the application of color constancy techniques as a preprocessing step to skin-color modeling proved to improve the performance. The NN-based color constancy techniques are very promising as they do not make any explicit assumptions about the scene content. Dynamic

adaptation techniques which transform the existing color models so as to adapt to the changing viewing conditions are also available. However, the problem with these approaches is that if we lose track of ground truth, the adaptive model might adapt to non-skin image regions. The problem domain factors such as the characteristics of the data set, the real-time considerations and the skin detection method should lead the choice of one of the illumination adaptation methods.

Most of the skin detection techniques discussed in literature are used a preprocessor for face detection and tracking systems. Though skin color analysis often produces high TPRs, it also produces high FPRs when the image contains cluttered background and shadows. To improve the accuracy of the classifier, various other features such as shape, spatial and motion information can be used along with skin-color information. In the past decade significant advancement has been made in skin detection using color information in the visual spectrum. However, to build an accurate classifier which can detect all the skin types under different illuminations, shadows, cluttered backgrounds and makeup is still an unsolved problem. For example, it is extremely hard to build a skin classifier which can distinguish human skin color from dyes designed to mimic skin color. Many of the problems encountered in visual spectrum can be overcome by using non-visual spectrum methods such as infrared (IR) and spectral imaging. The non-visual spectrum methods though immune to illumination conditions and makeup are very expensive and bulky, and hence are not suitable to many applications such as Tyflos [97] and iLearn [98]. Often the applications demand certain constraints on the skin detection methodology. When these methods are used in real-time, meeting computational and storage requirements is extremely important. Sometimes, accuracy may need to be sacrificed when the skin detection strategy is used only as a preprocessing step to face detection.

Acknowledgments

This research was partially supported by awards from NSF-ASU, AIIS Inc. and WSU/DAGSI.

References

[1] R. Chellappa, C. Wilson, S. Sirohey, Human and machine recognition of faces: a survey, Proc. IEEE 83 (5) (1995) 705–740.
[2] E. Hjelmas, B.K. Low, Face detection: a survey, J. Comput. Vision Image Understanding 83 (2001) 236–274.
[3] M.H. Yang, D.J. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 24 (1) (2002) 34–58.
[4] W. Zhao, R. Chellappa, P.J. Philips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surveys 85 (4) (2003) 299–458.
[5] D.A. Socolinsky, A. Selinger, J.D. Neuheisel, Face recognition with visible and thermal infrared imagery, Comput. Vision Image Understanding 91 (1–2) (2003) 72–114.
[6] S.G. Kong, J. Heo, B.R. Abidi, J. Paik, M.A. Abidi, Recent advances in visual and infrared face recognition—a review, Comput. Vision Image Understanding 97 (2005).
[7] C. Balas, An imaging colorimeter for noncontact tissue color mapping, IEEE Trans. Biomed. Eng. 44 (6) (1997) 468–474.
[8] E. Angelopoulou, R. Molana, K. Daniilidis, Multispectral skin color modeling, CVPR01, 2001.
[9] Z. Pan, G. Healey, M. Prasad, B. Tromberg, Face recognition in hyperspectral images, IEEE Trans. Pattern Anal. Mach. Intell. 25 (12) (2003).
[10] V. Vezhnevets, V. Sazonov, A. Andreeva, A survey on pixel-based skin color detection techniques, GRAPHICON03, 2003, pp. 85–92.
[11] J. Yang, W. Lu, A. Waibel, Skin-color modeling and adaptation, ACCV98, 1998.
[12] M.H. Yang, N. Ahuja, Gaussian Mixture model for human skin color and its application in image and video databases, Proceedings of SPIE: Conference on Storage and Retrieval for Image and Video Databases, vol. 3656, 1999, pp. 458–466.
[13] L.M. Bergasa, M. Mazo, A. Gardel, M.A. Sotelo, L. Boquete, Unsupervised and adaptive Gaussian skin-color model, Image Vision Comput. 18 (12) (2000) 987–1003.
[14] D. Brown, I. Craw, J. Lewthwaite, A SOM based approach to skin detection with application in real time systems, BMVC01, 2001.
[15] T.S. Caetano, D.A.C. Barone, A probabilistic model for the human skin-color, ICIAP01, 2001, pp. 279–283.
[16] N. Oliver, A. Pentland, F. Berard, Lafter: lips and face real time tracker, CVPR97, 1997.
[17] S.H. Kim, N.K. Kim, S.C. Ahn, H.G. Kim, Object oriented face detection using range and color information, AFGR98, 1998.
[18] K. Schwerdt, J.L. Crowely, Robust face tracking using color, AFGR00, 2000.
[19] N. Sebe, T. Cohen, T.S. Huang, T. Gevers, Skin detection, a Bayesian network approach, ICPR04, 2004.
[20] M. Soriano, J.B. MartinKauppi, S. Huovinen, M. Lääksonen, Adaptive skin color modeling using the skin locus for selecting training pixels, Pattern Recognition 36 (3) (2003) 681–690.
[21] M. Störring, T. Koèka, H.J. Anderson, E. Granum, Tracking regions of human skin through illumination changes, Pattern Recognition Lett. 24 (11) (2003).
[22] J.G. Wang, E. Sung, Frontal-view face detection and facial feature extraction using color and morphological operations, Pattern Recognition Lett. 20 (1999) 1053–1068.
[23] Poynton FAQ, Charles Poynton, ⟨http://www.poynton.com/PDFs/ColorFAQ.pdf⟩.
[24] C. Chen, S.P. Chiang, Detection of human faces in colour images, IEE Proc. Vision Image Signal Process. 144 (6) (1997) 384–388.
[25] H. Wu, Q. Chen, M. Yachida, Face detection from color images using a fuzzy pattern matching method, IEEE Trans. Pattern Anal. Mach. Intell. 21 (6) (1999) 557–563.
[26] C. Garcia, G. Tziritas, Face detection using quantized skin color regions merging and wavelet packet analysis, IEEE Trans. Multimedia 1 (3) (1999) 264–277.
[27] S. McKenna, S. Gong, Y. Raja, Modeling facial colour and identity with Gaussian mixtures, Pattern Recognition 31 (12) (1998) 1883–1892.
[28] D. Saxe, R. Foulds, Toward robust skin identification in video images, AFGR96, 1996.
[29] K. Sobottka, I. Pitas, A novel method for automatic face segmentation, facial feature extraction and tracking, Signal Process. Image Commun. 12 (1998) 263–281.
[30] Q.H. Thu, M. Meguro, M. Kaneko, Skin-color extraction in images with complex background and varying illumination, Sixth IEEE Workshop on Applications of Computer Vision, 2002.
[31] Y. Wang, B. Yuan, A novel approach for human face detection from color images under complex background, Pattern Recognition 34 (10) (2001) 1983–1992.
[32] Q. Zhu, K.-T. Cheng, C.-T. Wu, Y.-L. Wu, Adaptive learning of an accurate skin-color model, AFGR04, 2004.

[33] B.D. Zarit, J.B. Super, F.K.H. Quek, Comparison of five color models in skin pixel classification, ICCV99, 1999.
[34] J.C. Terillon, M. David, S. Akamatsu, Detection of human faces in complex scene images by use of a skin color model and of invariant Fourier–Mellin moments, ICPR98, 1998, pp. 350–1355.
[35] R.L. Hsu, M. Abdel-Mottaleb, A.K. Jain, Face detection in color images, IEEE Trans. Pattern Anal. Machine Intell. 24 (5) (2002) 696–706.
[36] D. Chai, A. Bouzerdoum, A Bayesian approach to skin color classification in YCbCr color space, IEEE TENCON00, vol. 2, 2000, pp. 421–424.
[37] D. Chai, K.N. Ngan, Locating facial region of a head-and-shoulders color image, ICFGR98, 1998.
[38] K.W. Wong, K.M. Lam, W.C. Siu, A robust scheme for live detection of human faces in color images, Signal Process. Image Commun. 18 (2) (2003) 103–114.
[39] J.J. de Dios, N. Garcia, Face detection based on a new color space YCgCr, ICIP03, 2003.
[40] Y. Dai, Y. Nakano, Face-texture model based on SGLD and its application in face detection in a color scene, Pattern Recognition 29 (6) (1996) 1007–1017.
[41] F. Marqués, V. Vilaplana, A morphological approach for segmentation and tracking of human face, ICPR 2000, 2000.
[42] E. Saber, A.M. Tekalp, Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions, Pattern Recognition Lett. 17 (8) (1998).
[43] G. Gomez, M. Sanchez, L.E. Sucar, On selecting an appropriate colour space for skin detection, Springer-Verlag: Lecture Notes in Artificial Intelligence, vol. 2313, 2002, pp. 70–79.
[44] J. Cai, A. Goshtasby, Detecting human faces in color images, Image Vision Comput. 18 (1999) 63–75.
[45] S. Kawato, J. Ohya, Automatic skin-color distribution extraction for face detection and tracking, Fifth International Conference on Signal Processing, vol. 2, 2000, pp. 1415–1418.
[46] G. Wyszecki, W.S. Stiles, Color Science, Wiley, New York, 1967.
[47] J. Brand, J. Mason, A comparative assessment of three approaches to pixel level human skin-detection, ICPR01 1 (2000) 1056–1059.
[48] M.J. Jones, J.M. Rehg, Statistical color models with application to skin detection, CVPR99, 1999.
[49] K. Sobottka, I. Pitas, Extraction of facial regions and features using color and shape information, ICPR96, 1996.
[50] D. Chai, K.N. Ngan, Face segmentation using skin-color map in videophone applications, IEEE Trans. Circuits Syst. Video Technol. 9 (4) (1999).
[51] F. Tomaz, T. Candeias, H. Shahbazkia, Fast and accurate skin segmentation in color images, CRV04, 2004.
[52] G. Gomez, E. Morales, Automatic feature construction and a simple rule induction algorithm for skin detection, Proceedings of Workshop on Machine Learning in Computer Vision, 2002, pp. 31–38.
[53] T.W. Yoo, I.S. Oh, A fast algorithm for tracking human faces based on chromatic histograms, Pattern Recognition Lett. 20 (10) (1999) 967–978.
[54] M.J. Jones, J.M. Rehg, Statistical color models with application to skin detection, J. Comput. Vision 46 (1) (2002) 81–96.
[55] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley InterScience, New York, 2002.
[56] M.M. Fleck, D.A. Forsyth, C. Bregler, Finding naked people, Proceedings of European Conference on Computer Vision, vol. 2, 1996, pp. 592–602.
[57] S. Marcel, and S. Bengio, Improving face verification using skin color information, ICPR02, 2002.
[58] S. Srisuk and W. Kurutach, "New robust face detection in color images, AFGR02, 2002, pp. 291–296.
[59] J. Yang, A. Waibel, A real-time face tracker, Proceedings of the third IEEE Workshop on Applications of Computer Vision, WACV96, 1996.
[60] M.H. Yang, N. Ahuja, Detecting human faces in color images, ICIP98, 1998.
[61] P. Kuchi, P. Gabbur, S. Bhat, S. David, Human face detection and tracking using skin color modeling and connected component operators, IETE J. Res., Special Issue on Visual Media Processing, May 2002.
[62] T.S. Caetano, S.D. Olabarriaga, D.A.C. Barone, Performance evaluation of single and multiple-Gaussian models for skin-color modeling, SIBGRAPI02, 2002.
[63] J.Y. Lee, S.I. Yoo, An elliptical boundary model for skin color detection, Proceedings of the International Conference on Imaging Science, Systems and Technology, 2002.
[64] J.C. Terrillon, M.N. Shirazi, H. Fukamachi, S. Akamatsu, Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images, CFGR00, 2000, pp. 54–61.
[65] H. Greenspan, J. Goldberger, I. Eshet, Mixture model for face-color modeling and segmentation, Pattern Recognition Lett. 22 (14) (2001) 1525–1536.
[66] T.S. Caetano, S.D. Olabarriaga, D.A.C. Barone, Do mixture models in chromaticity space improve skin detection?, Pattern Recognition 36 (12) (2003) 3019–3021.
[67] J. Karlekar, U.B. Desai, Finding faces in color images using wavelet transform, International Conference on Image Analysis and Processing, 1999.
[68] S.L. Phung, D. Chai, A. Bouzerdoum, A universal and robust human skin color model using neural networks, IJCNN01, 2001.
[69] H. Sahbi, N. Boujemaa, Coarse to fine face detection based on skin color adaptation, Workshop on Biometric Authentication, 2002, Lecture Notes in Computer Science, vol. 2359, 2002, pp. 112–120.
[70] M.J. Seow, D. Valaparla, V.K. Asari, Neural network-based skin color model for face detection, Proceedings of the 32nd Workshop on Applied Imagery Pattern Recognition, 2003.
[71] I. Anagnostopoulos, C. Anagnostopoulos, V. Loumos, E. Kayafas, A probabilistic neural network for human face identification based on fuzzy logic chromatic rules, IEEE MED03, 2003.
[72] B. Jedynak, H. Zheng, M. Daoudi, Statistical models for skin detection, IEEE Workshop on Statistical Analysis in Computer Vision, 2003.
[73] H. Zheng, M. Daoudi, B. Jedynak, From maximum entropy to belief propagation: an application to skin detection, Proceedings of British Machine Vision Conference, BMVC04, 2004.
[74] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, Mach. Learn. 29 (2) (1997) 131–163.
[75] S.L. Phung, A. Bouzerdoum, D. Chai, Skin segmentation using color pixel classification: analysis and comparison, IEEE Trans. Pattern Anal. Mach. Intell. 27 (1) (2005).
[76] S. Jayaram, S. Schmugge, M.C. Shin, L.V. Tsap, Effect of color space transformation, the illuminance component, and color modeling on skin detection, CVPR04, 2004, pp. 813–818.
[77] Z. Fu, J. Yang, W. Hu, T. Tan, Mixture clustering using multidimensional histograms for skin detection, ICPR04, 2004, pp. 549–552.
[78] E. Littmann, H. Ritter, Adaptive color segmentation: A comparison of neural and statistical methods, IEEE Trans. Neural Networks 8 (1) (1997) 175–185.
[79] A. Albiol, L. Torres, E.J. Delp, Optimum color spaces for skin detection, ICIP01, 2001.
[80] M.C. Shin, K.I. Chang, L.V. Tsap, Does colorspace transformation make any difference on skin detection? IEEE Workshop on Applications of Computer Vision, Orlando, FL, December 2002, pp. 275–279.
[81] A.M. Martinez, R. Benavente, The AR face database, CVC Technical Report #24, June 1998.
[82] E. Marszalec, B. Martinkauppi, M. Soriano, M. Pietikinen, A physics-based face database for color research, J. Electron. Imaging 9 (1) (2000) 32–38.
[83] A. Berman, L.G. Shapiro, A flexible image database system for content-based retrieval, Comput. Vision Image Understanding 75 (1–2) (1999) 175–195.

[84] G. Buchsbaum, A spatial processor model for object colour perception, J. Franklin Inst. 310 (1980) 1–26.
[85] D.H. Brainard, B.A. Wandell, Analysis of the retinex theory of color vision, J. Opt. Soc. America 3 (1986) 1651–1661.
[86] G.D. Finlayson, S. Hordley, Improving gamut mapping colour constancy, IEEE Trans. Image Process. 9 (10) (2000) 1774–1783.
[87] D.H. Brainard, W.T. Freeman, Bayesian color constancy, J. Opt. Soc. Am. A 14 (1997) 1393–1411.
[88] V. Cardei, A neural network approach to colour constancy, Ph.D. Thesis, Simon Fraser University, January 2000.
[89] K. Barnard, B. Funt, V. Cardei, A comparison of computational color constancy algorithms—Part I: theory and experiments with synthetic data, IEEE Trans. Image Process. 11 (9) (2002) 972–984.
[90] K. Barnard, L. Martin, A. Coath, B. Funt, A comparison of computational color constancy algorithms—Part II: Experiments with image data, IEEE Trans. Image Process. 11 (9) (2002) 985–996.
[91] J. Kovac, P. Peer, F. Solina, Human skin color clustering for face detection, EUROCON2003, 2003.
[92] A. Nayak, S. Chaudhuri, Self-induced color correction for skin tracking under varying illumination, ICIP03, 2003.
[93] P. Kakumanu, S. Makrogiannis, R. Bryll, S. Panchanathan, and N. Bourbakis, Image chromatic adaptation using ANNs for skin color adaptation, Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence, ICTAI04.
[94] K.M. Cho, J.H. Jang, K.S. Hong, Adaptive skin-color filter, Pattern Recognition 34 (5) (2001) 1067–1073.
[95] L. Sigal, S. Sclaroff, V. Athitsos, Skin color-based video segmentation under time-varying illumination, IEEE Trans. Pattern Anal. Mach. Intell. 26 (6) (2004).
[96] B. Martinkauppi, M. Soriano, M. Pietikäinen, Detection of skin-color under changing illumination: a comparative study, 12th Interenational Conference on Image Analysis and Processing, 2003.
[97] N. Bourbakis, D. Kavraki, An intelligent agent for blind people navigation, Proc. IEEE International Symposium on BIBE-01, 2001, pp. 230–235.
[98] S. Panchanathan et al., "iCare—a user centric approach to the development of assistive devices for blind and visually impaired, ICTAI03, 2003, pp. 641–648.
[99] D. Forsyth, A novel approach to color constancy, J. Comput. Vision 5 (1) (1990) 5–36.
[100] X. Zhu, J. Yang, A. Waibel, Segmenting hands of arbitrary color, AFGR00, 2000.
[101] H. Yao, W. Gao, Face detection and location based on skin chrominance and lip chrominance transformation from color images, Pattern Recognition 34 (8) (2001) 1555–1564.
[102] J.-C. Terillon, M.N. Shirazi, H. Fukamachi, S. Akamatsu, Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images, AFGR00, 2000, pp. 54–61.
[103] P. Kakumanu, A face detection and facial expression recognition method applicable to assistive technologies and biometrics, PhD Dissertaion, CSE Department, Wright State University, 2006.

About the Author—Kakumanu received his MS in 2003 and his PhD in 2006 from Wright State University. His research interests are in human computer interaction, assistive technology. He has published several journals and Conference papers.

About the Author—Makrogiannis received his PhD in 2000 from University of Patras, Greece. His research interest are in image segmentation, computer vision. He has published several Journals and Conference papers.

About the Author—Bourbakis received his PhD from University of Patras, Greece, 1983. His research interest are in Assistive Technology, Information Security, Intelligent Systems and Robotics, Computer Vision and Wearable Systems. He has published extensively in Internationally refereed Journals and Conference Proceedings.