A Survey on Pixel-Based Skin Color Detection Techniques

Vladimir Vezhnevets * Vassili Sazonov Alla Andreeva

Graphics and Media Laboratory †
Faculty of Computational Mathematics and Cybernetics
Moscow State University,
Moscow, Russia.

Abstract

Skin color has proven to be a useful and robust cue for face detection, localization and tracking. Image content filtering, content-aware video compression and image color balancing applications can also benefit from automatic detection of skin in images. Numerous techniques for skin color modelling and recognition have been proposed during several past years. A few papers comparing different approaches have been published [Zarit et al. 1999], [Terrillon et al. 2000], [Brand and Mason 2000]. However, a comprehensive survey on the topic is still missing. We try to fill this vacuum by reviewing most widely used methods and techniques and collecting their numerical evaluation results.

Keywords: image processing, color segmentation, color space model selection, skin detection

1 Introduction

Face detection and tracking has been the topics of an extensive research for the several past decades. Many heuristic and pattern-recognition based strategies have been proposed for achieving robust and accurate solution. Among feature-based face detection methods, the ones using skin color as a detection cue, have gained strong popularity. Color allows fast processing and is highly robust to geometric variations of the face pattern. Also, the experience suggests that human skin has a characteristic color, which is easily recognized by humans. So trying to employ skin color modelling for face detection was an idea suggested both by task properties and common sense.

When building a system, that uses skin color as a feature for face detection, the researcher usually faces three main problems. First, what colorspace to choose, second, how exactly the skin color distribution should be modelled, and finally, what will be the way of processing of color segmentation results for face detection. This paper covers the first two questions, leaving the third (an equally important one) for another discussion.

In this paper we discuss *pixel-based* skin detection methods, that classify each pixel as skin or non-skin individually, independently from its neighbors. In contrast, *region-based* methods [Kruppa et al. 2002], [Yang and Ahuja 1998], [Jedynak et al. 2002] try to take the spatial arrangement of skin pixels into account during the detection stage to enhance the methods performance.

Pixel-based skin detection has long history, but surprisingly few papers that provide surveys or comparisons of different techniques were published. [Zarit et al. 1999] have provided a comparison of five colorspaces (actually their chrominance planes) and two non-parametric skin modelling methods (lookup table and Bayes skin probability map). [Terrillon et al. 2000] have compared nine chrominance spaces and two parametric techniques (Gaussian and mixture of Gaussians models). [Brand and Mason 2000] have evaluated three different skin color modelling strategies. [Lee and Yoo 2002] also have compared two most popular parametric skin models in different chrominance spaces and have proposed a model of their own.

Our goal, in this paper, is to gather as much published techniques as we could find, describe their key ideas and try to find out and summarize their advantages, disadvantages and characteristic features. The paper is organized as follows. Section 2 is devoted to description of different colorspaces used for skin detection. Section 3 covers the existing skin color modelling methods. In section 4 numerical evaluation of some of the described methods is provided. In Sections 5 and 6 we discuss and compare the colorspaces and modelling methods. In Section 7 the conclusion are drawn.

2 Colorspaces used for skin modelling

Colorimetry, computer graphics and video signal transmission standards have given birth to many colorspaces with different properties. A wide variety of them have been applied to the problem of skin color modelling. We will briefly review the most popular colorspaces and their properties.

2.1 RGB

RGB is a colorspace originated from CRT (or similar) display applications, when it was convenient to describe color as a combination of three colored rays (red, green and blue). It is one of the most widely used colorspaces for processing and storing of digital image data. However, high correlation between channels, significant perceptual non-uniformity (see section 2.6 for perceptual uniformity explanation), mixing of chrominance and luminance data make RGB not a very favorable choice for color analysis and color-based recognition algorithms. This colorspace was used in [Brand and Mason 2000], [Jones and Rehg 1999].

2.2 Normalized RGB

Normalized RGB is a representation, that is easily obtained from the RGB values by a simple normalization procedure:

$$r = \frac{R}{R+G+B} \quad g = \frac{G}{R+G+B} \quad b = \frac{B}{R+G+B} \tag{1}$$

As the sum of the three normalized components is known ($r+g+b=1$), the third component does not hold any significant information and can be omitted, reducing the space dimensionality. The remaining components are often called "pure colors", for the dependance of r and g on the brightness of the source RGB color is diminished by the normalization. A remarkable property of this representation is that for matte surfaces, while ignoring ambient light, normalized RGB is invariant (under certain assumptions) to changes of surface orientation relatively to the light source [Skarbek

*e-mail: vvp@graphics.cmc.msu.ru

†www: http://graphics.cmc.msu.ru

and Koschan 1994]. This, together with the transformation simplicity helped this colorspace to gain popularity among the researchers [Brown et al. 2001], [Zarit et al. 1999], [Soriano et al. 2000], [Oliver et al. 1997], [Yang et al. 1998]

2.3 HSI, HSV, HSL - Hue Saturation Intensity (Value, Lightness)

Hue-saturation based colorspaces were introduced when there was a need for the user to specify color properties numerically. They describe color with intuitive values, based on the artist's idea of tint, saturation and tone. *Hue* defines the dominant color (such as red, green, purple and yellow) of an area, *saturation* measures the colorfulness of an area in proportion to its brightness [Poynton 1995]. The "intensity", "lightness" or "value" is related to the color luminance. The intuitiveness of the colorspace components and explicit discrimination between luminance and chrominance properties made these colorspaces popular in the works on skin color segmentation [Zarit et al. 1999], [McKenna et al. 1998], [Sigal et al. 2000], [Birchfield 1998], [Jordao et al. 1999]. Several interesting properties of Hue were noted in [Skarbek and Koschan 1994]: it is invariant to highlights at white light sources, and also, for matte surfaces, to ambient light and surface orientation relative to the light source. However, [Poynton 1995], points out several undesirable features of these colorspaces, including hue discontinuities and the computation of "brightness" (lightness, value), which conflicts badly with the properties of color vision.

$$H = \arccos \frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-B)(G-B))}} \quad (2)$$

$$S = 1-3\frac{\min(R,G,B)}{R+G+B} \quad (3)$$

$$V = \frac{1}{3}(R+G+B) \quad (4)$$

An alternative way of hue and saturation computation using log opponent values was introduced in [Fleck et al. 1996], where additional logarithmic transformation of RGB values aimed to reduce the dependance of chrominance on the illumination level.

The polar coordinate system of Hue-Saturation spaces, resulting in cyclic nature of the colorspace makes it inconvenient for parametric skin color models that need tight cluster of skin colors for best performance. A different representation of Hue-Saturation using Cartesian coordinates can be used [Brown et al. 2001]:

$$X = S\cos H, \;\; Y = S\sin H \quad (5)$$

2.4 TSL - Tint, Saturation, Lightness

A normalized chrominance-luminance TSL space is a transformation of the normalized RGB into more intuitive values, close to hue and saturation in their meaning.

$$S = [9/5(r'^2+g'^2)]^{1/2} \quad (6)$$

$$T = \begin{cases} arctan(r'/g')/2\pi+1/4, g'>0 \\ arctan(r'/g')/2\pi+3/4, g'<0 \\ 0, g'=0 \end{cases} \quad (7)$$

$$L = 0.299R+0.587G+0.114B \quad (8)$$

where $r' = r-1/3$, $g' = g-1/3$ and r, g come from (1). [Terrillon et al. 2000] have compared nine different colorspaces for skin modelling with a unimodal Gaussian joint pdf (only chrominance components of the colorspaces were used). They argue that normalized TSL space is superior to other colorspaces for this task. [Brown et al. 2001] has also employed this representation for their approach.

2.5 YCrCb

YC_rC_b is an encoded nonlinear RGB signal, commonly used by European television studios and for image compression work. Color is represented by *luma* (which is luminance, computed from nonlinear RGB [Poynton 1995]), constructed as a weighted sum of the RGB values, and two color difference values C_r and C_b that are formed by subtracting luma from RGB red and blue components.

$$\begin{aligned} Y &= 0.299R+0.587G+0.114B \\ C_r &= R-Y \\ C_b &= B-Y \end{aligned} \quad (9)$$

The transformation simplicity and explicit separation of luminance and chrominance components makes this colorspace attractive for skin color modelling [Phung et al. 2002], [Zarit et al. 1999] [Menser and Wien 2000], [Hsu et al. 2002], [Ahlberg 1999], [Chai and Bouzerdoum 2000].

2.6 Perceptually uniform color systems

The term "skin color" is not a physical property of an object, rather a perceptual phenomenon and therefore a subjective human concept. Therefore, color representation similar to the color sensitivity of human vision system should help to obtain high performance skin detection algorithm.

CIELAB and CIELUV are perceptually uniform colorspaces (reasonably perceptually unform, to be exact) that were proposed by G. Wyszecki and standardized by CIE (Commission Internationale de L'Eclairage). Perceptual uniformity means that a small perturbation to a component value is approximately equally perceptible across the range of that value ([Poynton 1995]). The well-known RGB colorspace is far from being perceptually uniform, the non-linear transformation to CIELAB and CIELUV try to correct the situation. The price for better perceptual uniformity is complex transformation functions from and to RGB space, demanding far more computation than most other colorspaces, described here. These colorspaces were used in [Zarit et al. 1999], [Yang and Ahuja 1999], [Schumeyer and Barner 1998], [Yang and Ahuja 1998]

Psychologist Farsnworth have proposed an even more perceptually uniform color system, derived from psychophysical experiments. It also uses nonlinear transforms from an RGB space. It was first used for the skin detection in [Chen et al. 1995].

2.7 RGB channels ratio

It was observed, that skin invariably contains a significant level of red. Using this observation, certain values of R/G ratio were used as skin presence indicators [Wark and Sridharan 1998]. Usefulness of other RGB-space ratios (R/B and G/B) for skin detection was tested and evaluated by [Brand and Mason 2000].

2.8 Other colorspaces

Besides YCrCb, several other linear transforms of the RGB space were employed for skin detection - YES [Saber and Tekalp 1998], YUV [Marques and Vilaplana 2000] and YIQ [Brand and Mason 2000], [C.Wang and M.Brandstein 1999]. Among less frequently used colorspaces, CIE-xyz [Terrillon et al. 2000] can be mentioned.

3 Skin modelling

The final goal of skin color detection is to build a decision rule, that will discriminate between skin and non-skin pixels. This is usually accomplished by introducing a metric, which measures distance (in general sense) of the pixel color to skin tone. The type of this metric is defined by the skin color modelling method.

3.1 Explicitly defined skin region

One method to build a skin classifier is to define explicitly (through a number of rules) the boundaries skin cluster in some colorspace. For example [Peer et al. 2003]:

$$\begin{aligned}&(R,G,B) \text{ is classified as skin if:}\\ &R > 95 \text{ and } G > 40 \text{ and } B > 20 \text{ and}\\ &max\{R,G,B\} - min\{R,G,B\} > 15 \text{ and}\\ &|R-G| > 15 \text{ and } R > G \text{ and } R > B\end{aligned} \tag{10}$$

The simplicity of this method have attracted (and still does) many researchers [Peer et al. 2003], [Ahlberg 1999], [Fleck et al. 1996], [Jordao et al. 1999]. The obvious advantage of this method is simplicity of skin detection rules that leads to construction of a very rapid classifier. The main difficulty achieving high recognition rates with this method is the need to find both good colorspace and adequate decision rules empirically. Recently, there have been proposed a method that uses machine learning algorithms to find both suitable colorspace and a simple decision rule that achieve high recognition rates [Gomez and Morales 2002]. The authors start with a normalized RGB space and then apply a constructive induction algorithm (see [Gomez and Morales 2002] for details) to create a number of new sets of three attributes being a superposition of r, g, b and a constant 1/3, constructed by basic arithmetic operations. A decision rule, similar to (10) that achieves the best possible recognition is estimated for each set of attributes. The authors prohibit construction of too complex rules, which helps avoiding data over-fitting, that is possible in case of lack of training set representativeness. They have achieved results that outperform Bayes skin probability map (see section 3.2.2) classifier in RGB space for their dataset.

3.2 Nonparametric skin distribution modelling

The key idea of the non-parametric skin modelling methods is to estimate skin color distribution from the training data without deriving an explicit model of the skin color. The result of these methods sometimes is referred to as construction of Skin Probability Map (SPM) [Brand and Mason 2000], [Gomez 2000] - assigning a probability value to each point of a discretized colorspace.

3.2.1 Normalized lookup table (LUT)

Several face detection and tracking algorithms [Chen et al. 1995], [Zarit et al. 1999], [Schumeyer and Barner 1998],[Sigal et al. 2000], [Soriano et al. 2000], [Birchfield 1998] use a histogram based-approach to skin pixels segmentation. The colorspace (usually, the chrominance plane only) is quantized into a number of bins, each corresponding to particular range of color component value pairs (in 2D case) or triads (in 3D case). These bins, forming a 2D or 3D histogram are referred to as the lookup table (LUT). Each bin stores the number of times this particular color occurred in the training skin images. After training, the histogram counts are normalized, converting histogram values to discrete probability distribution:

$$P_{skin}(c) = \frac{skin[c]}{Norm} \tag{11}$$

where $skin[c]$ gives the value of the histogram bin, corresponding to color vector c and $Norm$ is the normalization coefficient (sum of all histogram bin values [Jones and Rehg 1999], or maximum bin value present [Zarit et al. 1999]). The normalized values of the lookup table bins constitute the likelihood that the corresponding colors will correspond to skin.

3.2.2 Bayes classifier

The value of $P_{skin}(c)$ computed in (11) is actually a conditional probability $P(c|skin)$ - a probability of observing color c, knowing that we see a skin pixel. A more appropriate measure for skin detection would be $P(skin|c)$ - a probability of observing skin, given a concrete c color value. To compute this probability, the Bayes rule is used:

$$P(skin|c) = \frac{P(c|skin)P(skin)}{P(c|skin)P(skin) + P(c|\neg skin)P(\neg skin)} \tag{12}$$

$P(c|skin)$ and $P(c|\neg skin)$ are directly computed from skin and non-skin color histograms (11). The prior probabilities $P(skin)$ and $P(\neg skin)$ can also be estimated from the overall number of skin and non-skin samples in the training set [Jones and Rehg 1999], [Zarit et al. 1999], [Chai and Bouzerdoum 2000]. An inequality $P(skin|c) \geq \Theta$, where Θ is a threshold value, can be used as a skin detection rule [Jones and Rehg 1999]. Receiver operating characteristics (ROC) curve [Trees 1968] shows the relationship between correct detections and false detections for a classification rule as a function of the detection threshold. It turns out, that the ROC curve for $P(skin|c) \geq \Theta$ is invariant to choice of prior probabilities, due to nature of the Bayes model. This means that $P(skin)$ value affects only the choice of the threshold Θ.

One can avoid computing (12) explicitly, if what is really needed is the comparison of $P(skin|c)$ to $P(\neg skin|c)$, not their exact values. Using (12) the ratio of $P(skin|c)$ to $P(\neg skin|c)$ can be written as:

$$\frac{P(skin|c)}{P(\neg skin|c)} = \frac{P(c|skin)P(skin)}{P(c|\neg skin)P(\neg skin)} \tag{13}$$

Comparing (13) to a threshold produces the skin/non-skin decision rule. That after some manipulations, can be rewritten as:

$$\frac{P(c|skin)}{P(c|\neg skin)} > \Theta \qquad \Theta = K \times \frac{1-P(skin)}{P(skin)} \tag{14}$$

This shows, why the choice of prior probabilities does not affect the overall detector behavior - for any prior probability $P(skin)$ it is possible to choose the appropriate value of K, that gives the same detection threshold Θ. It is also clear, that maximum likelihood (ML) and maximum a posteriori (MAP) Bayes classification rules compared in [Zarit et al. 1999] are equivalent to (14) with different Θ values.

3.2.3 Self Organizing Map

Self-Organizing Map (or SOM), devised by Kohonen in 80's is now one of the most popular types of unsupervised artificial neural network. In [Brown et al. 2001] a SOM-based skin detector was proposed. Two SOM's - skin-only and skin + non-skin were trained from a set of about 500 manually labelled images. The detectors performance was tested on the authors training/test images set and famous Compaq skin database [Jones and Rehg 1999]. Several colorspaces (normalized RGB, Hue-Saturation, cartesian Hue-Saturation and chrominance plane of TSL) were tested with SOM detector. The results have shown, that SOM skin detectors do not exhibit vivid performance change when using different colorspaces. The SOM performance on the authors dataset is marginally better than Gaussian mixture model, while for the Compaq database the SOM performance is inferior to the RGB histograms used in [Jones and Rehg 1999]. The authors stress out that SOM method needs considerably less resource than histogram and mixture models and is efficiently implemented for run-time applications by the means of SOM hardware.

3.2.4 Non-parametric methods summary

Two clear advantages of the non-parametric methods are i. they are fast in training and usage and ii. they are theoretically independent to the shape of skin distribution (which is not true for explicit skin cluster definition and parametric skin modelling). The disadvantages are much storage space required and inability to interpolate or generalize the training data. If, for example, we consider RGB quantized to 8 bits per color, we'll need an array of 2^{24} elements to store skin probabilities. To reduce the amount of needed memory and to account for possible training data sparsity, coarser colorspace samplings are used - 128x128x128, 64x64x64 and 32x32x32. The evaluation of different RGB samplings in [Jones and Rehg 1999] has shown, that 32x32x32 shows the best performance.

3.3 Parametric skin distribution modelling

The most popular histogram-based non-parametric skin models require much storage space and their performance directly depends on the representativeness of the training images set. The need for more compact skin model representation for certain applications along with ability to generalize and interpolate the training data stimulates the development of parametric skin distribution models.

3.3.1 Single Gaussian

Skin color distribution can be modelled by an elliptical Gaussian joint probability density function (pdf), defined as:

$$p(c|skin) = \frac{1}{2\pi|\Sigma_s|^{1/2}} \cdot e^{-\frac{1}{2}(c-\mu_s)^T \Sigma_s^{-1}(c-\mu_s)} \tag{15}$$

Here, c is a color vector and μ_s and Σ_s are the distribution parameters (mean vector and covariance matrix respectively). The model parameters are estimated from the training data by (16):

$$\mu_s = \frac{1}{n}\sum_{j=1}^{n} c_j; \qquad \Sigma_s = \frac{1}{n-1}\sum_{j=1}^{n}(c_j-\mu_s)(c_j-\mu_s)^T \tag{16}$$

where n is the total number of skin color samples c_j. The $p(c|skin)$ probability can be used directly as the measure of how "skin-like" the c color is [Menser and Wien 2000], or, alternatively, the Mahalanobis distance from the c color vector to mean vector μ_s, given the covariance matrix Σ_s can serve for the same purpose [Terrillon et al. 2000]:

$$\lambda_s(c) = (c-\mu_s)^T \Sigma_s^{-1}(c-\mu_s) \tag{17}$$

Single Gaussian modelling method was also employed in [Hsu et al. 2002], [Ahlberg 1999], [Yang and Ahuja 1998], [Saber and Tekalp 1998].

3.3.2 Mixture of Gaussians

A more sophisticated model, capable of describing complex-shaped distributions is the Gaussian mixture model. It is the generalization of the single Gaussian, the pdf in this case is:

$$p(c|skin) = \sum_{i=1}^{k} \pi_i \cdot p_i(c|skin) \tag{18}$$

In (18) k is the number of mixture components, π_i are the mixing parameters, obeying the normalization constraint $\sum_{i=1}^{k}\pi_i = 1$, and $p_i(c|skin)$ are Gaussian pdfs, each with its own mean and covariance matrix. Model training is performed with a well-known iterative technique called the Expectation Maximization (EM) algorithm, which assumes the number of components k to be known beforehand. The details of training Gaussian mixture model with EM can be found, for example in [Yang and Ahuja 1999], [Terrillon et al. 2000]. The classification with a Gaussian mixture model is done by comparing the $p(c|skin)$ value to some threshold.

The choice of the components number k is important here. The model needs to explain the training data reasonably well with the given model on one hand, and avoid data over-fitting on the other. The number of components used by different researchers varies significantly - from 2 [Yang and Ahuja 1999] to 16 [Jones and Rehg 1999]. A bootstrap test for justification of $k = 2$ hypothesis was performed in [Yang and Ahuja 1999], in [Terrillon et al. 2000] $k = 8$ was chosen as a "good compromise between the accuracy of of estimation of the true distributions and the computational load for thresholding". [McKenna et al. 1998], [Oliver et al. 1997] have also used Gaussian mixture models.

3.3.3 Multiple Gaussian clusters

Approximation of skin color cluster with three 3D Gaussians in YCbCr space is described in [Phung et al. 2002]. A variant of k-means clustering algorithm for Gaussian clusters performs the model training. The pixel is classified as skin, if the Mahalanobis distance from the c color vector to the closest model cluster center is below a pre-defined threshold.

3.3.4 Elliptic boundary model

By examining skin and non-skin distributions in several colorspaces Lee and Yoo [Lee and Yoo 2002] have concluded that skin color cluster, being approximately elliptic in shape is not well enough approximated by the single Gaussian model. Due to asymmetry of the skin cluster with respect to its density peak, usage of the symmetric Gaussian model leads to high false positives rate. They propose an alternative they call an "elliptical boundary model" which is equally fast and simple in training and evaluation as the single Gaussian model and gives superior detection results on the Compaq database [Jones and Rehg 1999] compared both to single and mixture of Gaussians . The elliptical boundary model is defined as:

$$\Phi(c) = (c-\phi)^T \Lambda^{-1}(c-\phi) \tag{19}$$

The model training procedure has two steps - first, up to 5% of the training color samples with low frequency are eliminated to remove noise and negligible data. Then, model parameters (ϕ and Λ) are estimated by

$$\phi = \frac{1}{n}\sum_{i=1}^{n} c_i; \qquad \Lambda = \frac{1}{N}\sum_{i=1}^{n} f_i \cdot (c_i-\mu)(c_i-\mu)^T; \tag{20}$$
$$\mu = \frac{1}{N}\sum_{i=1}^{n} f_i c_i; \qquad N = \sum_{i=1}^{n} f_i$$

where n is the total number of distinctive training color vectors c_i of the training skin pixel set (not the total samples number!), and f_i is the number of skin samples of color vector c_i. Pixel with color c is classified as skin in case when $\Phi(c) < \theta$, where θ is a threshold value. The authors claim that their model approximates the skin cluster better, because the data skew does not affect the model centroid ϕ calculation.

3.3.5 Parametric methods summary

All described parametric methods (except described in Section 3.3.3) operate in colorspace chrominance plane, ignoring the luminance information.

Of course, since an explicit distribution model is used, a question of model validation arises. Obviously, the goodness of fit is more dependent on the distribution shape, and therefore colorspace used, for parametric than for non-parametric skin models. This is clearly visible in the results of [Terrillon et al. 2000], [Lee and Yoo 2002], where the model performance varies significantly from colorspace to colorspace.

Only several authors have included theoretical justification for the validity of models they used. [Yang et al. 1998] has shown that skin color distribution of a single person under fixed lighting conditions in normalized RGB space obeys Gaussian distribution. [Yang and Ahuja 1999] have justified the hypotheses of skin data normality in CIELuv space and validity of two-component Gaussian mixture model by statistical tests. Others relied whether on the observation of nearly elliptic shape of the skin chrominances cluster in the colorspace they used (to employ single Gaussian model or similar), or its clearly non-elliptical shape (to employ mixture of Gaussians or several Gaussian clusters) with further model performance evaluation as the acceptance criterion [Terrillon et al. 2000], [Lee and Yoo 2002], [McKenna et al. 1998].

3.4 Dynamic skin distribution models

A family of skin modelling methods was designed and tuned specifically for skin detection during face tracking. This task makes skin detection different from the static images analysis in several aspects. First, in principle, the skin model can be less general (more specific) - i.e. tuned for one concrete person, camera or lighting. Second, initialization stage is possible, when the face region is discriminated from background by different classifier or manually. This gives a possibility to obtain skin classification model, that is optimal for the given conditions (person, camera, lighting, background). Since there is no need for model generality, it is possible to reach higher skin detection rates with low false positives with this specific model, than with general skin color models, intended to classify skin in totally unconstrained images set (like in [Jones and Rehg 1999]). On the other hand, skin color distribution can vary with time, along with lighting or camera white balance change, so the model should be able to update itself to match the changing conditions. Also, model training and classification time becomes extremely important here, for the skin detection system must work at real-time, consuming little computing power.

To summarize the most important properties of skin color model for face tracking: first, it should be fast in both training and classification and second, it should be able to update itself to changing conditions. Minding these aspects, many researches turn to simple parametric skin modelling - it is easily updated to distribution change, is acceptably fast (except for many-component mixture of Gaussians) and needs little storage space. The high false positives rate - a usual companion of parametric skin modelling, is less a problem here. The need for specific, not general skin color model permits achievement of good classification performance. Among non-parametric models, the histogram-based LUT is popular for face tracking tasks, thanks to its simplicity and high training and working speed.

A number of methods for skin color distribution recalculation were proposed: online Expectation Maximization [Oliver et al. 1997], dynamic histograms [Soriano et al. 2000], [Stern and Efros 2002], [Sigal et al. 2000], Gaussian distribution adaptation [Yang et al. 1998].

Several authors have investigated how the color of a single person should be modelled and how it varies with lighting change. The hypothesis of unimodal Gaussian distribution of one person's skin color under fixed lighting was justified in [Yang et al. 1998]. A special study on skin color change under different lighting conditions was made by [M. Storring 1999] and [Martinkauppi et al. 2001]. An unusual method for automatic colorspaces switching during the face tracking was proposed in [Stern and Efros 2002]. See colorspaces discussion (Section 6) for more information on the two latter methods.

4 Comparative evaluation

For fair performance evaluation of different skin color modelling methods identical testing conditions are preferred. Unfortunately, many skin detection methods provide results on their own, publicly unavailable databases. The most famous training and test database for skin detection is the Compaq database [Jones and Rehg 1999]. In the table below the best results of different methods, reported by the authors, for this dataset are presented. Table 1 shows true positives (TP) and false positives (FP) rates for different methods configurations. Although different methods use slightly different separation of the database into training and testing image subsets and employ different learning strategies, the table should give an overall picture of the methods performance.

Method	TP	FP
Bayes SPM in RGB [Jones and Rehg 1999]	80% 90%	8.5% 14.2%
Bayes SPM in RGB [Brand and Mason 2000]	93.4%	19.8%
Maximum Entropy Model in RGB [Jedynak et al. 2002]	80%	8%
Gaussian Mixture models in RGB [Jones and Rehg 1999]	80% 90%	$\sim$ 9.5% $\sim$ 15.5%
SOM in TS [Brown et al. 2001]	78%	32%
Elliptical boundary model in CIE-xy [Lee and Yoo 2002]	90%	20.9%
Single Gaussian in CbCr [Lee and Yoo 2002]	90%	33.3%
Gaussian Mixture in IQ [Lee and Yoo 2002]	90%	30.0%
Thresholding of I axis in YIQ [Brand and Mason 2000]	94.7%	30.2%

Table 1: Performance of different skin detectors reported by the authors

The best performance (lower false positives for a given correct detection rate) is demonstrated by Bayes SPM and it's descendant - maximum entropy model [Jedynak et al. 2002]. The parametric modelling techniques (Gaussian, mixture of Gaussians, elliptic boundary model) are left behind together with SOM-based detector. High performance of the mixture of Gaussians used in [Jones and Rehg 1999] is due to the fact, that they actually modelled both $p(RGB|skin)$ and $p(RGB|\neg skin)$ pdfs (in contrast to other parametric skin modelling papers). They did not provide a clear indication on how exactly the final skin probability was computed from these pdfs, so we conclude that Bayesian rule was used (14). This, altogether with high number of mixture components (sixteen) makes this model an approximation of Bayes SPM. We believe that this the explanation of high performance of Gaussian mixture model of

Jones and Rehg. A fact worth noting is that simple thresholding of I component of YIQ space, proposed by [C.Wang and M.Brandstein 1999] and evaluated in [Brand and Mason 2000] shows result comparable to more sophisticated Gaussian and mixture of Gaussians skin models.

Another promising method, appeared recently, which is not included in this table, is automatic construction of a colorspace and an explicitly defined skin cluster in it [Gomez 2000], [Gomez and Morales 2002] (refer back to Section 3.1 for more details). The authors have achieved results that outperform Bayes SPM classifier in RGB space for their dataset, giving significantly lower false positives rate (around 6% against 22%) and almost equal false negatives (around 5%).

5 Methods discussion

The main advantage of the methods that use explicitly defined skin cluster boundaries (section 3.1) is the simplicity and intuitiveness of the classification rules. However, the difficulty with them is the need to find both good colorspace and adequate decision rules empirically. The recently proposed method that uses machine learning algorithms to find both suitable colorspace and simple decision rules [Gomez and Morales 2002] has shown a way to overcome these difficulties.

The non-parametric methods (section 3.2) are fast both in training and classification, independent to distribution shape and therefore to colorspace selection (see *colorspaces discussion* section for more information on the topic). But, they require much storage space and a representative training dataset.

The parametric methods (section 3.3) can also be fast, they have a useful ability to interpolate and generalize incomplete training data, they are expressed by a small number of parameters and need very little storage space. However, they can be really slow (like mixture of Gaussians) in both training and work, and their performance depends strongly on the skin distribution shape. Besides, most parametric skin modelling methods ignore the non-skin color statistics. This, together with dependance on skin cluster shape results in higher false positives rate, compared to non-parametric methods.

6 Colorspaces discussion

At a first glance, colorspace selection seems to be crucial for color-based skin detection. One important question is: what is the best colorspace for skin detection, or more generally - is there an optimal colorspace for skin-classification? Surprisingly, many papers on skin detection do not provide strict justification of their colorspace choice, probably because of possibility to obtain acceptable skin detection results on limited dataset with almost any colorspace.

Only few papers have been devoted to comparative analysis of different colorspaces used for skin detection [Zarit et al. 1999], [Terrillon et al. 2000], [Gomez 2000], [Gomez and Morales 2002], [Stern and Efros 2002]. Several authors have seriously considered the problem of colorspace selection, and have provided justifications for the optimality (or adequateness) of their choice for the skin model they employed [Yang and Ahuja 1999], [Yang et al. 1998], [M. Storring 1999], [Schumeyer and Barner 1998]. The colorspace 'goodness' for skin modelling is usually evaluated by two different families of measures. First is training and test set classification error, computed after color model parameter estimation. It is a well-known classifier performance evaluation principle, which clearly indicates the goodness-of-fit of the selected model to the given dataset. The second family of measures is skin and non-skin colors overlap in the given colorspace and compactness of the skin cluster. These measures are independent to color modelling strategy and are determined to evaluate the colorspace goodness 'in general'. They surely can provide an overall impression on the distribution of the skin and non-skin samples of the training set, but their feasibility for evaluation of the colorspace goodness seems doubtful to us.

Recently, there emerged several papers that seriously doubt any significant influence of colorspace selection on the final skin detection result [Shin et al. 2002], [Albiol et al. 2001]. In [Shin et al. 2002] the authors have used scatter matrices of skin and non-skin clusters and skin and non-skin histograms overlap as colorspace performance metrics. Their conclusion is that skin and non-skin color classes separability is highest in RGB space, and that dropping luminance component significantly worsens the separability. We do not quite agree the colorspace comparison strategy, carried out in [Shin et al. 2002]. Our strong belief is that valid colorspace comparison is the one carried not 'in general' (by assessing skin and non-skin colors overlap and skin cluster shape), but for a certain skin distribution model. The performance of parametric skin classifiers depends heavily on the colorspace choice - this can be observed by the results obtained in [Terrillon et al. 2000], [Lee and Yoo 2002]. The methods, that use explicitly defined skin region also benefit much by appropriate colorspace choice [Peer et al. 2003], [Gomez 2000], [Gomez and Morales 2002]. The non-parametric methods (Bayes SPM, SOM, LUT), on the contrary, are almost independent to the colorspace choice [Zarit et al. 1999], [Brown et al. 2001], [Albiol et al. 2001]. We believe, that skin and non-skin overlap damages heavily the performance of parametric skin color models [Terrillon et al. 2000], [Lee and Yoo 2002] and lookup-table (LUT) method [Zarit et al. 1999], because this overlap is not taken into account by the model. The independence on colorspace choice for most non-parametric models fits well with theoretical results obtained in [Albiol et al. 2001]. The authors state that for an optimal skin detector $D(x)$ in colorspace C, and for an invertible colorspace transformation rule $T : C \rightarrow C'$, there exists a classifier $D'(x')$ in C' colorspace, that has the same correct detection and false positive rates. The authors give an example of Bayes SPM, that performs almost equally in several colorspaces.

Many works on skin detection drop the luminance component of the colorspace. This decision seems logical, as the goal is to model what can be thought of "skin tone", which is more controlled by the chrominance than luminance coordinates. The dimensionality reduction, achieved by discarding luminance also simplifies the consequent color analysis. Another argument for ignoring luminance is that skin color differs from person to person mostly in brightness and less in the tone itself. The illumination conditions clearly affect the color of the objects in the scene. The goal of any color-based system is diminishing this influence to make color-based recognition robust to illumination change. It seems, that chrominance-only color analysis should render the system partially independent from the lighting conditions. The profit of luminance component removal, that seemed perfectly logical for many researchers before, was doubted by [Shin et al. 2002]. The tests the authors have performed have shown that luminance removal does not increase separability of skin and non-skin clusters. This is, of course, true because the projection of 3D data on a plane almost certainly smears skin and non-skin classes together. But we think, that dropping luminance is a matter of training data generalization. For a training dataset with sparse distribution of skin luminances (e.g. little number of face image under similar lighting conditions) the removal of luminance component helps constructing skin classifier that will also work for images with different lighting intensity. Also, the reduction of space dimensionality is very attractive in some cases.

An interesting study of skin color distribution behavior under changing lighting conditions was performed in [M. Storring 1999] and [Martinkauppi et al. 2001]. The authors have shown that for different lighting conditions the skin color from their dataset (of approximately 125 individuals) lies inside a definitely shaped colorspace region - the so-called *skin locus*, that can be modelled by one or two functions of up to quadratic order only. The locus is camera-specific and is used by [Soriano et al. 2000] as the skin color filter for dynamic skin histogram updating during face tracking. The locus may be found experimentally or, in principle, may be calculated. A database of illuminants, skin spectral reflectance and a knowledge of the camera sensitivities (for example, supplied by the manufacturer) can allow the user to compute the camera skin locus ([Martinkauppi and Soriano 2001]).

An adaptive colorspace switching method for face tracking was proposed in [Stern and Efros 2002]. The optimal colorspace for a given video frame is determined by a simple colorspace quality measure. The dynamic change of the colorspace is intended to contribute to robustness of the face tracking method. However, judging from the experimental data the authors have provided, among five colorspace chromaticity planes (normalized RGB, HS, YQ, and CrCb) and RG plane of the RGB space, the normalized RGB and HS planes performed almost equally and much better than the others. This suggests that little was gained by the adaptive colorspace switching, if compared to using solely HS or normalized RGB.

7 Conclusion

In this paper, we have provided the description, comparison and evaluation results of popular methods for skin modelling and detection. We tried to summarize the most notable and significant differences between the methods, their advantages and disadvantages. The most important conclusions we draw are listed below:

- Parametric skin modelling methods are better suited for constructing classifiers in case of limited training and expected target data set. The generalization and interpolation ability of these methods makes it possible to construct a classifier with acceptable performance from incomplete training data.

- The methods that are less dependent on the skin cluster shape and take into account skin and non-skin colors overlap (Bayes SPM, Maximum entropy model [Jedynak et al. 2002], automatically constructed colorspace and classification rules [Gomez and Morales 2002]) look more promising for constructing skin classifier for large target datasets.

- Excluding color luminance from the classification process cannot help achieving better discrimination of skin and non-skin colors, but can help to generalize sparse training data.

- Evaluation of colorspace goodness 'in general' by assessing skin/non-skin overlap, skin cluster shape, etc. regardless to any specific skin modelling method cannot give the impression of how good is the colorspace suited for skin modelling, because different modelling methods react very differently on the colorspace change.

8 Acknowledgements

We would like to express our gratitude to our colleague Alexey Ignatenko, who has provided us with a convenient statistical data graphical analysis tool, that was used for testing of skin color modelling methods.

References

AHLBERG, J. 1999. A system for face localization and facial feature extraction. Tech. Rep. LiTH-ISY-R-2172, Linkoping University.

ALBIOL, A., TORRES, L., AND DELP, E. J. 2001. Optimum color spaces for skin detection. In *Proceedings of the International Conference on Image Processing*, vol. 1, 122–124.

BIRCHFIELD, S. 1998. Elliptical head tracking using intensity gradients and color histograms. In *Proceedings of CVPR '98*, 232–237.

BRAND, J., AND MASON, J. 2000. A comparative assessment of three approaches to pixellevel human skin-detection. In *Proc. of the International Conference on Pattern Recognition*, vol. 1, 1056–1059.

BROWN, D., CRAW, I., AND LEWTHWAITE, J. 2001. A som based approach to skin detection with application in real time systems. In *Proc. of the British Machine Vision Conference, 2001.*

CHAI, D., AND BOUZERDOUM, A. 2000. A bayesian approach to skin color classification in ycbcr color space. In *Proceedings IEEE Region Ten Conference (TENCON'2000)*, vol. 2, 421–424.

CHEN, Q., WU, H., AND YACHIDA, M. 1995. Face detection by fuzzy pattern matching. In *Proc. of the Fifth International Conference on Computer Vision*, 591–597.

C.WANG, AND M.BRANDSTEIN. 1999. Multi-source face tracking with audio and visual data. In *IEEE MMSP*, 169–174.

FLECK, M., FORSYTH, D. A., AND BREGLER, C. 1996. Finding naked people. In *Proc. of the ECCV*, vol. 2, 592–602.

GOMEZ, G., AND MORALES, E. 2002. Automatic feature construction and a simple rule induction algorithm for skin detection. In *Proc. of the ICML Workshop on Machine Learning in Computer Vision*, 31–38.

GOMEZ, G. 2000. On selecting colour components for skin detection. In *Proc. of the ICPR*, vol. 2, 961–964.

HSU, R.-L., ABDEL-MOTTALEB, M., AND JAIN, A. K. 2002. Face detection in color images. *IEEE Trans. Pattern Analysis and Machine Intelligence 24*, 5, 696–706.

JEDYNAK, B., ZHENG, H., DAOUDI, M., AND BARRET, D. 2002. Maximum entropy models for skin detection. Tech. Rep. XIII, Universite des Sciences et Technologies de Lille, France.

JONES, M. J., AND REHG, J. M. 1999. Statistical color models with application to skin detection. In *Proc. of the CVPR '99*, vol. 1, 274–280.

JORDAO, L., PERRONE, M., COSTEIRA, J., AND SANTOS-VICTOR, J. 1999. Active face and feature tracking. In *Proceedings of the 10th International Conference on Image Analysis and Processing*, 572–577.

KRUPPA, H., BAUER, M. A., AND SCHIELE, B. 2002. Skin patch detection in real-world images. In *Annual Symposium for Pattern Recognition of the DAGM 2002, Springer LNCS 2449*, 109–117.

LEE, J. Y., AND YOO, S. I. 2002. An elliptical boundary model for skin color detection. In *Proc. of the 2002 International Conference on Imaging Science, Systems, and Technology*.

M. STORRING, H. ANDERSEN, E. G. 1999. Skin colour detection under changing lighting condition. In *Araujo and J. Dias (ed.) 7th Symposium on Intelligent Robotics Systems*, 187–195.

MARQUES, F., AND VILAPLANA, V. 2000. A morphological approach for segmentation and tracking of human faces. In *International Conference on Pattern Recognition (ICPR'00)*, vol. 1, 5064–5068.

MARTINKAUPPI, B., AND SORIANO, M. 2001. Basis functions of the color signals of skin under different illuminants. In *3rd Intl conference on Multispectral Color Science*, 21–24.

MARTINKAUPPI, B., LAAKSONEN, M., AND SORIANO, M. 2001. Behavior of skin color under varying illumination seen by different cameras at different color spaces. In *Machine Vision Applications in Industrial Inspection IX, Proceedings of SPIE*, vol. 4301, 102–112.

MCKENNA, S., GONG, S., AND RAJA, Y. 1998. Modelling facial colour and identity with gaussian mixtures. *Pattern Recognition 31*, 12, 1883–1892.

MENSER, B., AND WIEN, M. 2000. Segmentation and tracking of facial regions in color image sequences. In *Proc. SPIE Visual Communications and Image Processing 2000*, 731–740.

OLIVER, N., PENTLAND, A., AND BERARD, F. 1997. Lafter: Lips and face real time tracker. In *Proc. Computer Vision and Pattern Recognition*, 123–129.

PEER, P., KOVAC, J., AND SOLINA, F. 2003. Human skin colour clustering for face detection. In *submitted to EUROCON 2003 - International Conference on Computer as a Tool*.

PHUNG, S. L., BOUZERDOUM, A., AND CHAI, D. 2002. A novel skin color model in ycbcr color space and its application to human face detection. In *IEEE International Conference on Image Processing (ICIP'2002)*, vol. 1, 289–292.

POYNTON, C. A. 1995. Frequently asked questions about colour. In *ftp://www.inforamp.net/pub/users/poynton/doc/colour/ColorFAQ.ps.gz*.

SABER, E., AND TEKALP, A. 1998. Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions. *Pattern Recognition Letters 19*, 8, 669–680.

SCHUMEYER, R., AND BARNER, K. 1998. A color-based classifier for region identification in video. In *Visual Communications and Image Processing 1998, SPIE*, vol. 3309, 189–200.

SHIN, M. C., CHANG, K. I., AND TSAP, L. V. 2002. Does colorspace transformation make any difference on skin detection? In *IEEE Workshop on Applications of Computer Vision*.

SIGAL, L., SCLAROFF, S., AND ATHITSOS, V. 2000. Estimation and prediction of evolving color distributions for skin segmentation under varying illumination. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, 152–159.

SKARBEK, W., AND KOSCHAN, A. 1994. Colour image segmentation – a survey –. Tech. rep., Institute for Technical Informatics, Technical University of Berlin, October.

SORIANO, M., HUOVINEN, S., MARTINKAUPPI, B., AND LAAKSONEN, M. 2000. Skin detection in video under changing illumination conditions. In *Proc. 15th International Conference on Pattern Recognition*, vol. 1, 839–842.

STERN, H., AND EFROS, B. 2002. Adaptive color space switching for face tracking in multi-colored lighting environments. In *Proc. of the International Conference on Automatic Face and Gesture Recognition*, 249–255.

TERRILLON, J.-C., SHIRAZI, M. N., FUKAMACHI, H., AND AKAMATSU, S. 2000. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In *Proc. of the International Conference on Face and Gesture Recognition*, 54–61.

TREES, H. L. V. 1968. *Detection, Estimation, and Modulation Theory*, vol. I. Wiley.

WARK, T., AND SRIDHARAN, S. 1998. A syntactic approach to automatic lip feature extraction for speaker identification. In *ICASSP*, 3693–3696.

YANG, M.-H., AND AHUJA, N. 1998. Detecting human faces in color images. In *International Conference on Image Processing (ICIP)*, vol. 1, 127–130.

YANG, M., AND AHUJA, N. 1999. Gaussian mixture model for human skin color and its application in image and video databases. In *Proc. of the SPIE: Conf. on Storage and Retrieval for Image and Video Databases (SPIE 99)*, vol. 3656, 458–466.

YANG, J., LU, W., AND WAIBEL, A. 1998. Skin-color modeling and adaptation. In *Proceedings of ACCV 1998*, 687–694.

ZARIT, B. D., SUPER, B. J., AND QUEK, F. K. H. 1999. Comparison of five color models in skin pixel classification. In *ICCV'99 Int'l Workshop on recognition, analysis and tracking of faces and gestures in Real-Time systems*, 58–63.