Computational Strategies for Skin Detection

Simone Bianco, Francesca Gasparini, and Raimondo Schettini

University of Milano-Bicocca,
Viale Sarca 336, ed. U14, 20126 Milano, Italy
{simone.bianco,gasparini,schettini}@disco.unimib.it
http://www.ivl.disco.unimib.it

Abstract. In this paper we compare different computational strategies for skin detection. They differ in the type of data used in the training phase, the type of pre-processing done on the query image, and the level of visual information used. In particular, we define a high-level computational strategy, which uses a face detector in the pre-processing step. Two different implementations of it are proposed: one relies on an adaptive single gaussian model, the other a fixed threshold skin cluster detector on an illuminant-independent image representation. The experimental results on a heterogeneous dataset containing images acquired under uncontrolled lighting conditions show that the high-level strategies outperform low-level ones.

Keywords: Skin detection, skin segmentation, skin classification, skin cluster model, parametric skin model, non-parametric skin model.

1 Introduction

The detection of skin regions in color images is a preliminary step in many applications, such as image and video classification and retrieval in multimedia databases, semantic filtering of web contents (through the definition of medium-level features), human motion detection, human computer interaction and video-surveillance. It can also be useful in image processing algorithms, as well as in intelligent scanners, digital cameras, photocopiers, and printers. Many different methods for discriminating between skin and non-skin pixels are available in the literature. These can be grouped in three types of skin modeling: parametric, nonparametric, and explicit skin cluster definition methods, [8].

The simplest, and often applied, methods build what is called an explicit skin cluster classifier which expressly defines the boundaries of the skin cluster in certain color spaces, [18,11,14,10,15,16,17]. The hypothesis underlying these methods is that skin pixels exhibit similar color coordinates in an appropriately chosen color space.

Parametric Gaussian models [9,1,2] assume that skin color distribution can be modeled by an elliptical Gaussian joint probability density function. These parametric methods have the useful ability of interpolating and generalizing incomplete training data; they are expressed by a small number of parameters, and

S. Tominaga, R. Schettini, and A. Trémeau (Eds.): CCIW 2013, LNCS 7786, pp. 199–211, 2013.

require very little storage space. However their performance depends strongly on the skin distribution of the training images in the selected color space.

The key feature of non-parametric skin modeling methods is that the skin color distribution is estimated directly on the basis of the training data, without deriving an explicit model of the skin color. The result of these methods is sometimes referred to as a Skin Probability Map (SPM), [25,26]. We can take as an example the histogram-based non-parametric skin model, [13,27,28]. These nonparametric methods are quick trained and do not, theoretically, depend on the shape of the skin distribution (as, instead, explicit skin cluster definition and parametric modeling do).

In real world environment, the skin color strongly varies due to camera settings, illumination, peoples tans, and ethnic groups. The performance of skin classification is strongly dependent on the skin samples used to train the different methods and ideally a training set should be chosen to adapt for each different application.

In this paper we compare different computation strategies for skin detection. They differ in the type of data used in the training phase (i.e. sample images vs measured skin reflectances), the type of pre-processing done on the query image, and the type of information used (low-level vs high-level features). In particular, we propose a high-level computational strategy, which uses a face detector [29] in the pre-processing step. Within this strategy, we further propose two different implementations: the former exploits an adaptive skin detector, the latter exploits a fixed threshold skin cluster detector on an illuminant-independent image representation. Different strategies are here described with reference to the specific algorithms considered and objectively compared on a heterogeneous dataset of skin images.

2 Computational Strategies Considered

A skin detection strategy can be seen as a pipeline composed of three main steps: first a training phase, then an eventually pre-processing step and finally the detection through the skin model adopted. We here propose to distinguish the computational strategies for skin detection not with respect to the model adopted to classify skin vs no skin pixels, but instead with respect to the information used in the training phase. We distinguish between strategies based on a low-level training phase on skin databases, and high-level strategies where the training phase exploits information derived from an analysis of automatically detected faces.

2.1 Low-Level Strategies

Within the low-level strategies, we consider a further subdivision: methods trained on database of real images, and methods trained on a database of measured skin reflectances [30].

Training on Real Images. The generic pipeline of low-level strategies trained on real images is depicted in Figure 1. The model training is performed on the well known Compaq skin database [13]. This database is composed of images randomly picked from the World Wide Web, manually labeled into skin and non-skin pixels. No pre-processing algorithms are here applied before skin detection. In this work we have six skin detectors among those available in the literature, that can be grouped into three model types: explicit skin cluster models, parametric models, and non-parametric ones.

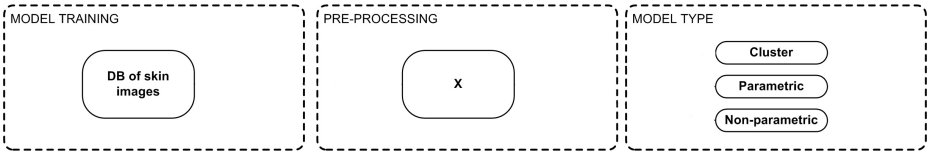

Fig. 1. Low level strategies trained on real images

The boundaries of the color skin cluster in a given color space are usually defined through simple, heuristically chosen decision rules. Gasparini et al. [3] have considered seven explicit skin cluster methods, working within different color spaces, and often cited in the literature. They have redefined the boundaries using genetic algorithms in which the fitness function is a weighted harmonic mean of precision and recall. To meet the widely varying requisites of different applications, the weighting coefficients were chosen to offer either high recall or high precision, or to satisfy a reasonable trade-off between the two.

We have here considered the following three methods among all the 21 possible ones presented in [3]:

- The best method in terms of Recall, with the boundaries obtained in [3]. This method works in the YCbCr color space and was introduced by Chai and Ngan in [11].
- The best method in terms of Precision, with the boundaries obtained in [3]. This method works in the HSI color space and was introduced by Hsieh et al. in [18].
- The best method in a tradeoff strategy, with the boundaries obtained in [3]. This method works in the HSV color space and was introduced by Tsekeridou and Pitias in [10].

We have here considered also two different non-parametric methods. The first one is a non-parametric histogram-based model developed by Conaire et al. [7]. The second one was introduced by Chai and Bouzerdoum [12] and uses the Bayes decision rule for minimum cost to classify pixels into skin color and non-skin color. Color statistics are collected from YCbCr color space. As a parametric model, we have considered a Gaussian Mixture Model with two components, as described by [2].

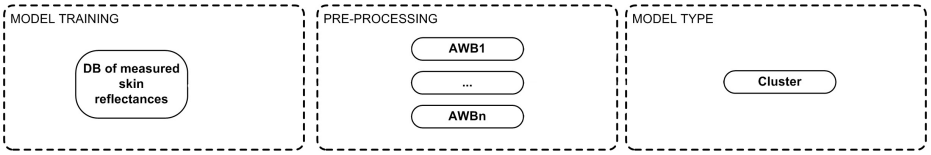

Fig. 2. Low-level strategies trained on measured skin reflectances

Training on Measured Skin Reflectances. The underlying idea of these methods is to use in the training, datasets obtained from a large set of measured skin reflectances. We considered the D65 CIE standard illuminant [5] and map the samples within the sRGB, YCbCr and HSV color spaces. Within these color spaces, we set the boundaries of the skin color cluster so that all the elements of the training set are included. The pipeline of this skin detection strategy is depicted in Figure 2. In the pre-processing step, a white balance algorithm is applied to discard the effect of the eventual illuminant. Finally the trained cluster classifier is applied to the balanced images.

In this work we have considered seven different color constancy algorithms. They can be obtained from a unique framework recently defined by Van de Weijer et al. [22]. These algorithms estimate the illuminant color $\mathbf{I}$ by implementing instantiations of the following equation:

$$\mathbf{I}(n,p,\sigma) = \frac{1}{k}\left(\iint |\nabla^n \boldsymbol{\rho}_\sigma(x,y)|^p \,\mathrm{d}x\,\mathrm{d}y\right)^{\frac{1}{p}}, \tag{1}$$

where n is the order of the derivative, p is the Minkowski norm, $\boldsymbol{\rho}_\sigma(x,y) = \boldsymbol{\rho}(x,y) \otimes G_\sigma(x,y)$ is the convolution of the image $\boldsymbol{\rho}(x,y)$ with a Gaussian filter $G_\sigma(x,y)$ with scale parameter σ, and k is a constant to be chosen such that the illuminant color $\mathbf{I}$ has unit length (using the $2-$norm). The integration is performed over all pixel coordinates. Different (n,p,σ) combinations correspond to different illuminant estimation algorithms, each based on a different assumption.

The values chosen for (n,p,σ) are reported in Table 1 and set as in [23].The algorithms are used in the original authors' implementation which is freely available online, [5].

Table 1. Values chosen for (n,p,σ) for the state-of-the-art algorithms which are instantiations of Eq.1

Algorithm	n	p	σ
Gray World (GW)	0	1	0
White Point (WP)	0	∞	0
Shades of Gray (SoG)	0	12	0
1st-order Gray Edge (GE1)	1	1	1
2nd-order Gray Edge (GE2)	2	1	2

The last algorithm considered is the Do Nothing (DN) algorithm which gives the same estimation for the color of the illuminant ($\mathbf{I} = [1\ 1\ 1]$) for every image, i.e. it assumes that the image is already correctly balanced.

2.2 High-Level Strategies

We here investigate high-level strategies for skin detection, where reliable pixels to be used in the training phase are extracted from automatically detected face regions. The flowchart of the two proposed strategies is depicted in Figure 3. The images for the training belong to a database of real images. In the pre-processing module, a face detector [4] is run on the input image to detect any faces. If no faces are detected, the input images can be processed with any other state of the art skin detector. If one or more faces are detected, a preliminary skin detection module is run on them to filter out any unreliable pixel. Reliable skin pixels are used to train the chosen skin detection model.

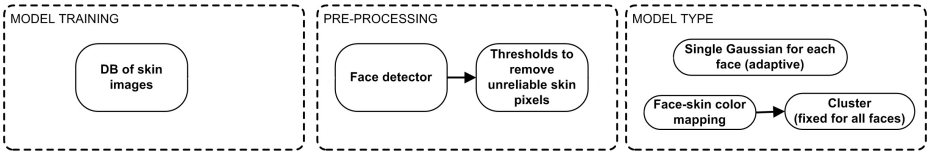

Fig. 3. High level

Pre-processing. Looping on all the faces detected, the face pixels are converted into the HSV color space. To select the reliable skin pixels, the same technique described in [20] is used. It is based on scale-space histogram filtering [19] to identify the highest peak location and width of the histogram of the hue component, within the hue interval corresponding to feasible skin colors. The boundaries of the reliable skin region are obtained from the training images. We here propose two different skin-detection model types: an adaptive single Gaussian model, and a color gamut mapping model.

Adaptive Single Gaussian Model (ASG). For each face detected in the image the color distribution of the reliable skin pixels is modeled with a single Gaussian in the HS plane of the HSV color space. This is an adaptive detector which builds a different model for each face. Each model is applied independently to the image under consideration. The results of the different detectors are combined in a recall-oriented scheme. The optimal threshold for the Gaussian models (which is fixed for all the detected faces) is found from the training images.

Color Gamut Mapping Model (CGM). All the images in the training dataset are processed and the reliable skin pixels found within all the detected faces are accumulated in the original image color space, i.e. sRGB. Similarly to [20], where the accumulated skin pixels were used to estimate the illuminant color with a gamut mapping approach, here the accumulated skin pixels are mapped to generate an illuminant-invariant skin gamut. Once this gamut has been obtained, a skin cluster detector with fixed thresholds is applied to the gamut mapped image. The optimal location of the illuminant-invariant skin gamut and the optimal thresholds for the skin cluster detector are found on the training images.

3 Experimental Results

All the experiments here reported were obtained using the Test Database for Skin Detection (TDSD) [24], which contains 554 images where skin pixels have been manually labeled. This database is a collection of skin images with at least one face, acquired under various lighting conditions and from different ethnic groups.

To quantify the performance of the skin detection methods presented, we use recall, precision and accuracy measures. Classification results are assigned as true positive (TP), false positive (FP) and false negative (FN). Recall is defined as follows:

$$\text{recall} = \frac{TP}{TP + FN} \tag{2}$$

Precision is defined as:

$$\text{precision} = \frac{TP}{TP + FP} \tag{3}$$

while Accuracy is defined as the ratio between the number of pixels correctly classified (both skin and no-skin) and the total number of pixels considered, i.e.:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

To summarize the performance of each strategy, the recall and precision values are combined into a single value using the F_1 measure:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{5}$$

The median values of the precision, recall, accuracy, and F_1 measure on the whole dataset are reported in Table 2.

The overall performance are finally compared using the Wilcoxon Signed-Rank Test on the F_1 distributions of each algorithm on the whole image database. This statistical test permits to compare the whole error distributions without limiting to punctual statistics. Let X and Y be random variables representing the F_1 measure distributions obtained on all the test images by the computational strategies C_X and C_Y; let μ_X and μ_Y be the median values of such random variables. The Wilcoxon signed-rank test can be used to test the null hypothesis $H_0 : \mu_X = \mu_Y$ against the alternative hypothesis $H_1 : \mu_X \neq \mu_Y$. We can test H_0 against H_1 at a given significance level α. We reject H_0 and accept H_1 if the probability of observing the error differences we obtained is less than or equal to α. We have used the alternative hypothesis $H_1 : \mu_X > \mu_Y$ with a significance level $\alpha = 0.05$.

We here report the results of the Wilcoxon test on the precision (Table 3), on the recall (Table 4), on the accuracy (Table 5), and on the F_1 measure (Table 6). A "+" sign in the (i, j)-position of the table means that the computational strategy i has been considered statistically better than the computational strategy j; a "-" sign that has been considered statistically worse, and a "=" sign that they have been considered statistically equivalent. The count of the number of times that a computational strategy has been considered statistically better than the others gives us a score which is reported in the last column of the tables.

The Wilcoxon scores are reported in Table 7, together with the average and median scores.

As a general comment, from the experiments comes that the high-level strategies outperform low-level ones. In particular, the Wilcoxon test on the F_1 measure (Table 6) ranks them as the best two strategies among the ones considered. The average and median of all the Wilcoxon scores, last two columns of Table 7, confirm the effectiveness of the high-level strategies, as ASG and CGM have the higher values. For what concerns low-level strategies, those trained with an uncontrolled database (first six methods) outperform the ones trained on the measured dataset of skin reflectances (labeled from 7 to 24). This is probably due to the fact that the skin reflectance database is not representative of all the possible reflectances of real world. However, each single cluster method trained on the skin reflectance database, improves when applied after a color balance algorithm, as emerges comparing in all the tests, methods 7, with methods from 10 to 14, method 8 with methods from 15 to 19, and method 9 with methods from 20 to 24.

Figure 4 depicts the results of applying the CGM strategy to two sample images of the TDSD: left column, the original images, middle column, the two skin masks, and finally, last column the corresponding skin pixels detected.

Table 2. Median values of the precision, recall, accuracy, and F_1 measure on the whole dataset

Strategy type	Strategy	Precision	Recall	Accuracy	F_1 measure
Low-level	HSI precision-oriented	0.8101	0.7043	0.8936	0.6804
	YCBCR recall-oriented	0.4389	0.9982	0.7268	0.6066
	HSV1 trade-off	0.7175	0.8214	0.8806	0.7019
	Bayesian	0.6385	0.9470	0.8689	0.7246
	Gaussian Mixture Model	0.4621	0.8242	0.7793	0.5378
	Histogram	0.5652	0.5428	0.8193	0.4971
	HSV D65+DN	0.4542	0.1528	0.8104	0.2167
	RGB D65+DN	0.2617	0.3879	0.6809	0.2926
	YCBCR D65+DN	0.2148	0.2246	0.7000	0.2070
	HSV D65+GW	0.5026	0.0830	0.8187	0.1371
	HSV D65+WP	0.4679	0.1624	0.8098	0.2318
	HSV D65+SoG	0.5440	0.1549	0.8140	0.2222
	HSV D65+GE1	0.5195	0.1564	0.8135	0.2218
	HSV D65+GE2	0.5046	0.1503	0.8120	0.2133
	RGB D65+GW	0.3398	0.5167	0.7123	0.3784
	RGB D65+WP	0.2731	0.4067	0.6782	0.2964
	RGB D65+SoG	0.3292	0.5520	0.6954	0.3675
	RGB D65+GE1	0.2987	0.4767	0.6943	0.3427
	RGB D65+GE2	0.2925	0.4332	0.6886	0.3252
	YCBCR D65+GW	0.4959	0.3615	0.8193	0.3803
	YCBCR D65+WP	0.2345	0.2559	0.7080	0.2293
	YCBCR D65+SoG	0.3873	0.3966	0.7733	0.3579
	YCBCR D65+GE1	0.2915	0.3033	0.7452	0.2715
	YCBCR D65+GE2	0.2813	0.2861	0.7385	0.2633
High-level	ASG	0.6860	0.9226	0.8869	0.7621
	CGM	0.8016	0.8210	0.9092	0.7760

Fig. 4. Left column, two test images belonging to the TDSD database. Middle column, the two skin masks obtained applying the CGM strategy. Last column, the corresponding skin pixels detected.

Table 3. The Wilcoxon test on the Precision for all the 26 skin detection strategies considered. Each algorithm score is reported in the last column.

Strategy	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Score
1) HSI precision-oriented	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	=	24
2) YCBCR recall-oriented	-	=	-	-	=	-	=	+	+	-	-	-	-	-	+	+	+	+	+	-	+	+	+	+	-	-	11
3) HSV1 trade-off	-	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	23
4) Bayesian	-	+	-	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	21
5) Gaussian Mixture Model	-	=	-	-	=	-	=	+	+	-	=	-	-	-	+	+	+	+	+	=	+	+	+	+	-	-	11
6) Histogram	-	+	-	-	+	=	+	+	+	+	+	=	+	+	+	+	+	+	+	+	+	+	+	+	-	-	19
7) HSV D65+DN	-	=	-	-	=	-	=	+	+	-	=	-	-	-	+	+	+	+	+	=	+	+	+	+	-	-	11
8) RGB D65+DN	-	-	-	-	-	-	-	=	+	-	-	-	-	-	-	=	-	-	-	+	-	=	=	-	-	-	2
9) YCBCR D65+DN	-	-	-	-	-	-	-	-	=	-	-	-	-	-	-	-	-	-	-	-	=	-	-	-	-	-	0
10) HSV D65+GW	-	+	-	-	+	-	+	+	+	=	=	-	=	=	+	+	+	+	+	=	+	+	+	+	-	-	14
11) HSV D65+WP	-	+	-	-	=	-	=	+	+	=	=	-	=	=	+	+	+	+	+	=	+	+	+	+	-	-	12
12) HSV D65+SoG	-	+	-	-	+	=	+	+	+	+	+	=	=	+	+	+	+	+	+	+	+	+	+	+	-	-	18
13) HSV D65+GE1	-	+	-	-	+	-	+	+	+	=	=	=	=	=	+	+	+	+	+	=	+	+	+	+	-	-	14
14) HSV D65+GE2	-	+	-	-	+	-	+	+	+	=	=	-	=	=	+	+	+	+	+	=	+	+	+	+	-	-	14
15) RGB D65+GW	-	-	-	-	-	-	-	+	+	-	-	-	-	-	=	+	=	+	+	-	+	-	+	+	-	-	8
16) RGB D65+WP	-	-	-	-	-	-	-	=	+	-	-	-	-	-	-	=	-	-	=	-	+	-	=	=	-	-	2
17) RGB D65+SoG	-	-	-	-	-	-	-	+	+	-	-	-	-	-	=	+	=	+	+	-	+	-	+	+	-	-	8
18) RGB D65+GE1	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	+	-	=	=	-	+	-	=	=	-	-	4
19) RGB D65+GE2	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	=	-	=	=	-	+	-	=	=	-	-	3
20) YCBCR D65+GW	-	+	-	-	=	-	=	+	+	=	=	-	=	=	+	+	+	+	+	=	+	+	+	+	-	-	12
21) YCBCR D65+WP	-	-	-	-	-	-	-	-	=	-	-	-	-	-	-	-	-	-	-	-	=	-	-	-	-	-	0
22) YCBCR D65+SoG	-	-	-	-	-	-	-	+	+	-	-	-	-	-	+	+	+	+	+	-	+	=	+	+	-	-	10
23) YCBCR D65+GE1	-	-	-	-	-	-	-	=	+	-	-	-	-	-	-	=	-	=	=	-	+	-	=	=	-	-	2
24) YCBCR D65+GE2	-	-	-	-	-	-	-	=	+	-	-	-	-	-	-	=	-	=	=	-	+	-	=	=	-	-	2
25) ASG	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	=	-	22
26) CGM	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	=	24

Table 4. The Wilcoxon test on the Recall for all the 26 skin detection strategies considered. Each algorithm score is reported in the last column.

Strategy	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Score
1) HSI precision-oriented	=	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	19
2) YCBCR recall-oriented	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	25
3) HSV1 trade-off	+	-	=	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	=	20
4) Bayesian	+	-	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	24
5) Gaussian Mixture Model	+	-	+	-	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	=	21
6) Histogram	-	-	-	-	-	=	+	+	+	+	+	+	+	+	=	+	=	+	+	+	+	+	+	+	-	-	16
7) HSV D65+DN	-	-	-	-	-	-	=	-	-	+	=	=	=	=	-	-	-	-	-	-	-	-	-	-	-	-	1
8) RGB D65+DN	-	-	-	-	-	-	+	=	+	+	+	+	+	+	-	=	-	-	-	+	+	=	+	+	-	-	11
9) YCBCR D65+DN	-	-	-	-	-	-	+	-	=	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	6
10) HSV D65+GW	-	-	-	-	-	-	-	-	-	=	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
11) HSV D65+WP	-	-	-	-	-	-	=	-	-	+	=	=	=	=	-	-	-	-	-	-	-	-	-	-	-	-	1
12) HSV D65+SoG	-	-	-	-	-	-	=	-	-	+	=	=	=	=	-	-	-	-	-	-	-	-	-	-	-	-	1
13) HSV D65+GE1	-	-	-	-	-	-	=	-	-	+	=	=	=	=	-	-	-	-	-	-	-	-	-	-	-	-	1
14) HSV D65+GE2	-	-	-	-	-	-	=	-	-	+	=	=	=	=	-	-	-	-	-	-	-	-	-	-	-	-	1
15) RGB D65+GW	-	-	-	-	-	=	+	+	+	+	+	+	+	+	=	+	=	+	+	+	+	+	+	+	-	-	16
16) RGB D65+WP	-	-	-	-	-	-	+	=	+	+	+	+	+	+	-	=	-	-	-	+	+	=	+	+	-	-	11
17) RGB D65+SoG	-	-	-	-	-	=	+	+	+	+	+	+	+	+	=	+	=	+	+	+	+	+	+	+	-	-	16
18) RGB D65+GE1	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	+	-	=	=	+	+	+	+	+	-	-	14
19) RGB D65+GE2	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	+	-	=	=	+	+	+	+	+	-	-	14
20) YCBCR D65+GW	-	-	-	-	-	-	+	-	+	+	+	+	+	+	-	-	-	-	-	=	+	-	=	=	-	-	8
21) YCBCR D65+WP	-	-	-	-	-	-	+	-	+	+	+	+	+	+	-	-	-	-	-	-	=	-	-	=	-	-	7
22) YCBCR D65+SoG	-	-	-	-	-	-	+	=	+	+	+	+	+	+	-	=	-	-	-	+	+	=	+	+	-	-	11
23) YCBCR D65+GE1	-	-	-	-	-	-	+	-	+	+	+	+	+	+	-	-	-	-	-	=	+	-	=	=	-	-	8
24) YCBCR D65+GE2	-	-	-	-	-	-	+	-	+	+	+	+	+	+	-	-	-	-	-	=	=	-	=	=	-	-	7
25) ASG	+	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	=	+	23
26) CGM	+	-	=	-	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	=	20

Table 5. The Wilcoxon test on the Accuracy for all the 26 skin detection strategies considered. Each algorithm score is reported in the last column.

Strategy	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Score
1) HSI precision-oriented	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	=	-	23
2) YCBCR recall-oriented	-	=	-	-	-	-	-	+	+	-	-	-	-	-	+	+	+	+	+	-	+	-	=	=	-	-	8
3) HSV1 trade-off	-	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	=	-	22
4) Bayesian	-	+	-	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	21
5) Gaussian Mixture Model	-	+	-	-	=	-	-	+	+	-	-	-	-	-	+	+	+	+	+	-	+	=	+	+	-	-	11
6) Histogram	-	+	-	-	+	=	=	+	+	=	=	=	=	=	+	+	+	+	+	=	+	+	+	+	-	-	13
7) HSV D65+DN	-	+	-	-	+	=	=	+	+	-	=	=	=	=	+	+	+	+	+	=	+	+	+	+	-	-	13
8) RGB D65+DN	-	-	-	-	-	-	-	=	-	-	-	-	-	-	-	=	-	=	=	-	-	-	-	-	-	-	0
9) YCBCR D65+DN	-	-	-	-	-	-	-	+	=	-	-	-	-	-	=	+	=	=	=	-	=	-	-	-	-	-	2
10) HSV D65+GW	-	+	-	-	+	=	+	+	+	=	+	=	=	=	+	+	+	+	+	=	+	+	+	+	-	-	15
11) HSV D65+WP	-	+	-	-	+	=	=	+	+	-	=	=	=	=	+	+	+	+	+	=	+	+	+	+	-	-	13
12) HSV D65+SoG	-	+	-	-	+	=	=	+	+	=	=	=	=	=	+	+	+	+	+	=	+	+	+	+	-	-	13
13) HSV D65+GE1	-	+	-	-	+	=	=	+	+	=	=	=	=	=	+	+	+	+	+	=	+	+	+	+	-	-	13
14) HSV D65+GE2	-	+	-	-	+	=	=	+	+	=	=	=	=	=	+	+	+	+	+	=	+	+	+	+	-	-	13
15) RGB D65+GW	-	-	-	-	-	-	-	+	=	-	-	-	-	-	=	+	=	+	+	-	=	-	-	-	-	-	4
16) RGB D65+WP	-	-	-	-	-	-	-	=	-	-	-	-	-	-	-	=	-	=	=	-	-	-	-	-	-	-	0
17) RGB D65+SoG	-	-	-	-	-	-	-	+	=	-	-	-	-	-	=	+	=	=	=	-	=	-	-	-	-	-	2
18) RGB D65+GE1	-	-	-	-	-	-	-	=	=	-	-	-	-	-	-	=	=	=	=	-	=	-	-	-	-	-	0
19) RGB D65+GE2	-	-	-	-	-	-	-	=	=	-	-	-	-	-	-	=	=	=	=	-	-	-	-	-	-	-	0
20) YCBCR D65+GW	-	+	-	-	+	=	=	+	+	=	=	=	=	=	+	+	+	+	+	=	+	+	+	+	-	-	13
21) YCBCR D65+WP	-	-	-	-	-	-	-	+	=	-	-	-	-	-	=	+	=	=	+	-	=	-	-	-	-	-	3
22) YCBCR D65+SoG	-	+	-	-	=	-	-	+	+	-	-	-	-	-	+	+	+	+	+	-	+	=	+	+	-	-	11
23) YCBCR D65+GE1	-	=	-	-	-	-	-	+	+	-	-	-	-	-	+	+	+	+	+	-	+	-	=	=	-	-	8
24) YCBCR D65+GE2	-	=	-	-	-	-	-	+	+	-	-	-	-	-	+	+	+	+	+	-	+	-	=	=	-	-	8
25) ASG	=	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	=	-	22
26) CGM	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	=	25

Table 6. The Wilcoxon test on the F_1 measure for all the 26 skin detection strategies considered. Each algorithm score is reported in the last column.

Strategy	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Score
1) HSI precision-oriented	=	+	=	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	21
2) YCBCR recall-oriented	-	=	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	20
3) HSV1 trade-off	=	+	=	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	21
4) Bayesian	+	+	=	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	22
5) Gaussian Mixture Model	-	-	-	-	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	19
6) Histogram	-	-	-	-	-	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	18
7) HSV D65+DN	-	-	-	-	-	-	=	-	=	+	=	=	=	=	-	-	-	-	-	-	=	-	-	-	-	-	1
8) RGB D65+DN	-	-	-	-	-	-	+	=	+	+	+	+	+	+	-	=	-	-	-	-	+	-	=	+	-	-	9
9) YCBCR D65+DN	-	-	-	-	-	-	=	-	=	+	=	=	=	=	-	-	-	-	-	-	-	-	-	-	-	-	1
10) HSV D65+GW	-	-	-	-	-	-	-	-	-	=	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
11) HSV D65+WP	-	-	-	-	-	-	=	-	=	+	=	=	=	=	-	-	-	-	-	-	=	-	-	-	-	-	1
12) HSV D65+SoG	-	-	-	-	-	-	=	-	=	+	=	=	=	=	-	-	-	-	-	-	=	-	-	-	-	-	1
13) HSV D65+GE1	-	-	-	-	-	-	=	-	=	+	=	=	=	=	-	-	-	-	-	-	=	-	-	-	-	-	1
14) HSV D65+GE2	-	-	-	-	-	-	=	-	=	+	=	=	=	=	-	-	-	-	-	-	=	-	-	-	-	-	1
15) RGB D65+GW	-	-	-	-	-	-	+	+	+	+	+	+	+	+	=	+	=	+	+	=	+	=	+	+	-	-	14
16) RGB D65+WP	-	-	-	-	-	-	+	=	+	+	+	+	+	+	-	=	-	-	=	-	+	-	+	+	-	-	10
17) RGB D65+SoG	-	-	-	-	-	-	+	+	+	+	+	+	+	+	=	+	=	+	+	=	+	=	+	+	-	-	14
18) RGB D65+GE1	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	+	-	=	=	-	+	=	+	+	-	-	12
19) RGB D65+GE2	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	=	-	=	=	-	+	-	+	+	-	-	11
20) YCBCR D65+GW	-	-	-	-	-	-	+	+	+	+	+	+	+	+	=	+	=	+	+	=	+	=	+	+	-	-	14
21) YCBCR D65+WP	-	-	-	-	-	-	=	-	+	+	=	=	=	=	-	-	-	-	-	-	=	-	-	-	-	-	2
22) YCBCR D65+SoG	-	-	-	-	-	-	+	+	+	+	+	+	+	+	=	+	=	=	+	=	+	=	+	+	-	-	13
23) YCBCR D65+GE1	-	-	-	-	-	-	+	=	+	+	+	+	+	+	-	-	-	-	-	-	+	-	=	=	-	-	8
24) YCBCR D65+GE2	-	-	-	-	-	-	+	-	+	+	+	+	+	+	-	-	-	-	-	-	+	-	=	=	-	-	8
25) ASG	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	=	-	24
26) CGM	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	=	25

Table 7. Scores for the Wilcoxon test on the precision, recall, accuracy and F_1 measure, together with their average and median values

Strategy	Precision	Recall	Accuracy	F_1measure	Average	Median
1) HSI precision-oriented	24	19	23	21	21.75	22
2) YCBCR recall-oriented	11	25	8	20	16	15.5
3) HSV1 trade-off	23	20	22	21	21.5	21.5
4) Bayesian	21	24	21	22	22	21.5
5) Gaussian Mixture Model	11	21	11	19	15.5	15
6) Histogram	19	16	13	18	16.5	17
7) HSV D65+DN	11	1	13	1	6.5	6
8) RGB D65+DN	2	11	0	9	5.5	5.5
9) YCBCR D65+DN	0	6	2	1	2.25	1.5
10) HSV D65+GW	14	0	15	0	7.25	7
11) HSV D65+WP	12	1	13	1	6.75	6.5
12) HSV D65+SoG	18	1	13	1	8.25	7
13) HSV D65+GE1	14	1	13	1	7.25	7
14) HSV D65+GE2	14	1	13	1	7.25	7
15) RGB D65+GW	8	16	4	14	10.5	11
16) RGB D65+WP	2	11	0	10	5.75	6
17) RGB D65+SoG	8	16	2	14	10	11
18) RGB D65+GE1	4	14	0	12	7.5	8
19) RGB D65+GE2	3	14	0	11	7	7
20) YCBCR D65+GW	12	8	13	14	11.75	12.5
21) YCBCR D65+WP	0	7	3	2	3	2.5
22) YCBCR D65+SoG	10	11	11	13	11.25	11
23) YCBCR D65+GE1	2	8	8	8	6.5	8
24) YCBCR D65+GE2	2	7	8	8	6.25	7.5
25) ASG	22	23	22	24	22.75	22.5
26) CGM	24	20	25	25	23.5	24.5

4 Conclusions

In this paper we have compared 26 computational strategies for skin detection. They differ in the type of data used in the training phase (i.e. real images vs measured skin reflectances), the type of pre-processing done on the query image, and the type of information used (low-level vs high-level features). In particular, we have defined a high-level computational strategy, which uses a face detector in the pre-processing step. Within this strategy, we have proposed two different implementations: the former exploits an adaptive skin detector, the latter exploits a fixed threshold skin cluster detector on an illuminant-independent image representation.

The experimental results on a dataset containing uncontrolled images show that the high-level strategies outperform low-level ones. In particular the Wilcoxon test on the F_1 measure ranks them as the best two strategies among the ones considered. The average and median Wilcoxon scores for all the metrics considered, further confirms the effectiveness of the high level approach.

References

1. Terrillon, J.C., Shirazit, M., Fukamachi, H., Akamatsu, S.: Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In: Proc. 4th Int. Conf. Automatic Face and Gesture Recognition, pp. 54–61 (2000)
2. Caetano, T., Olabarriaga, S.D., Barone, D.A.C.: Performance evaluation of single and multiple-Gaussian models for skin color modelling. In: Proc. Brazilian Symp. Computer Graphics and Image Processing, pp. 275–282 (2002)
3. Gasparini, F., Corchs, S., Schettini, R.: Recall or precision oriented strategies for binary classification of skin pixels. Journal of Electronic Imaging 17(2), 1–15 (2008)
4. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. of the Conference on Computer Vision and Pattern Recognition, pp. 511–518 (2001)
5. `http://lear.inrialpes.fr/people/vandeweijer/code/ColorConstancy.zip`
6. `http://www.cie.co.at/cie/`
7. Skin Detection, `http://clickdamage.com/sourcecode/index.html`
8. Vezhnevets, V., Sazonov, V., Andreeva, A.: A Survey on Pixel-Based Skin Color Detection Techniques. In: Proc. Graphicon-2003, Moscow, Russia, pp. 85–92 (September 2003)
9. Yang, M.H., Ahuja, N.: Gaussian Mixture Model for Human Skin Colour and its Applications in Image and Video Databases. In: SPIE/EI&T Storage and Retrieval for Image and Video Databases, pp. 458–466 (1999)
10. Tsekeridou, S., Pitas, I.: Facial feature extraction in frontal views using biometric analogies. In: Proc. of the IX European Signal Processing Conference, vol. I, pp. 315–318 (1998)
11. Chai, D., Ngan, K.N.: Face segmentation using skin color map in videophone applications. IEEE Transactions on Circuits and Systems for Video Technology 9(4), 551–564 (1999)
12. Chai, D., Bouzerdoum, A.: A Bayesian approach to skin color classification in YCbCr color space. In: Proceedings TENCON 2000, vol. 2, pp. 421–424 (2000)
13. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. International Journal of Computer Vision 46(1), 81–96 (2002)
14. Kovac, J., Peer, P., Solina, F.: 2D versus 3D colour space face detection. In: Proc. 4th EURASIP Conf. Video Image Processing and Multimedia Communications, pp. 449–454 (2003)
15. Hsu, R., Abdel Mottaleb, M., Jain, A.K.: Face detection in colour images. IEEE Trans. Pattern Analysis and Machine Intelligence 24, 696–706 (2002)
16. Garcia, C., Tziritas, G.: Face detection using quantized skin colour regions merging and wavelet packet analysis. IEEE Trans. Multimedia 1, 264–277 (1999)
17. Gomez, G., Morales, E.F.: Automatic feature construction and a simple rule induction algorithm for skin detection. In: Proc. of the ICML workshop on Machine Learning in Computer Vision, pp. 31–38 (2002)
18. Hsieh, I.S., Fan, K.C., Lin, C.: A statistic approach to the detection of human faces in colour nature scene. Pattern Recognition 35, 1583–1596 (2002)
19. Witkin, A.P.: Scale-space filtering: a new approach to multi-scale description. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 150–153 (1984)
20. Bianco, S., Schettini, R.: Color Constancy Using Faces. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 65–72 (2012)

21. Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics 1, 80–83 (1945)
22. Van de Weijer, J., Gevers, T., Gijsenij, A.: Edge-based color constancy. IEEE Transactions on Image Processing 16(9), 2207–2214 (2007)
23. Gijsenij, A., Gevers, T., Van de Weijer, J.: Computational color constancy: survey and experiments. IEEE Transactions on Image Processing 20(9), 2475–2489 (2011)
24. Zhu, Q., Cheng, K.T., Wu, C.T., Wu, Y.L.: Adaptive learning of an accurate skin-color model. In: Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 37–42 (2004)
25. Brand, J., Mason, J.: A comparative assessment of three approaches to pixel-level human skin detection. In: Proc. IEEE Int. Conf. Pattern Recognition, vol. 1, pp. 1056–1059 (2000)
26. Brand, J., Mason, J., Roach, M., Pawlewski, M.: Enhancing face detection in colour images using a skin probability map. In: Proc. Int.Conf. Intelligent Multimedia, Video and Speech Processing, pp. 344–347 (2001)
27. Zarit, B., Super, B.J., Quek, F.K.H.: Comparison of five color models in skin pixel classification. In: Proc. Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, pp. 58–63 (1999)
28. Sigal, L., Sclaroff, S., Athitsos, V.: Skin color-based video segmentation under time-varying illumination. IEEE Trans. Pattern Analysis and Machine Intelligence 26(7), 862–877 (2004)
29. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
30. ISO. Graphic technology - standard object colour spectra database for colour reproduction evaluation (socs). Technical Report ISO/TR 16066:2003(E) (2003)