Adaptive Skin Classification Using Face and Body Detection

Simone Bianco, *Member, IEEE*, Francesca Gasparini, *Member, IEEE*, and Raimondo Schettini, *Member, IEEE*

***Abstract*— In this paper, we propose a skin classification method exploiting faces and bodies automatically detected in the image, to adaptively initialize individual *ad hoc* skin classifiers. Each classifier is initialized by a face and body couple or by a single face, if no reliable body is detected. Thus, the proposed method builds an *ad hoc* skin classifier for each person in the image, resulting in a classifier less dependent from changes in skin color due to tan levels, races, genders, and illumination conditions. The experimental results on a heterogeneous data set of labeled images show that our proposal outperforms the state-of-the-art methods, and that this improvement is statistically significant.**

***Index Terms*— Skin classification, face detection, body detection.**

I. 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 [1]. Vezhnevets [2] identified three types of skin modeling on which skin detection methods are mainly based: parametric, nonparametric, and explicit skin cluster definition models. The hypothesis underlying these methods is that skin pixels exhibit similar color coordinates in an appropriately chosen color space, and that lighting conditions do not vary too much across the images in the training and test datasets. 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 [3]–[10].

Parametric models [11]–[13] 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 require very little storage space. However their performance depends strongly on the shape of skin color distribution of the training images in the selected color space.

In non-parametric skin modeling methods the skin color distribution is estimated directly on the basis of the training data, without deriving an explicit model of the skin color [14]–[16]. The result of these methods is sometimes referred to as a Skin Probability Map (SPM) [17], [18]. Non-parametric methods can be quickly trained and theoretically does not make any assumption on the shape of the skin color distribution.

All the methods considered when applied in real applications, may degrade their performance due to changes in camera settings, illumination, people tans, makeup, ethnic groups, etc. with respect to the training images. To solve the problem of different imaging conditions, a color constancy approach can be applied as a pre-processing step [7], [19]–[21]. As an alternative, dynamic adaptation techniques can be used, in which the existing skin color models are transformed to cope with changes in illumination conditions [16], [22]–[24]. Kakumanu et al. [1] presented a review of skin classification approaches based on color constancy and dynamic adaptation techniques. Khan et al. [25] analyze the effect of color constancy algorithms on several color based skin classifiers.

Adaptive approaches exploiting face detection have been proposed to cope not only with illuminant and environmental conditions but also with differences among acquired subjects. These methods are based on the assumption that at least one reliable face is present in the image and has been reliably detected. They differentiate among each other mainly in the way they select skin pixels from the detected face(s) to be used to train ad-hoc skin classifiers [26]–[28]. Bianco et al. [20] showed that skin classifiers initialized by reliable skin pixels extracted from faces outperform traditional methods, even when they are preceded by a color constancy pre-processing step.

In this paper we present an adaptive skin classification method where individual skin classifiers are initialized by a face and body couple or by a single face, if no reliable body is detected. The main contributions of this work are:

– The use of both face and body detection to provide more reliable initialization for the ad-hoc individual skin classifier with respect to that initialized using face detection alone. Different strategies for selecting skin pixels from detected bodies to be used as training sets are investigated.

Manuscript received August 13, 2014; revised November 22, 2014 and March 12, 2015; accepted August 5, 2015. Date of publication August 11, 2015; date of current version September 18, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Scott T. Acton.

The authors are with the Department of Informatics, Systems and Communication, University of Milano–Bicocca, Milan 20126, Italy (e-mail: simone.bianco@disco.unimib.it; gasparini@disco.unimib.it; schettini@disco.unimib.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2015.2467209

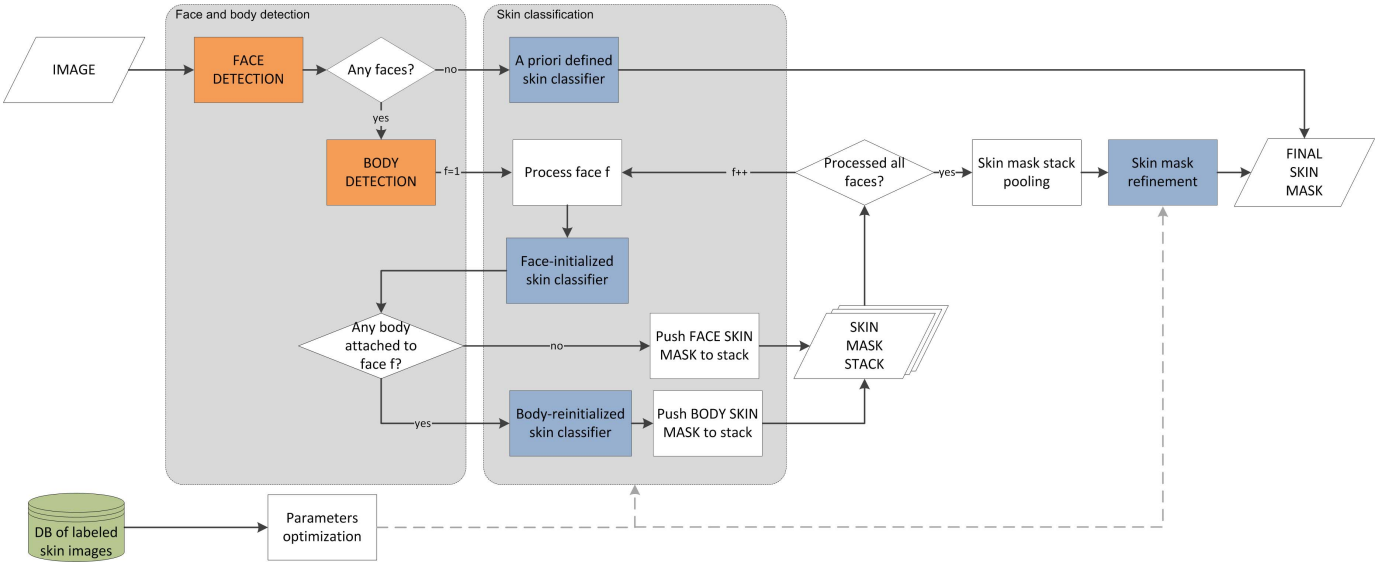


Fig. 1. Operation flowchart of our automatic adaptive skin classification method.

– The design of an adaptive computational method that does not make any assumption about the presence of reliable faces/bodies in the image. Given an input image the proposed method adaptively chooses between an a priori defined skin classifier and ad-hoc skin classifiers based on face and body detection.

An extensive comparison of the proposed method with respect to the state-of-the-art on a heterogeneous dataset containing images acquired under uncontrolled lighting conditions has been carried out. The statistical significance of the improvements obtained by our proposal are assessed using a non-parametric statistical test.

II. The Proposed Approach

The proposed skin classification method builds an ad-hoc skin classifier for each person automatically detected in the image. It exploits faces to initialize the individual ad-hoc skin classifiers, that are then reinitialized if related bodies are detected. The output of the individual classifiers are then combined to obtain the final skin mask. If the face detector does not find any face, an a-priori defined skin classifier in the state of the art is used. The flowchart of the proposed method is shown in Figure 1. There are two main blocks: the former concerning the detection of faces and bodies, the latter devoted to the skin classification. The output of the individual skin classifiers are pooled and refined to produce the final skin mask. All the parameters that are used in the processing blocks and that do not vary on the basis of the actual face and body detected, are found by optimization on a labeled dataset of training images. The list of symbols and functions used by the proposed method is reported in Table I.

TABLE I
List of Symbols and Functions

Symbol	Description
F	Number of detected faces
f	Current face index
I	Input image
M_f	Skin classification mask obtained from face f
B_f	Body detection mask attached to face f
S	Skin mask stack
$g(\cdot\|\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Single Gaussian model
$\boldsymbol{\mu}$	Mean vector
$\boldsymbol{\Sigma}$	Covariance matrix
$p(\cdot\|g)$	Probability wrt g
P	Max-pooled skin mask stack
R	Refined skin probability map

A. Face-Initialized Skin Classifier

A face detector [29] is run on the input image I. If no faces are detected, an a-priori defined skin classifier is used. Otherwise a loop on all the detected faces $f = \{1, \ldots, F\}$ is started. Given the current face f, all its pixels $\mathbf{x}$ are converted into the HSV color space and their luminance is normalized so that $\overline{V(\mathbf{x})} = 0.5$. To select the reliable skin pixels, an explicit skin cluster classifier is used. It filters out pixels that are too dark or too bright, and thus potentially clipped, if they satisfy the condition $V(\mathbf{x}) > t_1 \vee V(\mathbf{x}) < t_2$. Any pixel not belonging to the feasible saturation and hue region of skin colors, i.e. satisfying $S(\mathbf{x}) > t_3 \vee S(\mathbf{x}) < t_4$ or $t_5 < H(\mathbf{x}) < t_6$ is also filtered out.

For each face detected in the image the color distribution of the reliable skin pixels is modeled with a single Gaussian $g([H(\mathbf{x})\ S(\mathbf{x})]|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in the HS plane of the HSV color space, where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix. This is an adaptive skin classifier which builds a different model for each face, that we call Adaptive Single Gaussian (ASG). Each model is applied independently to the whole image I by computing the probability $p([H(\mathbf{x})\ S(\mathbf{x})]|g)$ $\forall x \in I$ and generating a binary mask M_f such that $M_f(\mathbf{x}) = 1$ if $p([H(\mathbf{x})\ S(\mathbf{x})]|g) > t_7$. The pseudo-code for the ASG classifier is reported in Algorithm 1. The optimal thresholds $[t_1, \ldots, t_7]$ for the ASG classifiers are found from the training images. These thresholds are fixed for all the detected faces. If the body detector does not find any body attached to

Algorithm 1 Pseudo-Code of the Adaptive Single Gaussian (ASG) Classifier

Result: The binarized skin mask M_f
Convert face f RGB values to HSV
Luminance normalize V s.t. $\overline{V} = 0.5$
Initialize empty list $s = \{\}$
Initialize skin mask $M_f = zeros(size(I))$
for $k = 1 : nFacePixels$ **do**
  $r = 1$
  if $V(\mathbf{x}_k) > t_1 \vee V(\mathbf{x}_k) < t_2$ **then**
    $r = 0$
  if $S(\mathbf{x}_k) > t_3 \vee S(\mathbf{x}_k) < t_4$ **then**
    $r = 0$
  if $t_5 < H(\mathbf{x}_k) < t_6$ **then**
    $r = 0$
  if $r == 1$ **then**
    Add $\mathbf{x}_k$ to s
if $\neg isempty(s)$ **then**
  Extract HS values for pixels listed in s
  Estimate Gaussian model $g([H(\mathbf{x}_s)\ S(\mathbf{x}_s)]|\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  Convert image RGB values to HSV
  for $k = 1 : nImagePixels$ **do**
    Compute probability $p([H(\mathbf{x}_k)\ S(\mathbf{x}_k)]|g)$
    if $p([H(\mathbf{x}_k)\ S(\mathbf{x}_k)]|g) > t_7$ **then**
      $M_f(\mathbf{x}_k) = 1$

the current face, the obtained skin mask M_f is pushed to the skin mask stack S, i.e. $S^{(f)} = M_f$, otherwise it is used for the re-initialization of the skin classifier as described below.

B. Body-Reinitialized Skin Classifier

Given the skin mask M_f generated from the current face f, and the body detection mask B_f associated to it, the ASG skin classifier $g([H(\mathbf{x}_b)\ S(\mathbf{x}_b)]|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is re-initialized using pixels $\mathbf{x}_b = \{\mathbf{x} \in I : M_f(\mathbf{x}) = 1 \wedge B_f(\mathbf{x}) = 1\}$.

In this work two different body detectors have been used to generate the body detection masks B_f. The former [30] outputs a stickman representation of the detected body, while the latter [31] outputs a contour for the detected body. However, the first detector could give a similar output to that of the second one, since it is based on a prior soft-labeling of pixels to body parts or background. Viceversa the output of the second detector could be transformed into a stickman representation by mapping labeled parts to detected contour.

The stickman representation needs to be converted into a mask to be used in our framework. To this end, for each body part type $b \in$ {head, torso, arm, forearm, thigh, lower leg}, the corresponding mask is obtained by dilation with a rectangular structuring element. The width and height of this element are proportional to the length l_b of the detected body part b: $[w, h] = [w_b l_b, h_b l_b]$. Two example images are reported in Figure 2 with the stickman detection overlaid and the mask obtained from it after dilation.

The second body detector used [31] gives as output a list of poselets for which the corresponding classifiers fired together with their confidence. We apply a threshold t_p to retain only the most confident detections. The final detected body mask is generated by summing all the retained detections, normalizing

Fig. 2. Two example images with the stickman detection overlaid and the mask B_f obtained from it after dilation.

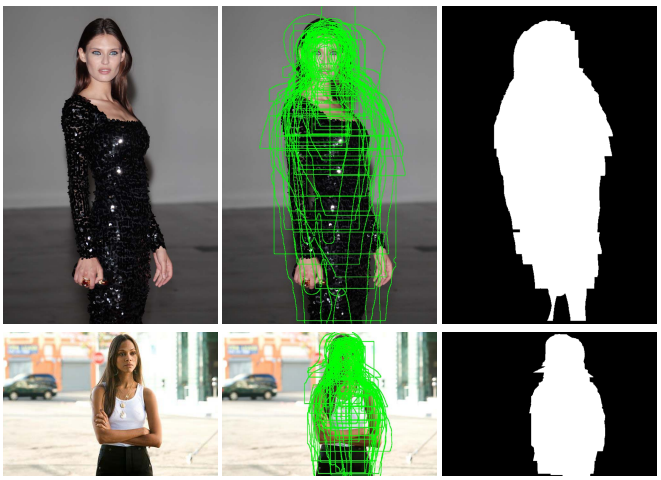


Fig. 3. The images of Figure 2 with the poselet detection overlaid and the mask B_f obtained from it after thresholding.

it by its maximum value, and binarizing it with a threshold t_b. The same original images of Figure 2 are reported in Figure 3 with the poselet detection overlaid and the mask obtained from it after thresholding.

Whatever is the body detector used, the skin mask obtained applying the re-initialized ASG classifier is pushed to the skin mask stack S.

C. Skin Mask Pooling and Refinement

When the loop on all the faces is complete, the final skin mask P for image I is obtained by max-pooling skin mask stack S:

$$P(\mathbf{x}) = \max_{f=1\ldots F} S^{(f)}(\mathbf{x}). \quad (1)$$

The final step of our proposed approach is the refinement of the max-pooled skin mask P using the cross-bilateral filter [32], [33]. The filtering expands the detected skin regions adding neighbor pixels in P which are not separated by strong edges. For each pixel $\mathbf{x}_p \in P$ the cross-bilateral filter output

is computed as:

$$P(\mathbf{x}_p) = \frac{1}{k(\mathbf{x}_p)} \sum_{\mathbf{x}_{p'} \in \Omega} g_d(\mathbf{x}_{p'} - \mathbf{x}_p) g_r(I(\mathbf{x}_p) - I(\mathbf{x}_{p'})) P(\mathbf{x}_{p'}) \tag{2}$$

where $k(\mathbf{x}_p)$ is a normalization term:

$$k(\mathbf{x}_p) = \sum_{\mathbf{x}_{p'} \in \Omega} g_d(\mathbf{x}_{p'} - \mathbf{x}_p) g_r(I(\mathbf{x}_p) - I(\mathbf{x}_{p'})) \tag{3}$$

The function $g_d(\cdot)$ sets the weight in the spatial domain based on the distance between the pixels, while the edge-stopping function $g_r(\cdot)$ sets the weight on the range based on intensity difference. Typically, both functions are Gaussians with widths controlled by the standard deviation parameters σ_d and σ_r respectively.

The difference with respect to standard bilateral filter [34] is that the edge-stopping function g_r is computed on a different image from the one that has actually to be filtered, i.e. $g_r(I(\mathbf{x}_p) - I(\mathbf{x}_{p'}))$ instead of $g_r(P(\mathbf{x}_p) - P(\mathbf{x}_{p'}))$.

We apply cross-bilateral filter to each RGB color channel separately. The outputs $P^{(k)}$, $k \in \{R, G, B\}$ of the three different cross-bilateral filters are then summed and normalized by its maximum value, i.e.:

$$R = \frac{\sum_{k \in \{R,G,B\}} P^{(k)}(\mathbf{x})}{\max_{\mathbf{x}} \sum_{k \in \{R,G,B\}} P^{(k)}(\mathbf{x})} \tag{4}$$

R can be seen as a skin probability map. To obtain the final skin classification mask, this map is then binarized using the threshold t_A and isolated detections are discarded by removing all connected components with area smaller than $t_B I_w I_h$.

The cross-bilateral filter parameters σ_r, σ_d and the thresholds t_A, t_B are found by optimization on the training images.

III. Experimental Setup

All the experimental results here reported were obtained using as training set the Compaq dataset [14], and as test set the Test Database for Skin Detection (TDSD) [35]. TSDS has been chosen as test set since containing more uncorrelated images than those available in video datasets [16], [36], and more full-body images than ECU [37] where most of the images are head-and-shoulder shots. TSDS contains a total of 554 images where skin pixels have been manually labeled. Each image contains at least one person. Several ethnic groups are considered in the dataset, and they can vary both intra- and inter-image. There are no restrictions on both face orientation and body pose. Moreover, the people in group shots may partially occlude each other. The images have been acquired under various lighting conditions in terms of both illuminant color and intensity. These conditions, that are assumed to be unknown, vary both across images and within a single image. Some examples of images belonging to the TDSD are reported in Figure 4.

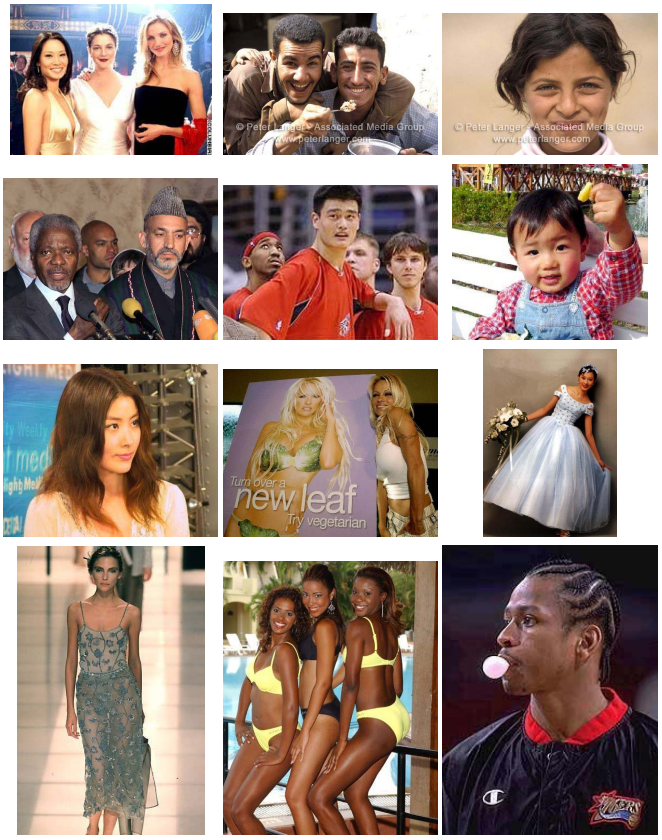


Fig. 4. Examples of images within the TDSD dataset.

A. Evaluation Procedure

To quantify the performance of our adaptive skin classification method and compare the results with those obtained by other methods in the state of the art, the following statistics are used:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{7}$$

by assigning pixel-level classification results as true positive (TP), false positive (FP), false negative (FN), and true negative (TN). To summarize the performance of each method, we used F_1-measure, which is defined as the harmonic mean of precision and recall.

To assess if the difference in performance among the different algorithms considered are statistically significant, we have used the paired Wilcoxon Signed-Rank Test (WST) [38]. This statistical test permits the comparison of the whole distributions of the performance measure. Given two algorithms, the WST is run on the corresponding Precision, Recall, and Accuracy distributions on the whole dataset. For each of the three different performance measures considered, a score is computed. This score counts the number of methods with respect to which the corresponding method has been considered significantly better.

B. Benchmarking Algorithms

To benchmark our method we have considered both pixel-based and face-based skin classifiers available in the literature. All the methods have been implemented by the authors.

Following Vezhnevets et al. [2] we group under the name of pixel-based skin classifiers parametric, nonparametric, and explicit skin cluster definition methods. As pixel-based skin classifiers we here consider:

- A parametric skin classification method based on a Gaussian mixture model in the RGB color space [11].
- A non-parametric skin classification method introduced by Chai and Bouzerdoum [39]. It 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.
- An explicit skin cluster definition method originally introduced by Tsekeridou and Pitias [6]. It works in the HSV color space defining top and bottom boundaries of the color skin cluster for each channel. In this work we use the boundaries redefined in [10] which resulted in the highest F_1-measure.

As face-based skin classifiers we have here considered four approaches: three of them are adaptive in the sense they build a skin color model for each detected person; the fourth one exploits faces to build an illuminant-invariant skin color model. The face-based classifiers considered are:

- A dynamic skin color classifier presented by Wimmer and Radig [26].
- An adaptive face-based classifier presented by Liao and Chi [27].
- An enhanced face-based adaptive skin color model presented by Hsieh et al. [28].
- A skin classifier based on a Color Gamut Mapping [20], hereafter called CGM. Similarly to [40] and [41], 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.

C. Investigated Instances of the Proposed Method

Four different instances of the proposed method are compared:

BSR: implements the strategy described in Section II using the skeleton representation [30] for the output of the detected bodies. It relies on faces and bodies automatically detected in the image, to adaptively initialize individual ad-hoc skin classifiers. Each classifier is initialized by a face and body couple or by a single face, if no reliable bodies are detected. If no faces are detected, the strategy uses the HSV F_1-measure pixel-based method.

BPR: differs from BSR by the body detector used: instead of the skeleton representation, it uses the poselet representation [31] for the output of the detected bodies.

BS: differs from BSR by the edge-stopping function g_r used in the cross-bilateral filter (equation 2): the bounding-boxes of the detected faces are converted into masks which are max-pooled with the body detection masks to give the joint face and body detection mask J. The edge-stopping function g_r used is then $g_r(J(\mathbf{x}_p) - J(\mathbf{x}_{p\prime}))$.

BP: is the same of BS but differs in the body detector used: instead of the skeleton representation, it uses the poselet representation [31] for the output of the detected bodies.

D. Parameters Optimization

Both the proposed method and the benchmarking solutions have been trained on an subset of 250 images taken from an independent dataset of annotated images collected by Jones and Rehg [14]. All the parameters of the proposed method, the face-based methods and pixel-based methods have been set to maximize the F_1-measure. Parameters are found by optimization using Particle Swarm Optimization (PSO) [42]. Given a skin classification method with a set of parameters $t_1, \ldots, t_N$ to be optimized, each possible solution is seen as a point $p \in \mathbb{R}^N$. The skin classifier with parameters p is then run on the whole training set, and the fitness function f computes the median F_1-measure. Since PSO is a population-based stochastic optimization algorithm, its first step consists in a random initialization of the particle position p_i and velocity v_i for each particle $i = 1, \ldots, N_p$. The fitness function f is then evaluated for each particle position p_i to obtain $fp_i = f(p_i)$. The best known position of each particle pb_i and the best known position $p^\star$ of the entire swarm are then initialized. After the initialization, the iterative process is started and repeated until the maximum number of iterations N_I has been reached. For each iteration j, particle positions are updated as

$$p_i^{(j)} = p_i^{(j-1)} + v_i^{(j)} \tag{8}$$

with

$$\begin{aligned} v_i^{(j)} = w^{(j)} v_i^{(j-1)} &+ c_1 U_1^{(j)} \left(pb_i^{(j-1)} - p_i^{(j-1)} \right) \\ &+ c_2 U_2^{(j)} \left(p^\star - p_i^{(j-1)} \right) \end{aligned} \tag{9}$$

where $[w^{(j)}, c_1, c_2]$ are weights that respectively control the importance of the inertia, the personal best influence, and the global best influence terms; $U_1^{(j)}$ and $U_2^{(j)}$ are two random numbers. The fitness function f is then evaluated for each particle position $p_i^{(j)}$ to obtain $fp_i^{(j)} = f(p_i^{(j)})$ and personal best positions pb_i are updated if $fp_i^{(j)} > fp_i^{(j-1)}$. Global best position is also updated if $\exists i$ such that $fp_i^{(j)} > p^\star$. In this work PSO is run with standard settings, i.e.: $N_p = 24$, $N_I = 100$, $w^{(1)} = 0.9$ with linear decay to $w^{(N_I)} = 0.4$, $c_1 = c_2 = 2$, and $U_1^{(j)} = U_2^{(j)} \sim \mathcal{U}(0, 1)$.

IV. Experimental Results

In this section we compare the performance of the different strategies withing the proposed adaptive method with those of the benchmarking methods on the TDSD dataset. We report in Table II their performance in terms of median precision, recall, and accuracy. The results are grouped with respect to

TABLE II

PERFORMANCE OF PIXEL-BASED, FACE-BASED, AND BOTH FACE- AND BODY-BASED SKIN CLASSIFIERS IN TERMS OF MEDIAN PRECISION, RECALL, AND ACCURACY. THE WST SCORES COMPUTED INDIVIDUALLY FOR EACH MEASURE ARE ALSO REPORTED

Skin Classifier Type	Method Name	Precision	WST Score	Recall	WST Score	Accuracy	WST Score
Pixel-based	HSV F_1-measure [10]	0.7171	4	0.8237	4	0.8805	1
	GMM	0.6984	2	0.7933	2	0.8775	1
	BAY [39]	0.7033	2	0.8643	7	0.8791	1
Face-based	Wimmer [26]	0.7883	6	0.7483	1	0.8868	5
	Liao [27]	0.5484	0	0.8927	9	0.8195	0
	Hsieh [28]	0.8515	10	0.7276	0	0.9095	6
	CGM [20]	0.7906	6	0.8325	5	0.9076	6
	ASG [20]	0.6605	1	0.9365	11	0.8841	1
Face- and body-based	BS [this work]	0.8401	8	0.8332	5	0.9225	9
	BP [this work]	0.8661	11	0.7899	2	0.9213	9
	BSR [this work]	0.7771	5	0.9298	10	0.9176	6
	BPR [this work]	0.8439	8	0.8643	7	0.9254	11

the type of skin classifier used: pixel-based, face-based, and both face- and body-based. For all the classifiers exploiting faces, the same face detector is adopted [29]. For a fair comparison, for all the face-based methods, when no faces are detected, the HSV F_1-measure pixel-based method [10] is used as it is the same pixel-based method used in our adaptive strategies. In Table II the WST scores are also reported: they are computed individually for the precision, recall, and accuracy measures. The values of precision, recall, and accuracy measures are color coded on the basis of the WST score: the more saturated the color, the higher the WST score. It can be noticed that face-based and face- and body-based classifiers obtain the highest WST scores. Furthermore, concerning accuracy, face- and body-based classifiers clearly outperform both face-based and pixel-based classifiers.

In Figure 5 we report in the precision-recall plane how the performance of the best pixel-based skin classifier (BAY, black circle) improves by using algorithms exploiting high-level cues: firstly adding face information (CGM, black square), and then adding body information. The two different body detectors used are respectively plotted in different colors: the red and blue triangles represent BPR and BSR, the red and blue stars BP and BS respectively. On the same plot iso-F_1 curves are also reported. From the plot it is possible to see that the addition of face information is able to increase the F_1-measure by 3.6% (CGM) with respect to the best pixel-based method (BAY). Adding body information always improves the F_1-measure for all the proposed strategies. In particular BSR improves F_1-measure by 7.2% with respect to BAY, while BPR by 7.9%. Using body information as done in BS and BP (red and blue stars) results in strategies more precision-oriented, while BSR and BPR (red and blue triangles) result in strategies more recall-oriented. The body detector used in BS and BSR (stickman: red star and red triangle) results in classifiers more recall-oriented, while the one used in BP and BPR (poselet: blue star and blue triangle) results in classifiers more precision-oriented.

Two example images taken from the TDSD dataset on which the proposed strategies reach the highest and lowest F_1-measure values are respectively reported in Figure 6 and 7.

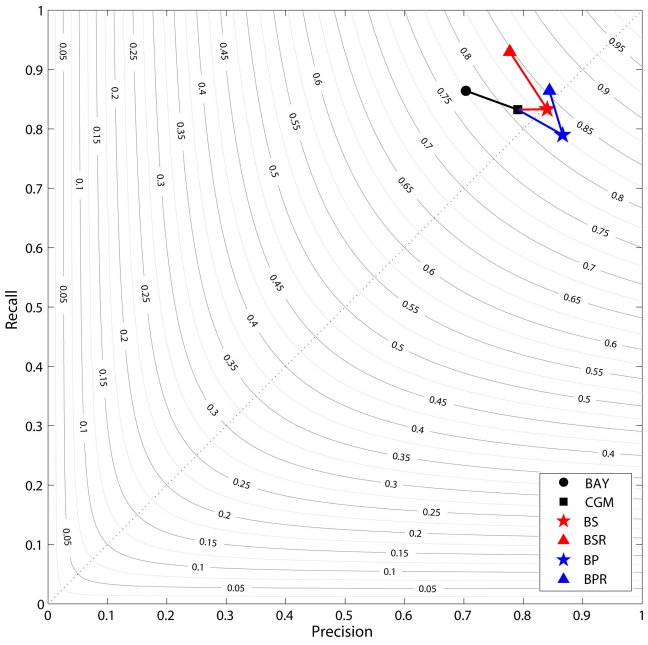


Fig. 5. Performance in terms of F_1-measure of the following skin classifiers: the best pixel-based skin classifier (BAY black circle); the best face-based skin classifier (CGM black square); and our four body-based skin strategies (BP blue star, BPR blue triangle, BS red star and BSR red triangle).

For each example we report the original image, the ground truth skin mask, the skin probability map R (see equation 4) for the proposed strategies BSR and BPR, and the corresponding final skin masks obtained by thresholding R as described in section II-C with the thresholds found by optimization on the training set.

For the images in Figure 6 it is possible to see that both the BSR and BPR strategies produce very good skin classification results although some false positive and false negative regions are present.

Concerning the worst results, a deeper analysis is needed. In order to support the analysis of the results, the face and body detections for both images of Figure 7 are reported in Figure 8.

The low performance for both BSR and BPR in the first image is caused by a low classification precision. In fact,

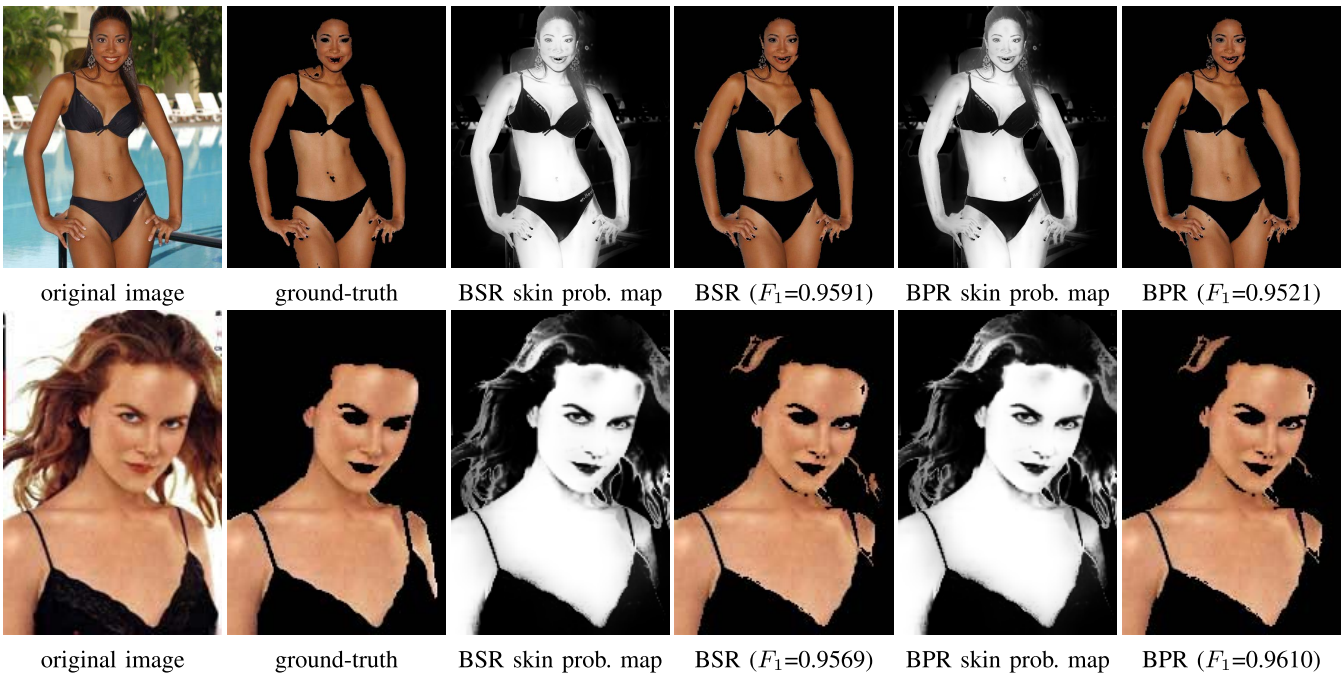


Fig. 6. Images with the highest F_1-measure for the BSR (first row), and BPR (second row).

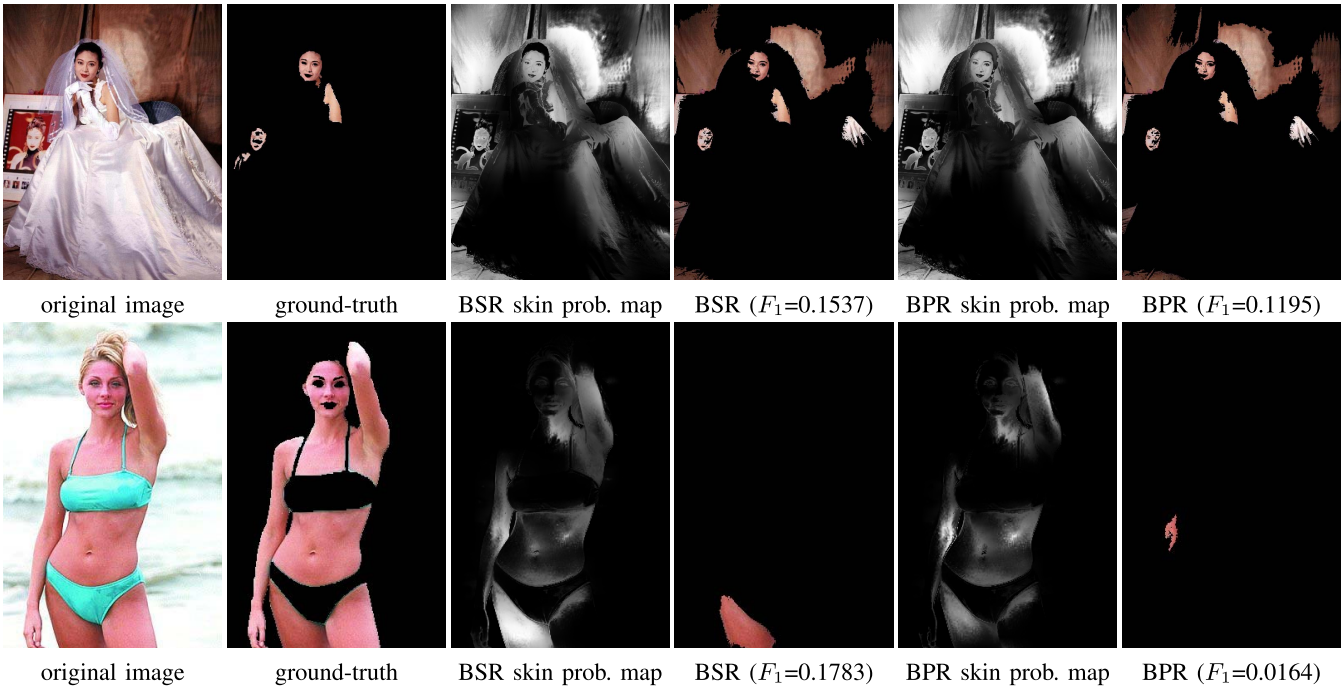


Fig. 7. Images with the lowest F_1-measure for the BSR (first row), and BPR (second row).

there are background pixels that are too similar to the skin tone of the detected face (see the corresponding skin probability maps reported in Fig. 7). For images where the background color is similar to the skin tone, the precision of the classifier can be increased by using BS or BP strategies, as can be seen in Figure 9 where the output of BS is reported.

The low performance in the second image is due to a low classification recall caused by the reddish tone of the skin. The most part of the face pixels are judged to not belong to the feasible hue region of skin colors. This can be seen in the corresponding skin probability maps reported in Figure 7, where probability on the right-side of the face is almost zero and on the left-side is very low. For this particular image, the exclusion of the constraint on the feasible hue region of skin colors (found by the optimization procedure on the training set) would generate much better results, as can be seen in Figure 10.

At first, such a constraint could seem a limitation. However, it has been designed to discard false positive face detections with unfeasible color that should not be used to initialize

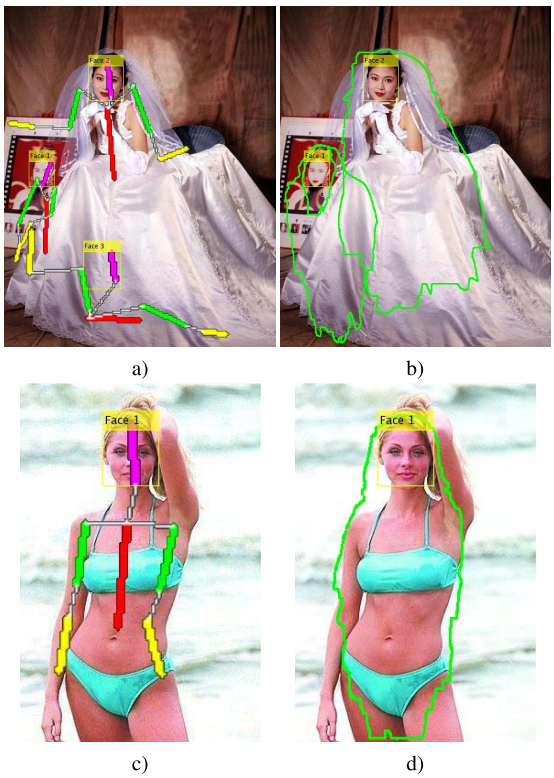


Fig. 8. Faces and bodies detections for the two images on which the proposed strategies reach the lowest F_1-measure values. Face and body detections used by BSR (a and c), and BPR (b and d).

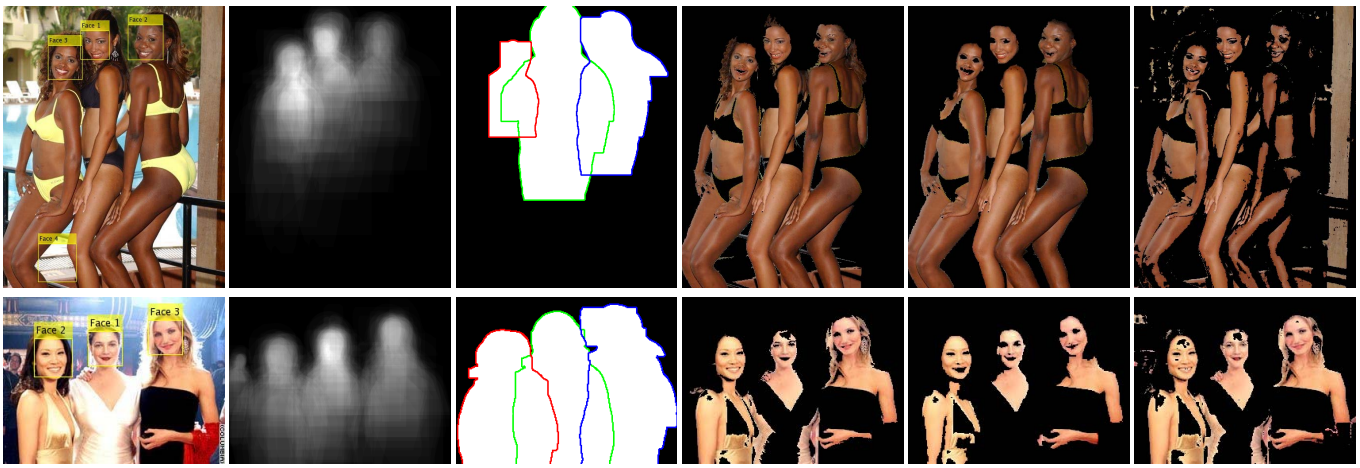
Fig. 9. Output of the BS skin classifier applied to the image of the first row of Figure 7.

Fig. 10. Output of the BSR skin classifier applied to the image of the second row of Figure 7 excluding the constraint on the feasible hue region of skin colors.

Fig. 11. Original image with detected faces (top left). Skin classification output: Wimmer [26] (top right); Liao [27] (center left); Hsieh [28] (center right); CGM [20] (bottom left); ASG [20] (bottom right).

the ASG. An example of the usefulness of this feature can be seen in Figure 7 (first row), where the skin probability map in the region of the false positive face detection (i.e. Face #3 in Figure 8.a) is almost zero. Another illustrative example is shown in Figure 11, where an input image with girls with painted faces is reported. The top left image contains the bounding boxes of the detected faces overlaid on the original image; the others contain the output of all the face-based skin classifiers considered in this paper. Comparing the outputs of the different face-based skin classifiers it is possible to see that CGM and ASG are the only ones which reach the highest precision by discarding red, white, and blue pixels as they not belong to the feasible hue region of skin colors.

Two additional examples in which more than a person is present are reported in Figure 12. They are relative to the BPR method, whose optimal parameters are reported in Table III. This has been chosen among the four proposals as being the one with the highest F_1-measure. For each example we report: a) the original image with the detected faces overlaid; each face region is used to initialize an ad-hoc individual skin classifier; b) a visualization of the detected bodies using poselets; c) the binarized body masks, where for better visualization color contours are used to identify each different body region. Each of these masks is used to reinitialize the ad-hoc individual skin classifier; d) the result obtained using the proposed BPR method; e) the ground truth; f) the result obtained by HSV F_1-measure method, which is the pixel-based method that is used in our proposals when no faces and bodies are detected. In the first example it is possible to see that a false positive face is detected (i.e. Face #4 in Fig. 12.a). This face is filtered out by the ASG skin classifier



a) b) c) d) e) f)

Fig. 12. Original image with detected faces (a); detected bodies using poselets (b); binary masks used to reinitialize the ad-hoc individual classifiers (for visualization color contours are used to identify the different body regions) (c); skin classification output of the proposed BPR method (d); ground truth (e); output of HSV F_1-measure [10] (f).

TABLE III
OPTIMAL PARAMETERS FOUND BY PSO FOR BPR

Parameter	Description	Value
t_1, t_2	boundaries on V	$0.38, 0.98$
t_3, t_4	boundaries on S	$0.13, 0.82$
t_5, t_6	boundaries on H	$0.03, 0.24$
t_7	face binarization	7.68
t_p	poselet confidence	0.00
t_b	poselet binarization	0.19
σ_r, σ_d	cross-bilateral filtering	$0.06, 0.05 \min[I_w, I_h]$
t_A, t_B	morphological operations	$0.60, 0.003$

and thus it is not used to reinitialize ASG (see Fig. 12.c). In the second example we can notice that four bodies are detected (Fig. 12.b). The left-most one is not a false positive detection, but it is not used to reinitialize ASG (see Fig. 12.c) since no corresponding face was detected (Fig. 12.a). The examples confirm that the performance of pixel-based skin classifiers can be improved by exploiting high-level cues, especially in the presence of skin-like backgrounds.

V. CONCLUSIONS

In this paper we have presented a fully automatic adaptive skin classification method that outperforms existing skin classifiers in case of images with a great variability in terms of illumination conditions, tan levels and races. Our method builds an ad-hoc skin classifier for each person in the image. The proposed method adaptively chooses between pixel-based, face-based, and both face- and body-based skin classifiers, on the basis of the detection results of both face and body detectors. In the experimental results we have shown that the performance of pixel-based skin classifiers improves incrementally by adding firstly face information and then body information. Four different strategies of our proposed method have been evaluated showing that skin classification methods that rely on body information outperform existing methods, whatever the body model adopted (BSR and BS versus BPR and BP) and the way to integrate body information (BSR and BPR versus BS and BP). Different body models and way to integrate body information result in skin classifiers more precision or recall oriented. Our experimental results report the performance of our proposals taking into account all the eventual face and body detector errors and the statistical significance of the improvements.

REFERENCES

[1] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern Recognit.*, vol. 40, no. 3, pp. 1106–1122, Mar. 2007.

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

[3] I.-S. Hsieh, K.-C. Fan, and C. Lin, "A statistic approach to the detection of human faces in color nature scene," *Pattern Recognit.*, vol. 35, no. 7, pp. 1583–1596, Jul. 2002.

[4] D. Chai and K. N. Ngan, "Face segmentation using skin-color map in videophone applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 551–564, Jun. 1999.

[5] J. Kovac, P. Peer, and F. Solina, "2D versus 3D colour space face detection," in *Proc. 4th EURASIP Conf. Focused Video/Image Process. Multimedia Commun.*, vol. 2. Jul. 2003, pp. 449–454.

[6] S. Tsekeridou and I. Pitas, "Facial feature extraction in frontal views using biometric analogies," in *Proc. 9th Eur. Signal Process. Conf.*, 1998, pp. 315–318.

[7] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 696–706, May 2002.

[8] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 264–277, Sep. 1999.

[9] G. Gomez and E. Morales, "Automatic feature construction and a simple rule induction algorithm for skin detection," in *Proc. ICML Workshop Mach. Learn. Comput. Vis.*, 2002, pp. 31–38.

[10] F. Gasparini, S. Corchs, and R. Schettini, "Recall or precision-oriented strategies for binary classification of skin pixels," *J. Electron. Imag.*, vol. 17, no. 2, p. 023017, Apr. 2008.

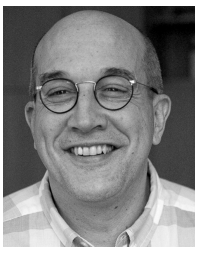
[11] M.-H. Yang and N. Ahuja, "Gaussian mixture model for human skin color and its applications in image and video databases," in *Electronic Imaging*. Bellingham, WA, USA: SPIE, 1999, pp. 458–466.

[12] J.-C. Terrillon, M. N. Shirazi, H. Fukamachi, and S. Akamatsu, "Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images," in *Proc. 4th IEEE Int. Conf. Automat. Face Gesture Recognit.*, Mar. 2000, pp. 54–61.

[13] T. S. Caetano, S. D. Olabarriaga, and D. A. C. Barone, "Performance evaluation of single and multiple-Gaussian models for skin color modeling," in *Proc. 15th Brazilian Symp. Comput. Graph. Image Process.*, 2002, pp. 275–282.
[14] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *Int. J. Comput. Vis.*, vol. 46, no. 1, pp. 81–96, Jan. 2002.
[15] B. D. Zarit, B. J. Super, and F. K. H. Quek, "Comparison of five color models in skin pixel classification," in *Proc. Int. Workshop Recognit., Anal., Tracking Faces Gestures Real-Time Syst.*, Sep. 1999, pp. 58–63.
[16] L. Sigal, S. Sclaroff, and V. Athitsos, "Skin color-based video segmentation under time-varying illumination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 7, pp. 862–877, Jul. 2004.
[17] J. Brand and J. S. Mason, "A comparative assessment of three approaches to pixel-level human skin-detection," in *Proc. 15th Int. Conf. Pattern Recognit.*, vol. 1. Sep. 2000, pp. 1056–1059.
[18] J. Brand, J. S. Mason, M. Roach, and M. Pawlewski, "Enhancing face detection in colour images using a skin probability map," in *Proc. Int. Symp. Intell. Multimedia, Video Speech Process.*, May 2001, pp. 344–347.
[19] F. Gasparini, S. Corchs, and R. Schettini, "A recall or precision oriented skin classifier using binary combining strategies," *Pattern Recognit.*, vol. 38, no. 1, pp. 2204–2207, Nov. 2005.
[20] S. Bianco, F. Gasparini, and R. Schettini, "Computational strategies for skin detection," in *Proc. 4th Int. Workshop Comput. Color Imag.*, LNCS vol. 7786. 2013, pp. 199–211.
[21] F. Solina, P. Peer, B. Batagelj, S. Juvan, and J. Kovač, "Color-based face detection in the '15 seconds of fame' art installation," in *Proc. Mirage, Conf. Comput. Vis./Comput. Graph. Collab. Model-Based Imag., Rendering, Image Anal. Graph. Special Effects*, Rocquencourt, France, Mar. 2003, pp. 38–47.
[22] T. S. Caetano, S. D. Olabarriaga, and D. A. C. Barone, "Do mixture models in chromaticity space improve skin detection?" *Pattern Recognit.*, vol. 36, no. 12, pp. 3019–3021, Dec. 2003.
[23] R. Hassanpour, A. Shahbahrami, and S. Wong, "Adaptive Gaussian mixture model for skin color segmentation," in *Proc. World Acad. Sci., Eng. Technol.*, vol. 31. 2008, pp. 1–6.
[24] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, "Adaptive skin color modeling using the skin locus for selecting training pixels," *Pattern Recognit.*, vol. 36, no. 3, pp. 681–690, Mar. 2003.
[25] R. Khan, A. Hanbury, J. Stöttinger, and A. Bais, "Color based skin classification," *Pattern Recognit. Lett.*, vol. 33, no. 2, pp. 157–163, Jan. 2012.
[26] M. Wimmer and B. Radig, "Adaptive skin color classificator," in *Proc. 1st ICGST Int. Conf. Graph., Vis. Image Process. (GVIP)*, vol. 1. 2005, pp. 324–327.
[27] W.-H. Liao and Y.-H. Chi, "Estimation of skin color range using achromatic features," in *Proc. 8th Int. Conf. Intell. Syst. Design Appl. (ISDA)*, vol. 2. Nov. 2008, pp. 493–497.
[28] C.-C. Hsieh, D.-H. Liou, and W.-R. Lai, "Enhanced face-based adaptive skin color model," *J. Appl. Sci. Eng.*, vol. 15, no. 2, pp. 167–176, 2012.
[29] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. 2001, pp. I-511–I-518.
[30] M. Eichner, M. Marin-Jimenez, A. Zisserman, and V. Ferrari, "2D articulated human pose estimation and retrieval in (almost) unconstrained still images," *Int. J. Comput. Vis.*, vol. 99, no. 2, pp. 190–214, Sep. 2012.
[31] L. Bourdev and J. Malik, "Poselets: Body part detectors trained using 3D human pose annotations," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 1365–1372.
[32] E. Eisemann and F. Durand, "Flash photography enhancement via intrinsic relighting," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 673–678, Aug. 2004.
[33] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 664–672, Aug. 2004.
[34] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 839–846.
[35] Q. Zhu, K.-T. Cheng, C.-T. Wu, and Y.-L. Wu, "Adaptive learning of an accurate skin-color model," in *Proc. 6th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2004, pp. 37–42.
[36] J. Stöttinger, A. Hanbury, C. Liensberger, and R. Khan, "Skin paths for contextual flagging adult videos," in *Proc. Adv. Vis. Comput.*, 2009, pp. 303–314.
[37] S. L. Phung, A. Bouzerdoum, and D. S. Chai, "Skin segmentation using color pixel classification: Analysis and comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 148–154, Jan. 2005.
[38] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, no. 6, pp. 80–83, Dec. 1945.
[39] D. Chai and A. Bouzerdoum, "A Bayesian approach to skin color classification in YCbCr color space," in *Proc. TENCON*, vol. 2. Sep. 2000, pp. 421–424.
[40] S. Bianco and R. Schettini, "Color constancy using faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 65–72.
[41] S. Bianco and R. Schettini, "Adaptive color constancy using faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1505–1518, Aug. 2014.
[42] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, vol. 4. Perth, WA, Australia, Nov./Dec. 1995, pp. 1942–1948.

Simone Bianco (M'12) received the B.Sc. and M.Sc. degrees in mathematics from the University of Milano–Bicocca, Italy, in 2003 and 2006, respectively, and the Ph.D. degree in computer science from the Department of Informatics, Systems and Communication, University of Milano–Bicocca, in 2010, where he is currently a Post-Doctoral Researcher. His research interests include computer vision, optimization algorithms, machine learning, and color imaging.

Francesca Gasparini (M'12) received the degree and the Ph.D. degree in nuclear engineering from the Polytechnic of Milan, Italy, in 1997 and 2000, respectively. Since 2001, she has been a Fellow with the ITC Imaging and Vision Laboratory, Italian National Research Council, Milan, where her research has focused on image enhancement, cast detection, and removal. She is currently an Assistant Professor of Computer Science with the Department of Informatics, Systems and Communication, University of Milano–Bicocca, where she is involved in image processing.



Raimondo Schettini (M'12) is currently a Full Professor with the University of Milano–Bicocca, Italy, and the Head of the Imaging and Vision Laboratory. He has authored over 250-refereed papers and six patents about color reproduction, and image processing, analysis, and classification. He is an IAPR fellow for his contributions to pattern recognition research and color image analysis.