Pattern Recognition 47 (2014) 1573–1585

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Face detection based on skin color likelihood

Yuseok Ban [a], Sang-Ki Kim [b], Sooyeon Kim [a], Kar-Ann Toh [a], Sangyoun Lee [a,*]

[a] School of Electrical & Electronic Engineering, Yonsei University, Seoul 120-749, Republic of Korea
[b] LG electronics Umyeon R&D campus, 16, Umyeon-dong, Seocho-gu, 137-900 Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 10 December 2012
Received in revised form
31 October 2013
Accepted 6 November 2013
Available online 25 November 2013

Keywords:
Face detection
Skin color emphasis
Likelihood
Boosting
Face pose variation
Complex background

ABSTRACT

We propose a face detection method based on skin color likelihood via a boosting algorithm which emphasizes skin color information while deemphasizing non-skin color information. A stochastic model is adapted to compute the similarity between a color region and the skin color. Both Haar-like features and Local Binary Pattern (LBP) features are utilized to build a cascaded classifier. The boosted classifier is implemented based on skin color emphasis to localize the face region from a color image. Based on our experiments, the proposed method shows good tolerance to face pose variation and complex background with significant improvements over classical boosting-based classifiers in terms of total error rate performance.

1. Introduction

Human face detection is among the most important topics in biometric research since it has a broad range of applications. Detection of face is often performed prior to recognition and tracking in biometric and surveillance systems. A variety of techniques have been proposed for face detection in the literature where they can be generally classified into the following categories [1]: knowledge-based methods, invariant feature methods, template matching methods and appearance-based methods.

Knowledge-based methods are rule-based methods which encode human knowledge of what constitutes a typical face. Usually, some rules are designed to capture the relationships among the facial components. Invariant feature methods adopt features such as facial components, texture, skin color and a multiple of these features for face detection. These methods aim to find common structural features which exist among faces under different ambient conditions. Template matching methods store several standard patterns of a face to describe the face either as a whole or as separate facial components. Appearance-based methods learn a model or a group of features from a set of training images to capture the representative variability of facial appearance.

Most of the face detection techniques incur a large number of false rejections due to severe face pose variation and false acceptances due to complex background. To address these issues, we propose a face detection method based on skin color emphasis and iterative boosting to selectively highlight the skin color information and deemphasize background information. Unlike other boosting-based methods using skin color, our method uses neither parametric curve fitting nor morphological operators. Skin color is used for skin color emphasis rather than skin color segmentation.

Our main contributions of this work include the tolerance of proposed system to face rotation and complex background. The boosted classifier reacts less sensitively to face pose variation as it concentrates on probabilistic distribution of facial skin color rather than the details of facial components in gray-level brightness. Also, non-skin color information including background is significantly reduced, so that skin color likelihood can be discriminatively learned.

The organization of this paper is as follows. Section 2 provides a review on related works in face detection using skin color information. Section 3 describes our proposed method in detail. Section 4 presents the experimental results of our method on several face databases. Finally, our conclusion is given in Section 5.

2. Related works

Many face detection methods based on a face model have been proposed to cope with varying conditions including face rotation and complex background. Wang and Yuan [2] proposed a human face detection from color images under complex conditions including arbitrary image background. They used an evolutionary computation technique to cluster skin-like color pixels and segment each face-like region. After the face-like regions are located,

* Corresponding author. Tel.: +82 2 2123 5768; fax: +82 2 313 2879.
E-mail addresses: van@yonsei.ac.kr (Y. Ban), boxboxsk@gmail.com (S.-K. Kim), sykim1221@yonsei.ac.kr (S. Kim), katoh@yonsei.ac.kr (K.-A. Toh), syleee@yonsei.ac.kr (S. Lee).

0031-3203/$ - see front matter
http://dx.doi.org/10.1016/j.patcog.2013.11.005

the wavelet decomposition is applied to each face-like region to detect the possible facial components and to check if there is an eye in the region. Regions in which an eye is detected or the facial components are distributed like a predefined face model are recognized as human faces. Yao and Gao [3] established a type of coordinate transformation which is able to improve chrominance of skin and lips. With the coordinates, they suggested a face detection method based on skin chrominance and lip chrominance transformations to deal with the varying pose of object and the complex background. Hsu et al. [4] presented a face detection algorithm for color images in the presence of varying lighting conditions and complex background. The algorithm is based on their novel lighting compensation technique and a nonlinear transform to the *YCbCr* color space. They first detected skin regions to generate face candidates that are then verified according to eye, mouth and boundary maps. Aldasouqi and Hassan [5] proposed a fast algorithm for detecting faces using morphology-based techniques in *HSV* color space. Sanjay Kr. Singh et al. [6] have combined *RGB*, *YCbCr* and *HSI* color spaces to get a new skin color based face detection algorithm. As presented above, model-based face detection methods commonly use transformation of color space and are based on single or multiple ranges of threshold and morphological operations in order to segment skin regions [7]. The advantage of explicitly defining the boundary of skin cluster is the simple skin detection rules which allow very rapid classification. However, to achieve a high recognition accuracy using this method, we need to find a specifically adequate threshold levels and appropriate decision rules in an empirical way [8].

Many face detection methods based on boosting algorithm have been also suggested. Viola and Jones [9] proposed the boosting-based face detection from learning a sequence of Haar-like features. The differences in average intensities between two rectangular regions are encoded by Haar-like features. The cascade structure of classifiers is built using boosting algorithm which chooses distinctive features [9]. Lienhart et al. [10] extended the work of Viola and Jones using an extended set of Haar features for different views of faces. Despite of the usefulness of Haar-like features, the complete set of the features has to include a mass of redundant information, and the use of pixel brightness shows limitation against varying conditions such as face rotation and complex background. Zhang et al. [11] used AdaBoost learning to select a set of local regions and their weights with respect to Local Binary Pattern (LBP) features for face detection. Many face detection techniques have difficulty in finding face under conditions of large variation in face pose and complex background, and so does AdaBoost using LBP features. Yan-Wen Wu et al. [12] used AdaBoost algorithm combined with skin color segmentation, and the segmentation is obtained by single Gaussian model fitting and morphological operations on binary image. Furthermore, Gaussian mixture models have been suggested for modeling the skin color distribution [13]. Kai-Biao Ge et al. [14] suggested an AdaBoost algorithm combined with skin segmentation and LBP based face description. Although parametric curve fitting such as Gaussian fitting or elliptical fitting enables incomplete training data to be generalized and interpolated, the result highly depends on the shape of the curve [8]. Additionally, either general facial shape information or specific facial component information can be lost via skin color segmentation.

In this paper, we propose a boosting-based face detection method using skin color information without any parametric fitting or morphological operation. Skin color information is used not for skin color segmentation but for skin color emphasis. A cascaded classifier based on AdaBoost is combined with skin color emphasis, resulted in achieving improved face detection

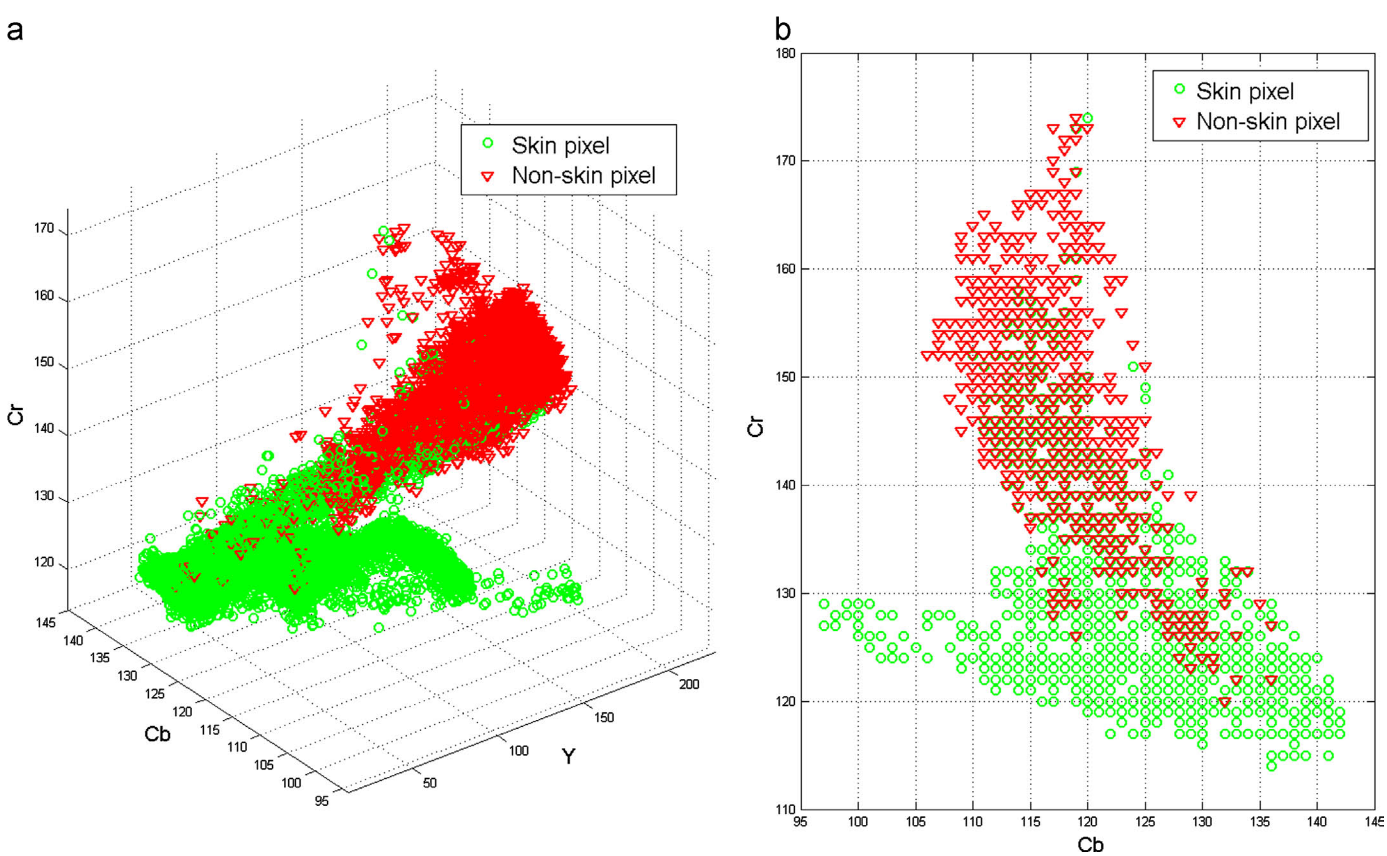

Fig. 1. Examples of skin and non-skin color distributions on (a) *YCbCr* space and (b) *CbCr* space.

performance against complex background and face rotation. AdaBoost is adopted here since it is based on a cascaded structure of sequential classifiers which can effectively identify the non-skin color information such as complex background. The first cascade stage pays less attention to non-skin color, and so do the later successive stages. On the other hand, the skin color is effectively emphasized through these cascaded stages. Once trained, the face can be efficiently detected with fast computational speed.

3. Proposed method

3.1. Skin likelihood

The $YCbCr$ space can be easily obtained from the RGB space by a simple matrix operation. Eq. (1) shows the actual conversion from RGB to $YCbCr$ according to [15]

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.168 & -0.331 & 0.5 \\ 0.5 & -0.418 & -0.081 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{1}$$

The $YCbCr$ space is perceptually uniform, and it separates luminance and chrominance presenting compactness of the skin distribution cluster [4] as shown in Fig. 1a. Human skin forms a relatively tight cluster in color space even when different races are considered [16,17], hence learning the probability of skin color through the chromaticity information of $YCbCr$ space could be helpful. Based on the probability of skin color, we can emphasize the color that belongs to the skin, while ignoring the color that does not. To build a two dimensional histogram of skin colors, $h_{skin}(Cb, Cr)$, a large set of pixels containing skin colors is used as shown in Fig. 1b. All the pixels within an image can be used to define a second histogram of the entire colors, $h_{total}(Cb, Cr)$. The probability that a given color belongs to the skin is obtained by applying Bayes rule to each pixel of an image using the two histograms [18].

We use the following terms to apply the Bayes rule:

$h_{skin}(Cb, Cr)$: Histogram of skin colors in an image
$h_{total}(Cb, Cr)$: Histogram of entire colors in an image
N_{skin}: Sum over Cb and Cr of $h_{skin}(Cb, Cr)$
N_{total}: Sum over Cb and Cr of $h_{total}(Cb, Cr)$

The probability of skin color given a (Cb, Cr) color vector can be approximated as the ratio between normalized histograms of skin color and non-skin color which are based on manually labeled ground truths of skin pixels (see Fig. 2).

The probability of a skin-like pixel is approximated by the fraction of observed skin-like pixels as follows:

$$p(skin) \cong N_{skin}/N_{total} \tag{2}$$

By using Bayes rule, the probability of skin given a (Cb, Cr) color vector is described as

$$p(skin|Cb, Cr) = \frac{p(Cb, Cr|skin) \times p(skin)}{p(Cb, Cr)} \tag{3}$$

Eq. (3) gives a lookup table that directly converts a (Cb, Cr) pixel value into probability of whether it belongs to skin color. Throughout the creation of the lookup table, the two histograms can be quantized into adequate levels. In the lookup table, we place default values of 0 for all pixels for which $h_{total}(Cb, Cr)$ is zero. Finally, the skin likelihood for each pixel $(Cb(i,j), Cr(i,j))$ at position (i,j) can be approximated by

$$P_{skin}(i,j) = p(skin|Cb(i,j), Cr(i,j)) \tag{4}$$

We have found the adequate number of quantization levels of histogram, l^*, by minimizing the following objective function which is the summation of false acceptance rate and false rejection rate in accordance with 2^k number of bins of the two histograms,

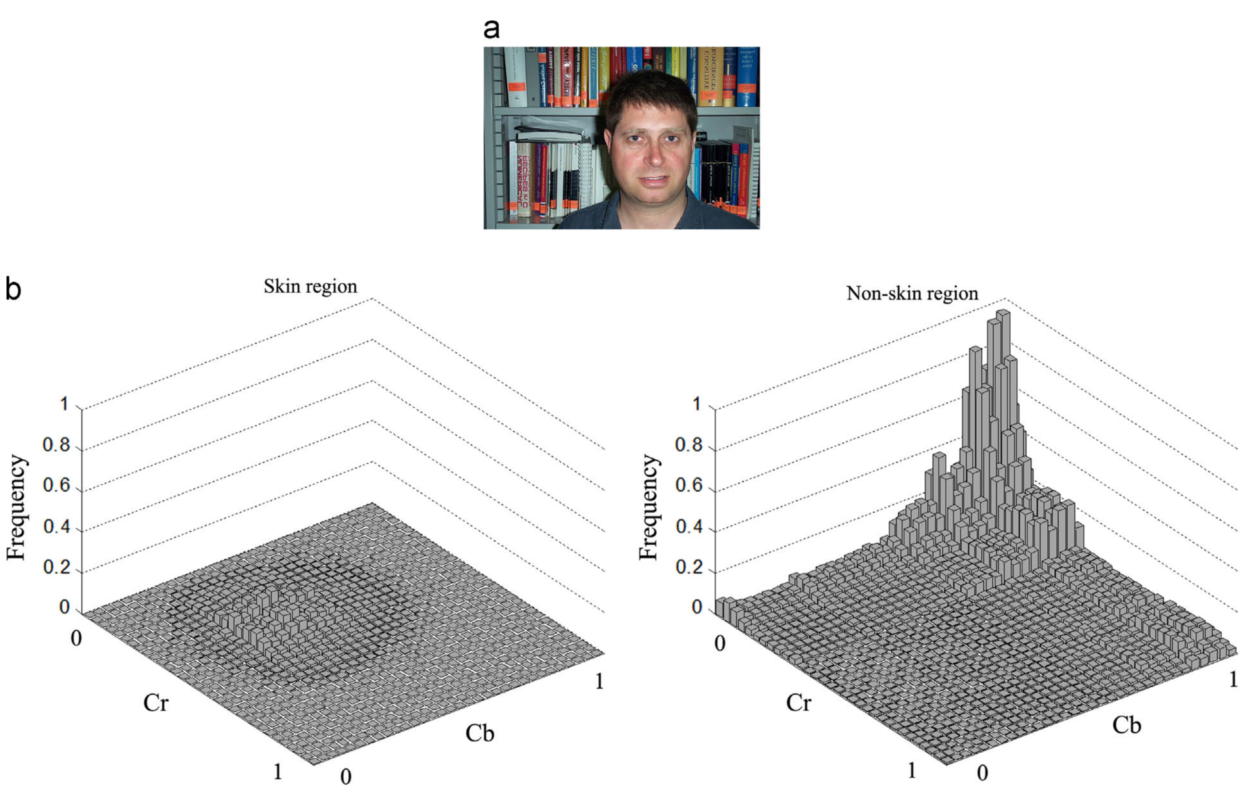

Fig. 2. (a) A face example on containing skin color and (b) its histograms of Cb and Cr spaces of skin region and non-skin region.

$h_{skin}(Cb, Cr)$ and $h_{total}(Cb, Cr)$.

$$l^* = \arg\min_{l \in 2^k} \left\{ \frac{1}{m^-} \sum_{j=1}^{m^-} \delta_{FN,T(l)}(\mathbf{x}_j) + \frac{1}{m^+} \sum_{i=1}^{m^+} \delta_{FP,T(l)}(\mathbf{x}_i) \right\} \quad (5)$$

where k is a positive integer, m^- and m^+ respectively denote the number of negative and positive examples, $\delta_{FN,T(l)}(\mathbf{x}_j)$ and $\delta_{FP,T(l)}(\mathbf{x}_i)$ using l^* number of bins of histogram correspond to '1' whenever the test data $\mathbf{x}_j$ and $\mathbf{x}_i$ are false negative and false positive respectively.

A number of authors have adopted Gaussian models where further image processing methods such as morphological operations are utilized to find the facial features [12,13]. Our method captures multiple types of complexion similar to Gaussian models where multiple types of complexion are considered during skin color modeling. For example, CMU PIE and Pointing'04 contain faces ranging from dark skin tone to light skin tone (see Section 4.1). Using the two histograms, $h_{skin}(Cb, Cr)$ and $h_{total}(Cb, Cr)$, we can prevent loss of information resulted from over-generalization and distortion from interpolation caused by Single Gaussian Model (SGM) [12] or Gaussian Mixture Models (GMM) [13]. In contrast to Gaussian models which used only skin color distribution for modeling, both skin color histogram and non-skin color histogram are used to create a discrete and simple Bayesian lookup table based on the probability of skin color and non-skin color. This leads to faster computation compared to continuous curve fitting methods based on parametric formulations such as Gaussian models (see Table 7).

The formulation for learning and utilizing the Bayesian lookup table are described in Eqs. (6) and (7) respectively. In Eq. (6), $f_1(\cdot)$ accumulates the histograms of skin regions and builds the lookup table.

$$T = f_1(\mathbf{I}, \mathbf{B}), \quad \mathbf{I} = (\mathbf{i}_1 \quad \mathbf{i}_2 \quad \ldots \quad \mathbf{i}_N), \quad \mathbf{B} = (\mathbf{b}_1 \quad \mathbf{b}_2 \quad \ldots \quad \mathbf{b}_N) \quad (6)$$

where T is the lookup table, $\mathbf{I}$ is the image sequence containing the two channels of Cb and Cr, and $\mathbf{B}$ is the binary mask of skin regions.

In Eq. (7), $f_2(\cdot)$ calculates the probability of skin based on the lookup table. Applying this equation to each pixel, we finally obtain a skin likelihood image.

$$y_n^{[p]} = f_2(\mathbf{x}_n^{[p]}, T), \quad 0 \le y_n^{[p]} \le 1, \quad \mathbf{x}_n^{[p]} = \begin{pmatrix} x_{n,Cb}{}^{[p]} \\ x_{n,Cr}{}^{[p]} \end{pmatrix} \quad \forall n = 1, 2, \ldots, N \ \forall p = 1, 2, \ldots, P \quad (7)$$

where $y_n^{[p]}$ is the probability of skin of pth pixel of the nth image, $\mathbf{x}_n^{[p]}$ is the vector component of pth pixel of the nth image including the two channels of Cb and Cr, T is the lookup table, N is the number of images, and P is the number of pixels within an image.

3.2. Skin boosted classifier

The training scheme of skin boosted classifier is described in Table 1. Skin color information is accentuated so that it can be distinctively learned by skin boosted cascade training via the procedure shown in Fig. 3. The procedure is to learn not only the characteristic of facial structure but also the characteristic of skin color distribution of face. Background information can be also effectively minimized from the beginning of the first cascade stage of the iterative boosting [19]. Therefore our skin boosted classifier learning aims both to select complementary weak classifiers and simultaneously to determine the associated weights that skin color information is emphasized while non-skin color information is reduced [9,20].

In Table 1, a set of training examples as $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$ is given for training, and $y_i \in \{+1, -1\}$ is the class label of $x_i \in R^n$. $H_{skin}(x')$ is skin strong classifier, h_m is weak classifier, α_m is weight, and M is the total number of combinational features.

Haar-like features provide a way for detailed analysis of edge and texture of skin. The rectangular Haar-like feature is calculated as the difference of the sums of skin likelihood within white rectangles and black rectangles. Each feature type can express existence or absence of characteristics of the distribution of skin likelihood in the probabilistic image (see Fig. 4). LBP feature is extracted by binarising the gradients of center point to its 8 neighboring points pixel-wise [21] (see Fig. 5a). Subsequently, the histogram of the binary pattern is used as a skin likelihood descriptor. As such, LBP eventually describes the distribution of facial skin which can be seen as a composition of micropatterns such as spot, flat, line end, edge and corner [22]. The face region is scanned by local patches, and the texture feature is extracted from each patch independently with respect to the regional histogram. The features are then concatenated to form a global description of the face having highlighted skin color. The skin texture description of a single patch describes the appearance of the patch and the combination of descriptions of all patches represents the global geometry of the face with highlighted skin in the probabilistic image (see Fig. 5b). The combinations of our method with the rectangular summation of Haar-like features and the histogram computation of LBP are compared in terms of detection accuracy performance and computational time.

Table 1
Training scheme of skin boosted classifier.

1. Apply Bayesian lookup table $x_n^{[p]} = f_2(\mathbf{x}_n^{[p]}, T)\ \forall n = 1, 2, \ldots, N\ \forall p = 1, 2, \ldots, P$
2. Start with initial weight $w_i = 1/N,\ i = 1, 2, \ldots, N,\ H_{skin}(x') = 0$
3. Repeat for $m = 1, 2, \ldots, M$
   (a) Fit the regression function by weighted least squares fitting of Y to X'.
   (b) Update $H_{skin}(x') \leftarrow H_{skin}(x') + \alpha_m h_m(x')$
   (c) Update $w_i \leftarrow w_i e^{-y_i \alpha_m h_m(x_i)}$ and normalization
4. Output the skin strong classifier

$$H_{skin}(x') = \begin{cases} 1 & \text{when } \sum_{m=1}^{M} \alpha_m h_m(x') \ge \frac{1}{2} \sum_{m=1}^{M} \alpha_m \\ 0 & \text{otherwise} \end{cases}$$

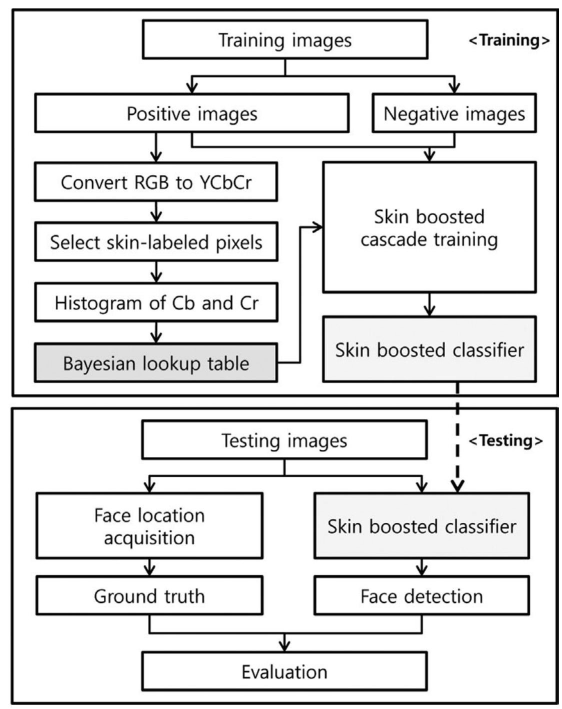

Fig. 3. Systematic flow.

A weak classifier is designed to select the single rectangular feature or the single LBP histogram bin that best separates positive and negative examples [9]. The description of the weak classifier tends to rather concentrate on skin region than to average over all image areas including complex background region. The skin strong classifier, which is a linear combination of weighted weak classifiers selected through the training scheme, eventually focuses on the characteristic of probabilistic distribution of skin color. On the contrary, probabilistic distribution of non-skin color is taken into less consideration.

4. Experiment

4.1. Experimental setup

To evaluate our proposed method, we used color face images under complex background and largely varying pose of face in yaw, roll and pitch directions. In addition, varying lighting condition, different races, facial expression and slight changes in appearance such as moustache or glasses are also basically considered. We used 393 images of Pointing'04 database [23] (Fig. 6) and 1200 images of IMM database [24] (Fig. 7) to test our method against face pose variation in the presence of varying lighting condition. Then, 250 images of Caltech database [25] is used to test the condition of complex background (Fig. 8), while 1020 images of CMU PIE database [26] is used for the test under the combination of complex background and face rotation (Fig. 9) in the presence of different races condition. Bao database [27] includes the combination of multiple faces, complex background, pose variation, different races, facial expression, appearance change such as glasses, moustache and varying lighting conditions over the 221 images of different sizes (Fig. 10). Finally, FDDB [28] includes 2845 images containing a wide range of difficulties (Fig. 11). FDDB database additionally considers face occlusion, out-of-focused face, motion blur and severely varying lighting conditions compared to Bao database (Table 2).

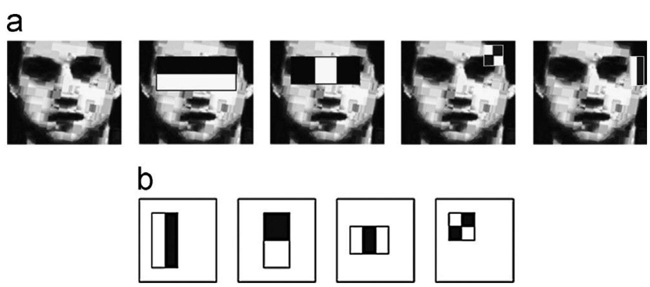

Fig. 4. (a) Haar-like representation and (b) Haar-like feature extraction of skin color likelihood.

Our skin boosted classifier is trained using positive samples and negative samples with a ratio of 1 positive sample to 4 negative samples. AdaBoost is applied for cascading strong classifiers, and Haar-like features and LBP features are used to build weak classifiers. Including Viola et al. [19], many classical algorithms started at the base scale of 24×24 pixels in which faces are detected, so the smallest size of the scanning window is set to 24×24. Viola et al. [19] uses a rescaling factor of 1.25 while Lienhart et al. [10] uses a rescaling factor of 1.2. We set the rescaling factor to 1.2 because window-scanning using a rescaling factor of 1.2 can be more elaborate than that of using a rescaling

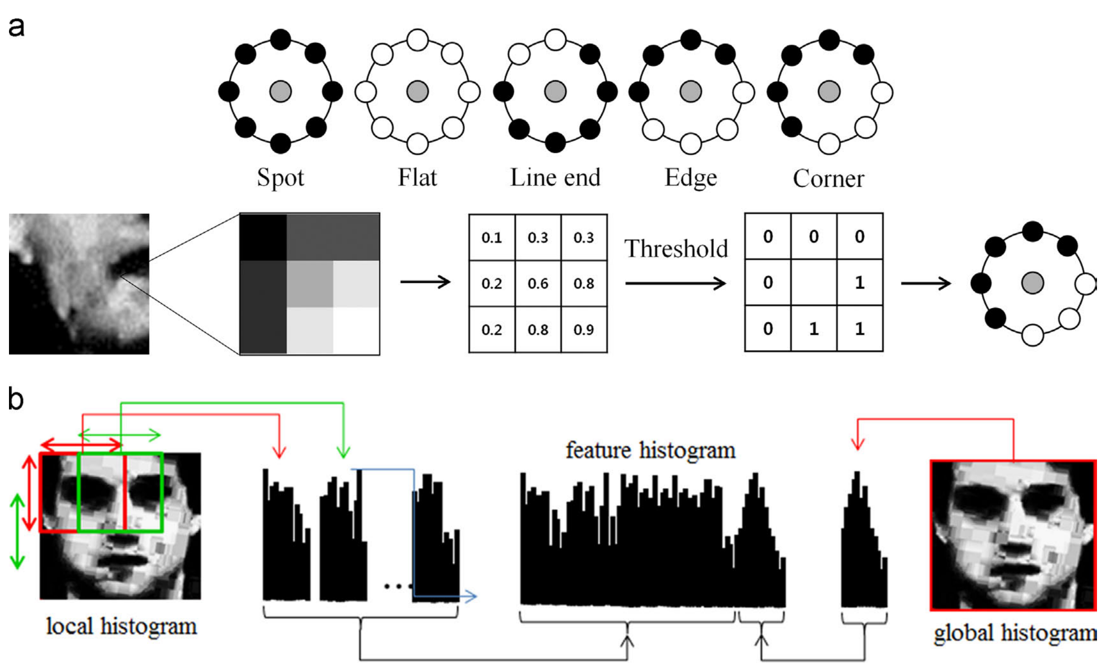

Fig. 5. (a) LBP representation and (b) LBP feature extraction of skin color likelihood.

Fig. 6. Pointing'04 database.

Fig. 7. IMM database.

Fig. 8. Caltech database.

Fig. 9. CMU PIE database.

Fig. 10. Bao database.

Fig. 11. FDDB database.

factor of 1.25. Every error rate is achieved by computing the average of 10 fold cross-validation results for better reliability.

4.2. Performance measure

First, to deal with how good a face detection result is, we applied the overlap measure A [29] as given by

$$A = \frac{|A_g \cap A_c|}{\sqrt{|A_g| \times |A_c|}} \tag{8}$$

where A_g and A_c are the ground truth bounding box and the computed bounding box respectively, and their magnitudes, $|A_g|$ and $|A_c|$, are defined as area of the bounding box in number of pixels.

The measure A represents the ratio of the overlapping area between A_g and A_c to the geometric mean of A_g and A_c. When A_g and A_c completely overlap, A equals 1. To implement this quality measure, the ground truth is extracted manually by the rule depicted in Fig. 12. Requisitely, we measure the margin between nose and mouth, d_{nm}, and the margin of mouth width from the left tip to the right tip, d_{mw}. These margins are scaled by $\alpha$to produce four boundaries. The upper boundary is located above the margin, $d_{nm}\alpha$, from the eyes, the lower boundary is located below the margin, $d_{nm}\alpha$, from mouth, the left boundary is located left the margin, $d_{mw}\alpha$, from eye 1, and the right boundary is located right the margin, $d_{mw}\alpha$, from eye 2.

Second, for quantitative evaluation of face localization result, a confusion matrix is constructed as shown in Fig. 13. If the positive prediction area overlaps the ground truth area by more than a threshold of 95%, it is considered as True Positive (TP), otherwise it is considered as False Positive (FP). Also we take negative testing image with no positive prediction as True Negative (TN) and positive testing image with no positive prediction as False Negative (FN) as presented in Eq. (9). False Acceptance Rate (FAR) is the

Table 2
Composition of database.

Dataset	Condition	Number of images
Caltech	Complex background	250
Pointing'04	Face rotation	393
CMU PIE	Complex background	136
	Small variation in face rotation, complex background	476
	Large variation in face rotation, complex background	408
IMM	Face rotation	1200
Bao	Multiple faces, complex background, face rotation	221
FDDB	Combination of various conditions in the wild	2843

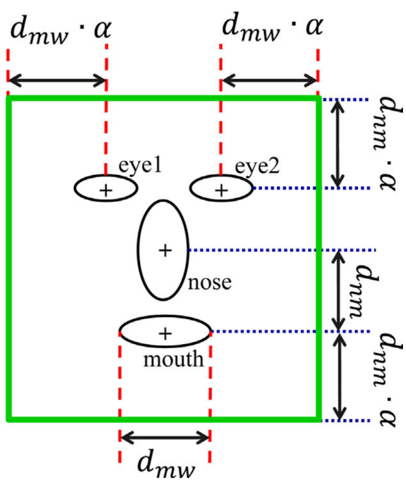

Fig. 12. Margins for ground truth.

		Ground truth	
		P	N
Prediction	$\hat{P}$	TP	FP
	$\hat{N}$	FN	TN

Fig. 13. Confusion matrix.

rate that a negative example is falsely detected or the positive example does not overlap the ground truth area by more than the threshold value. False Rejection Rate (FRR) is the rate that a positive example is falsely rejected. According to [30], these rates are written as

$$FAR = \frac{1}{m^-}\sum_{j=1}^{m^-}\delta_{FN}(\mathbf{x}_j), \quad FRR = \frac{1}{m^+}\sum_{i=1}^{m^+}\delta_{FP}(\mathbf{x}_i) \tag{9}$$

where m^- and m^+ denote the numbers of negative and positive examples respectively, $\delta_{FN}(\mathbf{x}_j)$ corresponds to a '1' whenever the test data $\mathbf{x}_j$ is FN, and $\delta_{FP}(\mathbf{x}_i)$ corresponds to a '1' whenever the test data $\mathbf{x}_i$ is FP. With the FAR and FRR in place, the Half Total Error Rate (HTER) [31] can be written as

$$HTER = \frac{FAR+FRR}{2} = \frac{1}{2m^-}\sum_{j=1}^{m^-}\delta_{FN}(\mathbf{x}_j) + \frac{1}{2m^+}\sum_{i=1}^{m^+}\delta_{FP}(\mathbf{x}_i) \tag{10}$$

4.3. Results

The experimental results are shown in Fig. 14–19. Our proposed method improves the boosting-based conventional methods [9–14] under face pose variation and complex background conditions in respect to half total error rate. We also compare the time complexity of each method by measuring the computational time and the number of training cascade stages. Two main advantages could be achieved using our suggested method. First, the classifier reacts less sensitively to variation in face pose, because it tends to focus on structural distribution of skin color of the presented face rather than the details of facial components in gray-level brightness. Second, small likelihood value is assigned to background region producing tolerance against complex background. According to these advantages, we could efficaciously detect a face in the presence of complex background and face pose variation in yaw, roll and pitch directions. Compared to conventional methods, our proposed method noticeably reduces false rejection rates against face pose variation and reduces false acceptance rates against complex background as presented in Tables 3 and 4 respectively.

4.3.1. Face pose variation

Conventional face detection method using Haar-like features calculates the summations of gray-level brightness in rectangular areas, whereas our suggested method computes the summations of likelihood that a given color belongs to the skin in the areas. Similarly, LBP features of our proposed method label the pixels by using not the gray-level intensities but the skin likelihood of their neighborhood. The rectangular summation of Haar-like feature has been reported to tolerate face rotation beyond a certain range [19]. As presented in Table 3, our proposed method further improves the tolerance against varying face pose.

Pointing'04 test dataset contains various face pose conditions, and our method shows considerably improved performance against severe changes in face pose regarding all the directions of yaw, roll and pitch. Skin color likelihood could be well combined with both the rectangular summation of Haar-like features and the histogram computation of LBP features. The highlighted skin color helps the skin boosted classifier to concentrate on the structural distribution of skin color of face rather than the details of facial components in gray-level brightness. As a result, our proposed method reduces false rejections against severe face pose variation also on other test datasets including IMM, CMU PIE as shown in Table 4.

4.3.2. Complex background

Caltech test dataset includes face in the presence of complex background. Our proposed method substantially reduces the false acceptance rate by focusing on the frequencies of skin color likelihood as shown in Table 3. Skin boosted classifier effectively extracts distinctive features by taking background information into less consideration. CMU PIE test dataset concurrently contains complex background and face rotation conditions. The overall results on CMU PIE test dataset show the effectiveness of our proposed method not only against complex background but also against the combination of complex background and face rotation. Bao test dataset mainly contains multiple faces with a combination of complex background, pose variation, different races and varying lighting conditions. Based on the experimental results, our method effectively deemphasizes complex background even when the proportion of background region increases and multiple faces are presented. FDDB test dataset includes complex background with an extensive combination of various conditions in the wild additionally considering face occlusion, severely varying lighting, motion blur and out-of-focused face conditions additionally compared to the Bao test dataset. Based on skin color emphasis, our

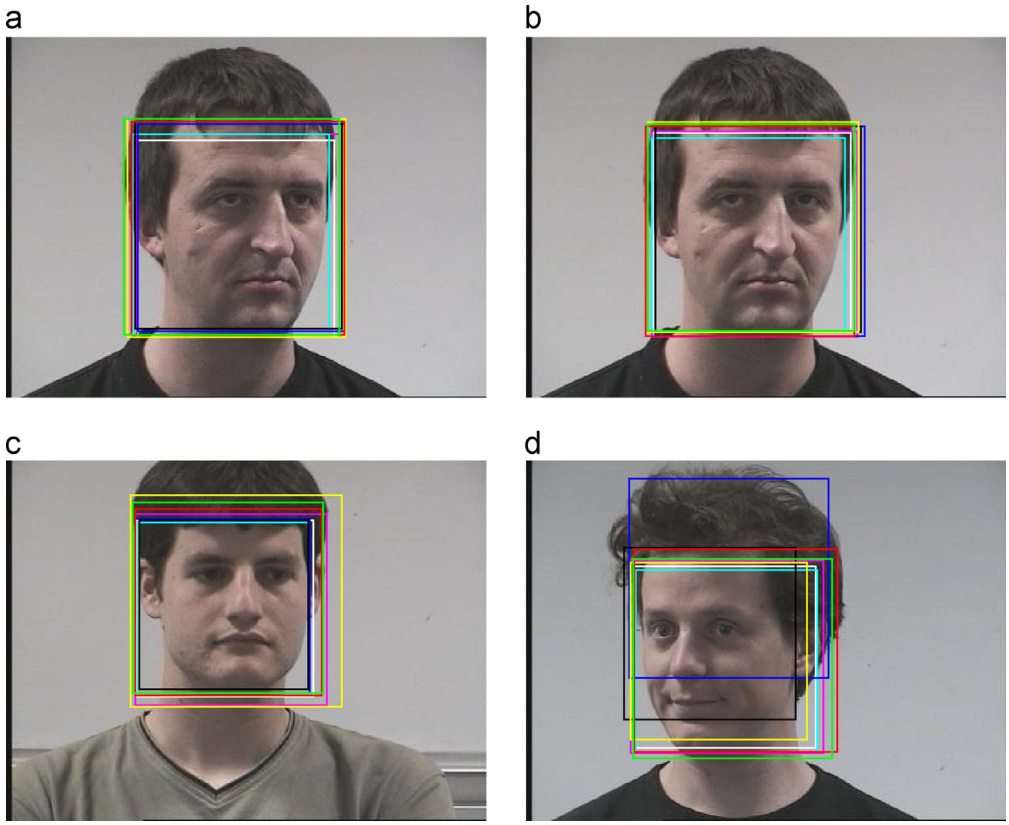

Fig. 14. Detection results of Pointing'04 dataset.

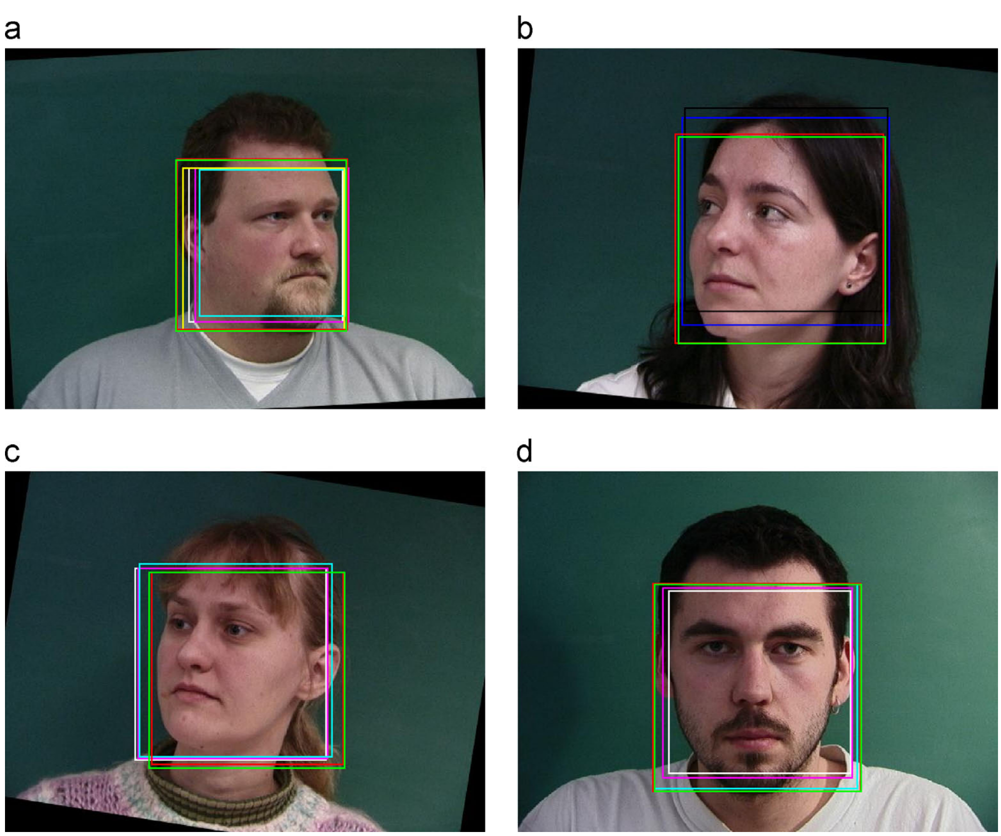

Fig. 15. Detection results of IMM dataset.

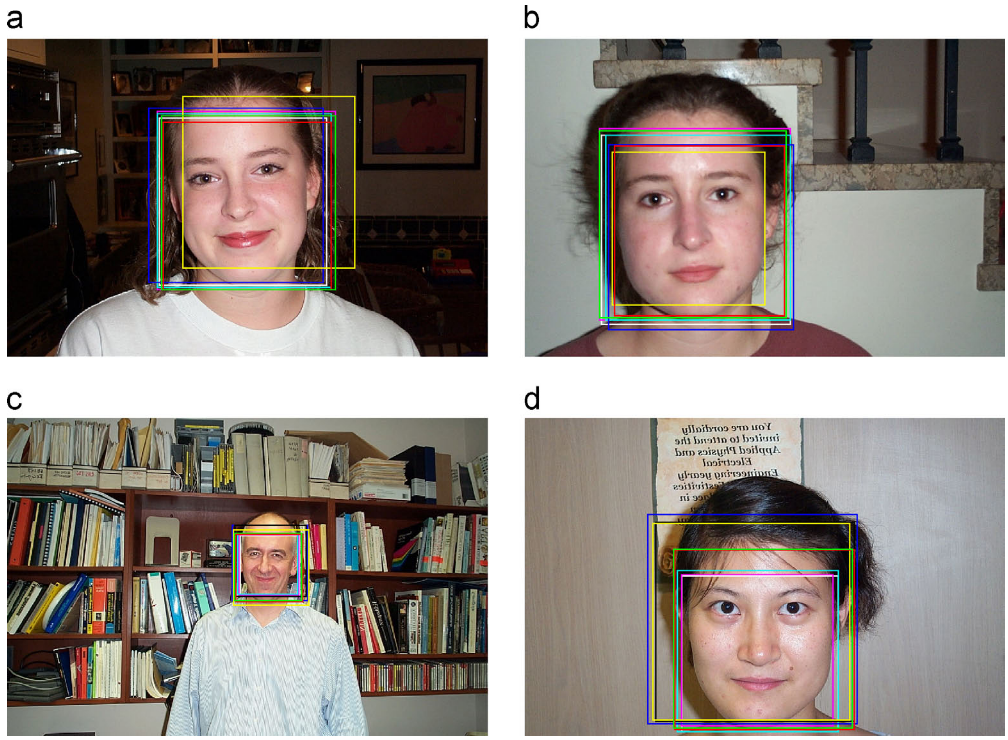

Fig. 16. Detection results of Caltech dataset.

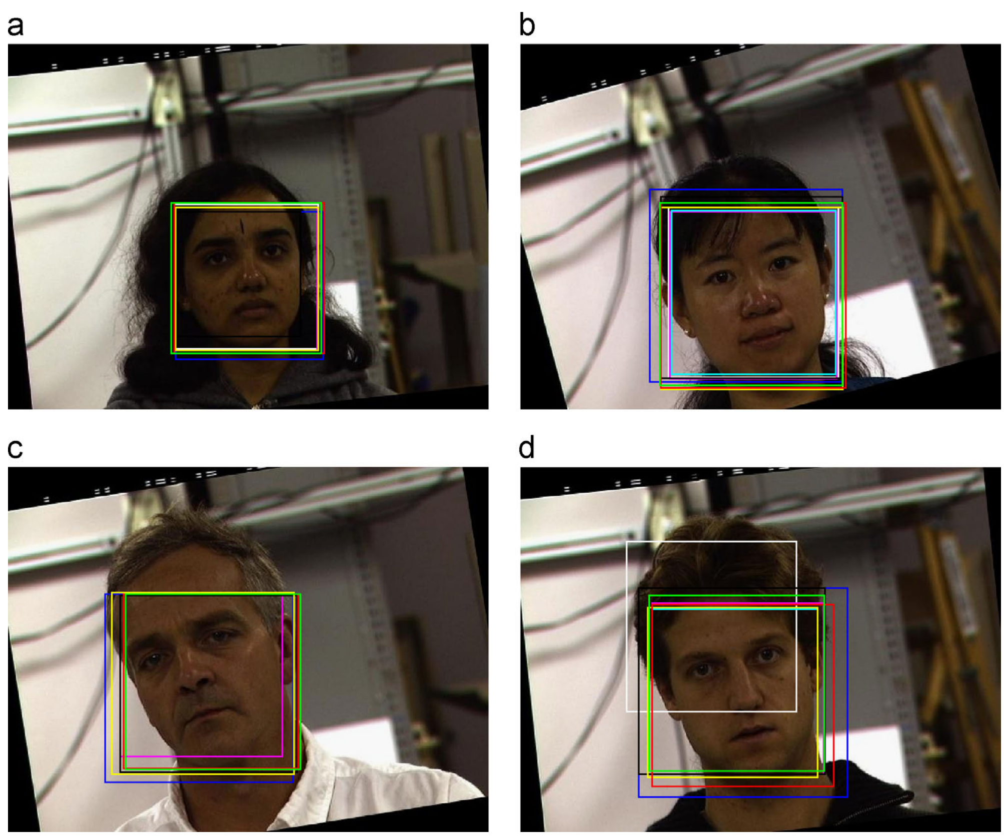

Fig. 17. Detection results of CMU PIE dataset.

method effectively finds faces in the examples which contain complex background, pose variation and multiple faces. However, methods using skin color are prone to face occlusion problem because occlusion basically makes it difficult to handle skin color. Face detection performance degrades as the structural distribution of skin color of a face varies widely when it is occluded. Overall, we observe that our method performs well under complex background in the presence of face pose variation and multiple faces conditions,

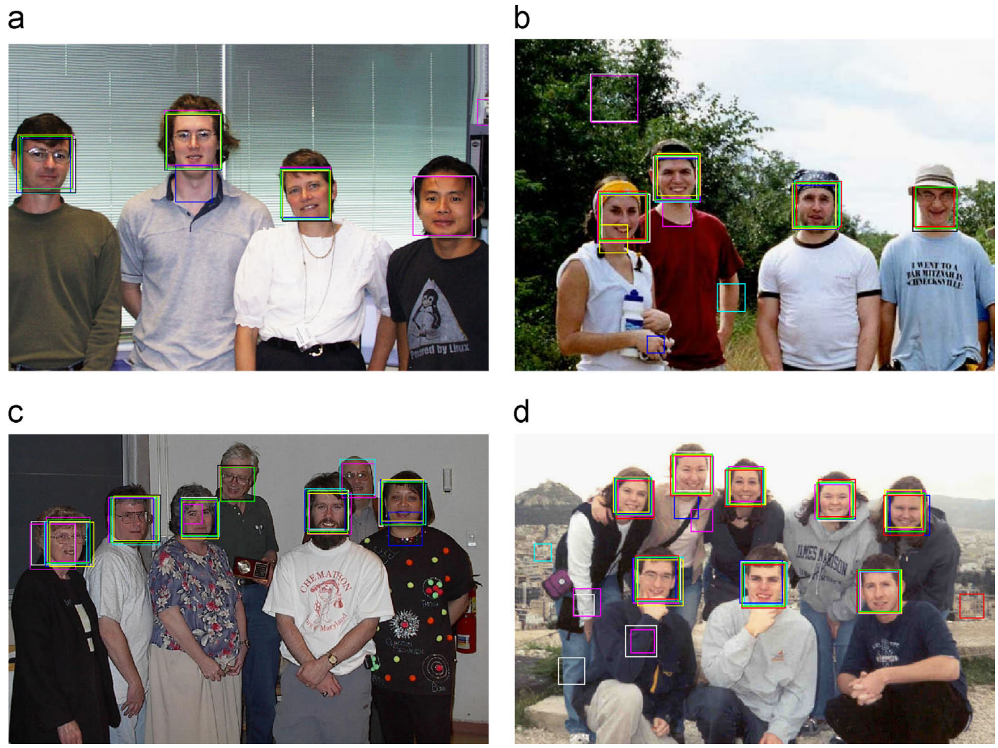

Fig. 18. Detection results of Bao dataset.

Fig. 19. Detection results of FDDB dataset.

whereas it does not perform well when face occlusion is considered and when skin colors vary too broadly which can be dependent to the color of surrounding environment or severe lighting condition (Table 5).

4.3.3. Time complexity

Training procedure for boosting becomes lighter due to reduced information of non-skin color as background. Skin strong classifier of the first cascade stage pays less attention to information of non-skin

Table 3
False acceptance rate.

	Dataset	Viola et al.	Lienhart et al.	Zhang et al.	Yan-Wen et al.	Green Span et al.	Kai-Biao et al.	Proposed method w/Haar	Proposed method w/LBP
Mean	Pointing'04	0.4182	0.4206	0.3624	0.4573	0.4380	0.4535	0.3985	0.4017
	IMM	0.4370	0.4028	0.4071	0.4590	0.4308	0.4691	0.4239	0.4473
	Caltech	0.5758	0.4119	0.3655	0.3813	0.3256	0.3529	0.2172	0.2172
	CMU PIE	0.5560	0.4024	0.4004	0.6932	0.3933	0.4128	0.3129	0.3070
	Bao	0.5102	0.5369	0.4889	0.4099	0.3788	0.3937	0.3213	0.3029
	FDDB	0.5334	0.5474	0.5045	0.5593	0.5042	0.5676	0.5319	0.5276
Standard deviation	Pointing'04	0.0051	0.0061	0.0051	0.0018	0.0019	0.0053	0.0018	0.0020
	IMM	0.0022	0.0018	0.0052	0.0011	0.0012	0.0810	0.0095	0.0088
	Caltech	0.0124	0.0141	0.0134	0.0108	0.0105	0.0074	0.0005	0.0019
	CMU PIE	0.0052	0.0030	0.0065	0.0057	0.0059	0.0019	0.0007	0.0029
	Bao	0.0293	0.0198	0.0282	0.0114	0.0189	0.0181	0.0022	0.0317
	FDDB	0.0448	0.0436	0.0462	0.0254	0.0419	0.0326	0.0033	0.0927

Table 4
False rejection rate.

	Dataset	Viola et al.	Lienhart et al.	Zhang et al.	Yan-Wen et al.	Green Span et al.	Kai-Biao et al.	Proposed method w/Haar	Proposed method w/LBP
Mean	Pointing'04	0.4139	0.3455	0.3400	0.2551	0.2290	0.2826	0.1805	0.1951
	IMM	0.6525	0.5637	0.5485	0.3267	0.3465	0.4328	0.2764	0.2911
	Caltech	0.3769	0.3663	0.3884	0.4042	0.3781	0.3895	0.3569	0.3326
	CMU PIE	0.4843	0.4488	0.5496	0.4203	0.4230	0.6089	0.4238	0.4412
	Bao	0.3975	0.3859	0.4311	0.4136	0.3562	0.3996	0.3519	0.3525
	FDDB	0.5539	0.5131	0.5748	0.7865	0.7248	0.7753	0.7513	0.7669
Standard deviation	Pointing'04	0.0467	0.0829	0.1082	0.1381	0.1315	0.0787	0.0370	0.0368
	IMM	0.1051	0.1230	0.1748	0.1813	0.2116	0.2159	0.1522	0.1643
	Caltech	0.0361	0.0552	0.0221	0.0186	0.0201	0.0590	0.0106	0.0181
	CMU PIE	0.0852	0.1524	0.1388	0.0450	0.0474	0.0953	0.0058	0.0678
	Bao	0.0488	0.0859	0.0247	0.0280	0.0276	0.0602	0.1562	0.0195
	FDDB	0.0815	0.146	0.0614	0.0434	0.0416	0.116	0.2485	0.0322

Table 5
Half total error rate.

	Dataset	Viola et al.	Lienhart et al.	Zhang et al.	Yan-Wen et al.	Green Span et al.	Kai-Biao et al.	Proposed method w/Haar	Proposed method w/LBP
Mean	Pointing'04	0.4161	0.3831	0.3512	0.3562	0.3335	0.3681	0.2895	0.2984
	IMM	0.5448	0.4833	0.4778	0.3929	0.3887	0.4510	0.3502	0.3692
	Caltech	0.4764	0.3891	0.3770	0.3928	0.3519	0.3712	0.2871	0.2749
	CMU PIE	0.5202	0.4256	0.4750	0.5568	0.4082	0.5109	0.3684	0.3741
	Bao	0.4539	0.4614	0.4600	0.4118	0.3675	0.3967	0.3366	0.3277
	FDDB	0.5437	0.5303	0.5397	0.6729	0.6145	0.6715	0.6416	0.6473

color, and so do the later successive stages. As a result of focusing particularly on information of skin color, we could maintain a small number of training cascade stages as shown in Table 6. Moreover, we compare the computational times of skin color modeling and transformation based on the results using SGM, GMM and histogram. Although histogram is quite time consuming for building the lookup table in the training phase, it shows much faster computation in the testing phase compared to SGM and GMM due to simple table lookup (see Table 7). We also measure the processing time to test a single image. Instead of the advantages to face pose variation and complex background, additional use of skin color information needs additional processing time. However, our method efficiently retains certain processing speed both by using simple Bayesian table lookup and by placing less emphasis on non-skin color region while improving the detection performance against face pose variation and complex background (Table 8).

5. Conclusion

We proposed a boosting-based face detection method based on the likelihood of skin color in this paper. Our method emphasizes skin color information while simultaneously deemphasizes non-skin color information. Skin color emphasis could be well combined with iterative boosting algorithm. The proposed method shows improvement over those conventional methods of [9–14] against severely varying face pose and complex background. Our proposed method substantially reduces the half total error rate and maintains both a small number of training cascade stages and certain processing speed. Two important aspects have been observed. Firstly, skin color is effectively highlighted so that the skin boosted classifier concentrates on structural distribution of skin color of face rather than the details of facial components in gray-level brightness. This provides tolerance against the face pose variation in yaw, roll and pitch

Table 6
Number of training cascade stages.

	Dataset	Viola et al.	Lienhart et al.	Zhang et al.	Yan-Wen et al.	Green Span et al.	Kai-Biao et al.	Proposed method w/Haar	Proposed method w/LBP
Mean	Pointing '04	9.1000	9.2000	9.0000	5.1000	6.0000	7.0000	5.0000	7.4000
	IMM	8.2000	8.1000	8.0000	7.0000	9.0000	7.2000	6.1000	8.0000
	Caltech	8.8000	8.9000	8.1000	6.8000	7.2000	7. 1000	5.6000	7.1000
	CMU PIE	11.9000	12.2000	12.1000	6.2000	7.8000	8.2000	7.7000	8.3000
	Bao	12.2000	12.8000	12.7000	9.0000	9.6000	9.8000	8.2000	8.9000
	FDDB	12.9000	13.0000	13.1000	10.0000	10.2000	9.9000	8.2000	9.1000
Standard deviation	Pointing '04	0.3162	0.4216	0.0000	0.3162	0.0000	0.0000	0.0000	0.5164
	IMM	0.4216	0.3162	0.0000	0.4714	0.0000	0.4216	0.3162	0.3162
	Caltech	0.4216	0.5676	0.3162	0.4216	0.4216	0.5676	0.5164	0.3162
	CMU PIE	0.5676	0.4216	0.5676	0.4216	0.4216	0.4216	0.4830	0.4830
	Bao	0.4216	0.4216	0.4830	0.4714	0.5164	0.4216	0.4216	0.5676
	FDDB	0.5676	0.4714	0.5676	0.4714	0.4216	0.5676	0.4216	0.3162

Table 7
Computational time per image for skin color modeling and transformation (s).

	Dataset	Modeling				Transformation		
		Training images	SGM	GMM	Histogram	SGM	GMM	Histogram
Mean	Pointing'04	353	2.558	27.714	51.013	0.1984	0.7257	0.0951
	IMM	1080	10.605	69.288	87.156	0.1723	0.7223	0.0980
	Caltech	225	5.624	82.881	165.459	0.2719	1.0300	0.1372
	CMU PIE	918	33.696	304.956	339.66	0.2692	0.8112	0.0981
	Bao	198	4.935	76.863	155.478	0.7969	1.5235	0.1794
	FDDB	2560	39.790	311.756	380.191	0.8797	1.7230	0.1991
Standard deviation	Pointing'04	353	0.045	0.125	2.319	0.0046	0.0047	0.0042
	IMM	1080	0.303	0.441	4.587	0.0049	0.0046	0.0051
	Caltech	225	0.281	0.872	13.788	0.0130	0.0108	0.0111
	CMU PIE	918	1.162	3.244	28.305	0.0091	0.0086	0.0081
	Bao	198	0.330	0.926	15.678	0.0205	0.0182	0.0128
	FDDB	2560	0.486	4.627	11.984	0.0302	0.0235	0.0186

Table 8
Computational time per half-sized image for testing (s).

	Dataset	Viola et al.	Lienhart et al.	Zhang et al.	Yan-Wen et al.	Green Span et al.	Kai-Biao et al.	Proposed method w/Haar	Proposed method w/LBP
Mean	Pointing'04	0.0057	0.0190	0.0289	0.2539	1.0503	0.0823	0.1026	0.1189
	IMM	0.0151	0.0152	0.0217	0.1901	0.7427	0.0701	0.1178	0.1195
	Caltech	0.0283	0.0234	0.0426	0.2189	0.3668	0.0743	0.1823	0.1896
	CMU PIE	0.0201	0.0159	0.0267	0.2234	0.8749	0.0895	0.1244	0.1219
	Bao	0.2108	0.2060	0.3074	0.5911	1.2151	0.1354	0.3739	0.3992
	FDDB	0.2561	0.2502	0.3756	0.6828	1.3120	0.2083	0.4962	0.4883
Standard deviation	Pointing'04	0.0014	0.0004	0.0003	0.0017	0.0016	0.0009	0.0019	0.0004
	IMM	0.0019	0.0013	0.0010	0.0020	0.0041	0.0029	0.0034	0.0005
	Caltech	0.0047	0.0010	0.0014	0.0046	0.0041	0.0056	0.0068	0.0097
	CMU PIE	0.0021	0.0005	0.0006	0.0034	0.0018	0.0034	0.0061	0.0011
	Bao	0.2212	0.2365	0.0714	0.1956	0.1990	0.1065	0.1448	0.1395
	FDDB	0.0528	0.0576	0.0650	0.1592	0.1837	0.1728	0.1444	0.1058

directions. Secondly, non-skin information is remarkably reduced with respect to the complex background. Skin boosted classifier effectively extracts distinctive features by deemphasizing background information and selectively accentuating the skin color information at the same time. Consequently, our method captures the characteristic of face based on the relative distribution of skin color likelihood of face rather than segmenting out skin region, which results in less number of falsely accepted examples against complex background.

Future work includes studying face detection using skin color likelihood based on facial landmarks to locally deal with face occlusion and face pose estimation problems in details. Another topic is to find feature descriptor that provides good representation of skin color in the presence of severe lighting condition.

Conflict of interest statement

None declared.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (NRF-2011-0016302).

References

[1] Ming-Hsuan Yang, D.J. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 34–58.
[2] Yanjiang Wang, Baozong Yuan, A novel approach for human face detection from color images under complex background, Pattern Recognition 34 (2001) 1983–1992.
[3] Hongxun Yao, Wen Gao, Face detection and location based on skin chrominance and lip chrominance transformation from color images, Pattern Recognition 34 (2001) 1555–1564.
[4] Rein-Lien Hsu, M. Abdel-Mottaleb, A.K. Jain, Face detection in color images, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 696–706.
[5] Iyad Aldasouqi, Mahmoud Hassan, Smart human face detection system, Int. J. Comput. 5 (2011) 210–221.
[6] Sanjay Kr. Singh, D.S. Chauhan, Mayank Vatsa, Richa Singh, A robust skin color based face detection algorithm, Tamkang J. Sci. Eng. 6 (2003) 227–234.
[7] Kakumanu, Makrogiannis, Bourbakis, A survey of skin-color modeling and detection methods, Pattern Recognition 40 (2007) 1106–1122.
[8] Vladimir Vezhnevets, Vassili Sazonov, Alla Andereeva, A survey on pixel-based skin color detection techniques, in: Proceedings of the GraphiCon, 2003, pp. 85–92.
[9] P. Viola, M. Jones, Rapid object detection using boosted cascade of simple features, in: Proceedings of the 2001 IEEE Computer Society Conference on CVPR, vol. 1, 2001, pp. 511–518.
[10] R. Lienhart, J. Maydt. An extended set of Haar-like features for rapid object detection, in: Proceedings of ICIP, 2002, pp. 900–903.
[11] G. Zhang, X. Huang, S.Z. Li, Y. Wang, X. Wu, Boosting local binary pattern (LBP)-based face recognition, in: Proceedings of the Advances in Biometric Person Authentication, Lecture Notes in Computer Science, vol. 3338, 2004, pp. 179–186.
[12] Yan-Wen Wu, Xue-Yi Ai, Face detection in color images using AdaBoost algorithm based on skin color information, in: Proceedings of the First International Workshop on Knowledge Discovery and Data Mining, 2008, WKDD 2008, 2008, pp. 339–342.
[13] Greenspan Hayit, Jacob Goldberger, Itay Eshet, Mixture model for face-color modeling and segmentation, Pattern Recognition Lett. 22 (2001) 1525–1536.
[14] Kai-Biao Ge, Jing Wen, Bin Fang, Adaboost algorithm based on MB-LBP features with skin color segmentation for face detection, in: Proceedings of the 2011 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), 2011, pp. 40–43.
[15] JPEG File Interchange Format Version 1.02.
[16] McKenna, J. Stephen, Shaogang Gong, Yogesh Raja, Modelling facial colour and identity with gaussian mixtures, Pattern Recognition 31 (1998) 1883–1892.
[17] M. Hunke, A. Waibel, Face locating and tracking for human–computer interaction, in: Proceedings of the 28th Asilomar Conference on Signals, Systems and Computers, vol. 2, 1994, pp. 1277–1281.
[18] Karl Schwerdt, James L. Crowley, Robust face tracking using color, in: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 90–95.
[19] Paul A. Viola, Michael J. Jones, Robust real-time face detection, Int. J. Comput. Vision 57 (2004) 137–154.
[20] Yoav Freund, Robert E. Schapire, A decision-theoretic generation of on-line learning and application to boosting, J. Comput. Syst. Sci. 55 (1997) 119–139.
[21] Zhang Xiaozheng, Yongsheng Gao, Face recognition across pose: a review, Pattern Recognition 42 (2009) 2876–2896.
[22] Ahonen Timo, Abdenour Hadid, Matti Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 2037–2041.
[23] N. Gourier, D. Hall, J.L. Crowley, Estimating face orientation from robust detection of salient facial features, in: Proceedings of Pointing 2004, International Workshop on Visual Observation of Deictic Gestures, ICPR, 2004, pp. 17–25.
[24] M.B. Stegmann, B.K. Ersbøll, R. Larsen., FAME — a flexible appearance modeling environment, IEEE Trans. Med. Imaging 22 (2003) 1319–1331.
[25] 〈http://www.vision.caltech.edu/html-files/archive.html〉.
[26] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression database, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 1615–1618.
[27] R. Frischholz, Bao face database at the face detection homepage, in 〈http://www.facedetection.com〉.
[28] Vidit Jain, Erik Learned-Miller, FDDB: A Benchmark for Face Detection in Unconstrained Settings, Technical Report UM-CS-2010-009, Department of Computer Science, University of Massachusetts, Amherst, 2010.
[29] M. Soriano, B. Marinkauppi, S. Huovinen, M. Laaksonen, Adaptive skin color modeling using the skin locus for selecting training pixels, Pattern Recognition 36 (2003) 681–690.
[30] Kar-Ann Toh, Jaihie Kim, Sangyoun Lee, Biometric scores fusion based on total error rate minimization, Pattern Recognition 41 (2008) 1066–1082.
[31] S. Bengio, J. Mariéthoz, A statistical significance test for person authentication, in: Proceedings of Odyssey04—The Speaker and Language Recognition Workshop, 2004, pp. 237–244.

Yuseok Ban received his B.S. degree in Electrical and Electronic Engineering from Yonsei University, Seoul, Korea. Currently, he is a candidate of Ph.D. degree in Electrical and Electronic Engineering from Yonsei University, Seoul, Korea. His research interests include pattern recognition, biometrics and computer vision.

Sang-Ki kim received his B.S. and Ph.D. degrees in Electrical and Electronic Engineering from Yonsei University, Seoul, Korea in 2004 and 2011 respectively. He is currently a senior research engineer at LG Electronics. His research interests are in the fields of pattern recognition and computer vision with special focus on face and gesture recognition.

Sooyeon Kim received her B.S. degree in Electrical and Electronic Engineering from Yonsei University, Seoul, Korea. Currently, she is a candidate of Ph.D. degree in Electrical and Electronic Engineering from Yonsei University, Seoul, Korea. Her research interests include face and gesture recognition.

Kar-Ann Toh is a full time professor in the School of Electrical and Electronic Engineering at Yonsei University, South Korea. He received the Ph.D. degree from Nanyang Technological University (NTU), Singapore. He worked for 2 years in the aerospace industry prior to his post-doctoral appointments at research centres in NTU from 1998 to 2002. He was affiliated with Institute for Infocomm Research in Singapore from 2002 to 2005 prior to his current appointment in Korea. His research interests include biometrics, pattern classification, optimization and neural networks. He is a co-inventor of a US patent and has made several PCT filings related to biometric applications. Besides being an active member in publications, Dr. Toh has served as a member of technical program committee for international conferences related to biometrics and artificial intelligence. He is currently an associate editor of Pattern Recognition Letters and a senior member of the IEEE.

Sangyoun Lee received his B.S. and M.S. degrees in Electronic Engineering from Yonsei University, Seoul, South Korea in 1987, 1989 respectively. He received his Ph.D. degree in Electrical and Computer Engineering from Georgia Tech., Atlanta, Georgia, in 1999. He was a Senior Researcher in Korea Telecom from 1989 to 2004. He is now a faulty member of the School of Electrical and Electronic Engineering, Yonsei University, Korea. His research interests include pattern recognition, computer vision, video coding and biometrics.