

Learning-based Face Detection by Adaptive Switching of Skin Color Models and AdaBoost under Varying Illumination

Deng-Yuan Huang¹, Chun-Jih Lin¹, Wu-Chih Hu²

¹Department of Electrical Engineering, Dayeh University, Taiwan
168, University Rd., Dacun, Changhua 51519, Taiwan
kevin@mail.dyu.edu.tw

²Department of Computer Science and Information Engineering,
National Penghu University of Science and Technology, Taiwan
300, Liu-Ho Rd., Makung, Penghu 88050, Taiwan
wchu@npu.edu.tw

Received February 2010; revised August 2010

ABSTRACT. *Face detection has a variety of applications, such as face recognition, face expression analysis, and video conferencing. However, most existing methods for face detection are sensitive to lighting variation. In this paper, a novel scheme invariant with illumination based on adaptive switching of skin color models (ASSM) with lighting compensation for face detection is proposed. The skin-tone pixels detected are connected by the proposed fast 8-connected component labeling method into a more compact skin cluster. An optimal skin color model is thus adaptively selected using a well-defined quality measure. Possible face candidates are further validated by a cascaded AdaBoost detector. Experimental results indicate that robust face detection can be achieved for various lighting conditions, such as dim light, side light, and back light. A detection time of 60 ms for each frame is achieved with the aid of the ASSM method. A detection rate of 94.4% was obtained for a test video sequence.*

Keywords: skin color model, AdaBoost, face detection

1. **Introduction.** Face detection is of importance in a wide variety of applications, such as facial expression analysis [1], speaker recognition [2], VoIP (voice over IP) security [3], and identity authentication. A significant degree of separability usually exists between skin-tone pixels and their background. Face detection schemes often adopt the extraction of skin-tone colors [4, 5, 6] since they are invariant with scales, poses, and facial expressions. However, it is difficult to robustly detect skin-tone pixels in the presence of a complex background and varying illumination using color-based approaches.

Skin-tone colors are widely used as a primary feature of a face. Ishii et al. [4] extracted face candidates from skin color parts detected using a back-propagation neural network which learned from the YCbCr color information of face images. Since search regions are confined within the skin color parts, the performance of face detection can be significantly improved. Skin-color models are constructed using a large number of skin-tone pixels for detecting possible candidates of faces. Jones et al. [5] constructed color models for skin and non-skin classes from a dataset of nearly one billion labeled pixels using a statistical method. The classes used to build a skin pixel detector show good separability. They

also found that a histogram model is superior to a mixture model in terms of accuracy and computational cost for skin detection.

Hsu et al. [6] modeled skin color using a parametric ellipse in a 2D color space and extracted facial features by constructing feature maps for the eyes, mouth, and face boundary. In their work, they proposed a lighting compensation technique that uses a reference white to normalize the color appearance. They regarded pixels with the top 5 percent of the luma (nonlinear gamma-corrected luminance) values in the image as the reference white only if the number of these pixels is sufficiently large. The red (R), green (G), and blue (B) components of a color image are adjusted so that the average gray value of these reference white pixels is linearly scaled to 255.

Compact skin clusters can be obtained in the YCbCr color space under a wide range of lighting variation [7, 8]. Garcia et al. [7] applied quantized skin color region merging and wavelet packet analysis to face detection. They found that skin color pixels can form a single but compact cluster in both the YCbCr and HSV color spaces, even though skin-tone color distributions for human races vary widely. This result shows that the cluster of skin color is less compact in the HSV space than it is in the YCbCr space, and that the HSV space is more sensitive to lighting variation. Similar results have been found elsewhere [8]. However, skin color models that use the YCbCr space frequently misclassify non-skin pixels at low luminance as skin-tone pixels, and vice versa [6], due to their nonlinear dependence on luma.

In general, skin-tone pixels have a distinct shape in the normalized color space (r, g). Soriano et al. [9] used the color space (r, g) to perform face detection with high accuracy under daylight, incandescent light, fluorescent light, and a combination of these light sources. They applied an adaptive histogram back-projection approach, where the skin color model is updated by pixels in the regions which fall in the so-called skin locus. In the locus, the shape of the range of skin-tone pixels is quite different from that in the (r, g) color space.

Bergasa et al. [10] performed a study of the distribution of human skin color in various color spaces (RGB, normalized RGB, HSI, SCT, YQQ) and came to the conclusion that the best space for this application is normalized RGB. In this space, the human skin color forms a compact class; the color differences between people can be reduced by working with chromaticity to eliminate intensity. For some lighting conditions, the skin color distribution can be modeled by a Gaussian function in the (r, g) color space [11].

Since a single skin color model cannot deal with similar skin color pixels in the background and under a wide range of lighting conditions, several adaptive color space switching methods have been presented [12, 13]. Stern et al. [12] proposed an adaptive color space switching method for face tracking that is a combination of color-space models (CSMs) and color-distribution models (CDMs). In the CSMs, the color spaces of RG, rg, HS, IQ, and CbCr are used, and in the CDMs, the probability model (Pr), the possibility model (Ps), a Gaussian with one component (G1), a Gaussian with three components (G3), and a Gaussian with a dynamic number of components (GD) are used. They found that the tracking performance can be significantly enhanced by switching the color spaces throughout the tracking sequences.

More recently, Chang et al. [13] proposed a support vector machine (SVM) based adaptive color switching method for a face tracking system. They used four color spaces, i.e., YCbCr, normalized RGB, XYZ, and YIQ, and employed the Laws textures of L5E5, R5E5, and W5W5 as discriminative features of faces to train their SVMs for each color space. A quality measure was then presented to adaptively select an optimal color space under varying illumination.

Many color constancy schemes have been suggested, but so far their performance is not satisfactory. In this paper, we propose a novel scheme based on adaptive switching of skin color models (ASSM) with lighting compensation for face detection under unconstrained scene conditions. Detected skin-tone pixels are connected by the proposed fast 8-connected component labeling method into a more compact cluster, which is further validated by the cascaded AdaBoost algorithm to determine whether it is a human face. Extensive experiments show that the performance of face detection can be greatly enhanced by switching the skin color models throughout the tracking sequence.

The remainder of this paper is organized as follows. In Section 2, the method of adaptive switching for skin color models is described. Section 3 contains the detailed procedures of the proposed fast 8-connected component labeling method. Section 4 presents the results of face detection in a video sequence under varying illumination. The conclusion and future work are given in Section 5.

2. Proposed face detection method. A flow chart of the proposed method for face detection in a video sequence is shown in Figure 1. First, the input video sequence is resized to 80×60 pixels by bilinear interpolation. Second, the skin-tone pixels are detected by the proposed ASSM method, which includes all the possible combinations of three skin color models (i.e., YCbCr [6, 7, 8], Soriano's model [9], and the Gaussian mixture model [10, 11]) and three lighting compensation models (i.e., reference white [6], modified reference white [14], and gray world [15]). Third, detected skin-tone pixels are connected by the proposed fast 8-connected component labeling method into a more compact cluster. Fourth, the quality measures r^k are estimated using the skin clusters for the nine skin color models. The optimal skin color model (that with maximum r^k) is thus selected. Finally, possible face candidates are validated by the cascaded AdaBoost algorithm to determine whether they are human faces.

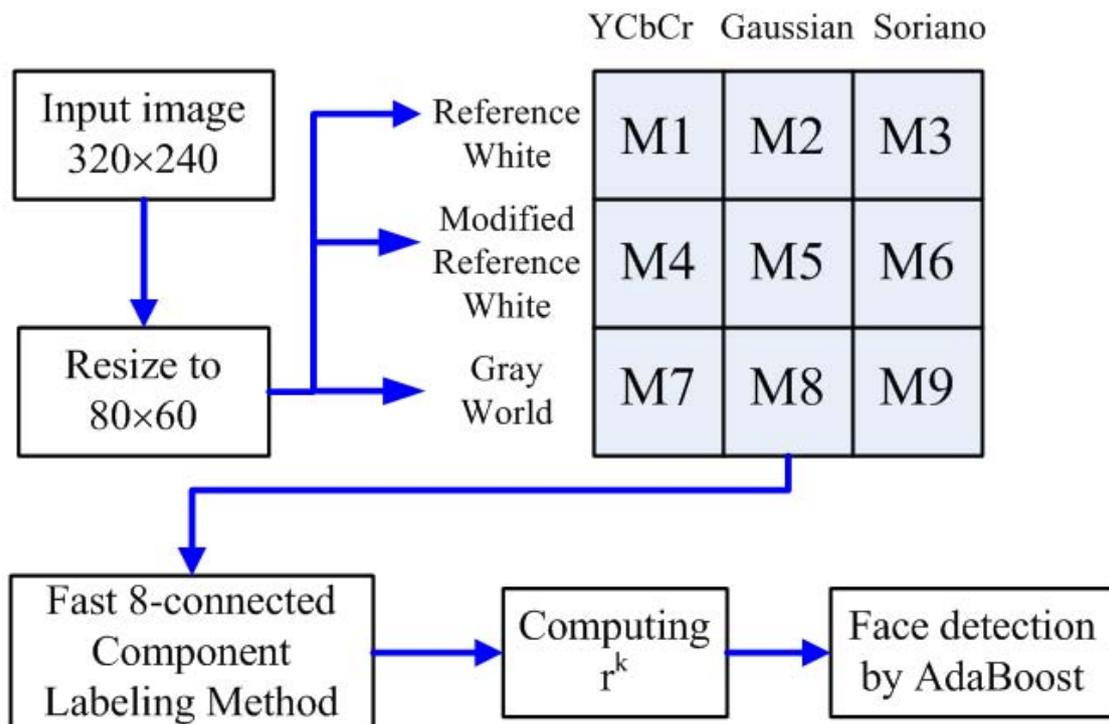


FIGURE 1. Flow chart for proposed face detection method

2.1. Skin color models. The three skin color models, i.e., YCbCr [6, 7, 8], Soriano's model [9], and the Gaussian mixture model [10, 11], used in the proposed method are described below.

2.1.1. YCbCr skin color model. YCbCr is a family of color spaces that is used as a part of the color image pipeline in response to increasing demand for digital approaches in handling video information. It has become a widely used model in digital video. Since the skin-tone color depends on luminance, a conversion of RGB to the YCbCr color space to make the skin cluster luma-independent enables robust detection of dark and light skin-tone colors. Therefore, the YCbCr color space leads to better performance under varying illumination conditions when compared to the RGB color space. The corresponding skin cluster can be determined as:

$$\begin{aligned} (Y, Cb, Cr) \text{ is classified as skin pixels if :} \\ 60 \leq Y \leq 255 \\ 100 \leq Cb \leq 125 \\ 135 < Cr \leq 170 \\ \text{where } Y, Cb, Cr \in [0, 255] \end{aligned}$$

2.1.2. Soriano's skin color model. Normalized RGB is easily obtained from the RGB values by a simple normalization procedure, i.e., $r=R/(R+G+B)$, $g=G/(R+G+B)$, and $b=B/(R+G+B)$. The b component is usually ignored since it does not hold any significant information. The remaining (r, g) components are often called "pure colors", since the dependence of (r, g) on the brightness of the source RGB color can be greatly reduced by the normalization procedure. Another attractive property of normalized RGB is that for matte surfaces (nonspecular object materials), it is invariant to changes of surface orientation relative to the light source when ignoring ambient light [17]. For Soriano's skin color model, the normalized RG color space is first applied, and a pair of quadratic functions is used to define the upper and lower bounds of the skin locus [9]. To prevent grayish and whitish pixels from being labeled as skin, pixels are excluded from skin membership if they fall within a circle with radius 0.02 around the white point ($r=g=0.33$). Here, we slightly modify this model with an additional RGB constraint, C_{RGB} . Therefore, the skin cluster of value S can be determined in terms of chromaticity (r, g) and the source RGB space in Eq.(1), where the dot (\cdot) means the logical operator "and".

$$\begin{aligned} S = \begin{cases} 1, & (g < g_u) \cdot (g > g_d) \cdot (R_W > 0.0004) \cdot C_{RGB} \\ 0, & \text{otherwise} \end{cases} \\ \text{where} \\ g_u = -1.3767r^2 + 1.0743r + 0.1452 \\ g_d = -0.776r^2 + 0.5601r + 0.1766 \\ R_W = (r - 0.33)^2 + (g - 0.33)^2 \\ C_{RGB} = (R > 130) \cdot (B > 55) \cdot (G > B) \cdot ((R - G) > 25) \end{aligned} \quad (1)$$

2.1.3. Gaussian mixture skin color model. The Gaussian mixture model, which is a generalization of the single Gaussian joint probability density function, is commonly used to describe a complex-shaped distribution of skin-tone pixels. Since the Gaussian mixture model is parametric, the representation of skin color distribution is more compact than those of most popular histogram-based non-parametric skin color models. In addition, the Gaussian mixture model has the ability to interpolate and generalize incomplete training data. It is expressed by only a small number of parameters and needs very little storage space. The Gaussian mixture model can also be viewed as a form of a generalized radial

basis function (RBF) network. It is used to describe the skin cluster in the RGB color space as:

$$p(x|skin) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right] \quad (2)$$

where

$$x \in \{R, G\}, 51 \leq R \leq 102, 51 \leq G \leq 153$$

where μ_x and σ_x are the mean value and standard deviation estimated from the specified skin color regions of R and G channels, respectively. In Eq.(2), the probability of $p(x|skin)$ is used to measure how skin-like the color x is. Let $G_{max} = \mu_G + 2\sigma_G$, $G_{min} = \mu_G - 2\sigma_G$, $R_{max} = \mu_R + 2\sigma_R$, and $R_{min} = \mu_R - 2\sigma_R$ be the upper and lower bounds of possible skin-tone pixels for R and G channels, respectively. The white point (R=G=84) is excluded from the skin color membership. Thus, the corresponding skin cluster can be expressed by Eq.(3), where the dot (\cdot) means the logical operator "and".

$$S(R, G) = \begin{cases} 1, & (G \leq G_{max}) \cdot (G \geq G_{min}) \cdot (R \leq R_{max}) \cdot (R \geq R_{min}) \cdot (R_W > 26) \\ 0, & otherwise \end{cases}$$

where

$$R_W = (R - 84)^2 + (G - 84)^2 \quad (3)$$

2.2. Lighting compensation methods. The three lighting compensation methods, i.e., reference white [6], modified reference white [14] and gray world [15], used in the proposed method are described below.

2.2.1. Reference white lighting compensation. Lighting compensation is commonly used to normalize the skin color appearance since the reflection of skin-tone color greatly depends on the lighting conditions. Reference white, first presented by Hsu et al. [6], is the most popular method. The method removes bias color pixels from a color image because these pixels appear as "real white". In applications of face detection, reference white detects few non-face pixels and most skin-tone facial pixels [6].

The top 5% of the luma values in the image is regarded as the reference white if the number of these pixels is sufficiently large (>100). The R, G, and B components of a color image are then adjusted so that the average gray value of the reference white pixels is linearly scaled to 255. Let $i \in [l_u, 255]$ be the top 5% gray levels and f_i be the pixel number with gray level i in the image. Thus, the modified RGB components can be estimated as:

$$M_{top} = \sum_{i=l_u}^{255} i \cdot f_i / \sum_{i=l_u}^{255} f_i$$

$$\chi_{new} = \chi_{old} / M_{top} \times 255, \text{ where } \chi \in \{R, G, B\}$$

2.2.2. Modified reference white. This modified version of reference white was proposed by Xu [14]. The bottom 5% gray levels are also considered. Let $i \in [l_u, 255]$ and $i \in [0, l_d]$ be the top 5% and bottom 5% gray levels in the image, respectively. The modified RGB components can be calculated as:

$$\chi_{new} = (\ln(\chi_{old}) - \ln(l_d)) / (\ln(l_u) - \ln(l_d)) \times 255$$

where $\chi \in \{R, G, B\}$

2.2.3. Gray world lighting compensation. White balance is a fundamental function in the processing pipeline of a digital camera. It is usually used to correct the color of pixels under varying luminous scenarios. The white balance can be set manually or automatically. Of these two approaches, automatic white balance is more preferred for consumer digital cameras. To set white balance automatically, the gray world assumption [15] is commonly used. Gray world is a lighting compensation method which seeks to equalize the mean of the R, G, and B channels. The gray world assumption is based on the observation that for a typical scene, the average intensity of the R, G, and B channels should be equal. Let M and N be the image height and width, respectively. The averages of RGB channels and gray levels, χ_{AVG} and μ_{AVG} , are first calculated, respectively, as follows:

$$\begin{aligned}\chi_{AVG} &= \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N I_{\chi}(x, y), \text{ where } \chi \in \{R, G, B\} \\ \mu_{AVG} &= \frac{1}{3} (R_{AVG} + G_{AVG} + B_{AVG})\end{aligned}$$

The scale ratios and modified pixels of the original RGB channels, A_{χ} and \hat{I}_{χ} , are then estimated, respectively.

$$A_{\chi} = \mu_{AVG} / \chi_{AVG}, \text{ and } \hat{I}_{\chi} = A_{\chi} \cdot I_{\chi}(x, y)$$

2.3. Face detection using AdaBoost algorithm. Haar-like features (see Figure 2) are widely used by the AdaBoost algorithm with the concept of an integral image for efficient feature computation. AdaBoost learning selects a small number of weak classifiers which represent the local discriminative features of faces, and then combines them into a strong classifier to decide whether a detected cluster is a human face. AdaBoost can deal with very large sets of weak classifiers due to its greedy characteristics. To significantly improve computational efficiency and reduce the false positive rate (FPR), a sequence of strong classifiers is concatenated as a so-called cascaded detector. More details of computing the integral image and training the cascade AdaBoost detector can be found in [16].



FIGURE 2. Set of Haar-like features used in the AdaBoost method

To train the cascade AdaBoost detector, a training set in CBCL (Center for Biological and Computational Learning at MIT) face database #1 [18] was used, which consists of grayscale images of 2,429 faces and 4,548 non-faces with a size of 19×19 pixels. The face images have various illumination conditions, facial expressions (e.g., open/closed eyes, smiling/not smiling) and facial details (e.g., glasses/no glasses). Some typical face and non-face examples from the CBCL face database are shown in Fig. 3(a) and 3(b), respectively.

2.4. Quality measure. The skin-tone pixels detected by ASSM are connected by the proposed fast 8-connected component labeling method into a more compact cluster. Only those clusters with greater than 50 pixels are considered. Possible face regions are then fitted with a rectangle W_r or an ellipse W_e , as shown in Figure 4. To evaluate the quality measures for all the skin color models, a quality measure r^k representing the k^{th} skin color model is proposed. It is defined as:

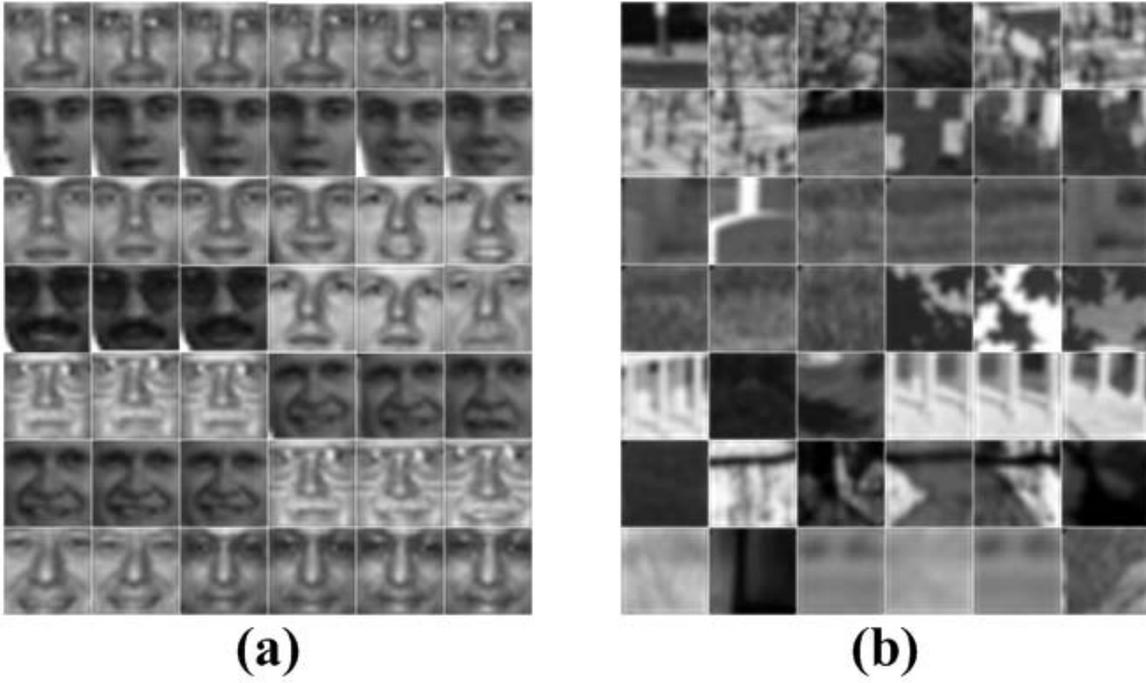


FIGURE 3. Some typical training examples in the CBCL face database. (a) Face images; (b) non-face images

$$r^k = \omega_1 r_1^k + \omega_2 r_2^k \quad (4)$$

where

$$r_1^k = \sum_{i=1}^{N_C} \left(\sum_{(x,y) \in W_r} p(x,y) / W_r \right) / N_C \quad (5)$$

and

$$r_2^k = \sum_{i=1}^{N_C} ((S_e^{W_e} + S_p^{W_e} + S_a^{W_e})/3) / N_C \quad (6)$$

where $p(x,y)$ is a detected skin-tone pixel in W_r or W_e ; ω_1 and ω_2 are weights which are both set to 0.5 since the two models of r_1 and r_2 are equally important, as determined from experiments; and N_C is the total number of all the 8-connected components with skin-tone pixels greater than 50. $S_e^{W_e}$, $S_p^{W_e}$, and $S_a^{W_e}$ represent the sensitivity (i.e., true positive rate= $TP/(TP+FN)$), the specificity (i.e., true negative rate= $TN/(TN+FP)$), and the spatial accuracy (i.e., $1-(FP+FN)/(TP+FN)$), respectively. These values are estimated from the elliptical regions, W_e , where TP , TN , FP , and FN represent the total number of pixels for true positive, true negative, false positive and false negative, respectively.

The quality measure (r^k) is normalized with a value of $0 \leq r^k \leq 1$. A higher value of r^k implies that a more compact cluster can be achieved by its corresponding skin color model. Figure 5 shows the results of detected skin-tone pixels for all possible combinations. An optimal combination of YCbCr + modified reference white ($r^k = 0.612$) is selected by the ASSM method.

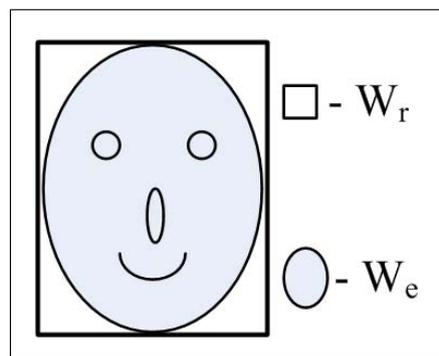


FIGURE 4. Estimation of quality measure by rectangular and elliptical windows

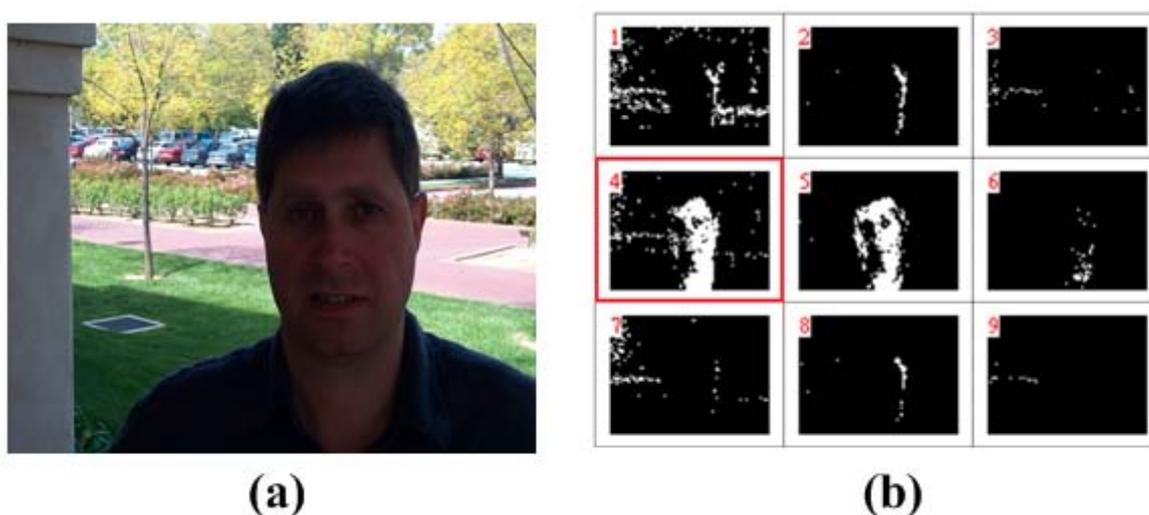


FIGURE 5. Optimal skin color model selected by the ASSM method. (a) Original image; (b) selected skin color model ($k = 4$) with the maximum quality measure ($r^k = 0.612$)

3. Proposed fast 8-connected component labeling method. To improve the performance of the conventional 8-connected component labeling method, which requires a large number of recursive procedures, a fast version that scans the whole image only once (top to bottom and left to right) is proposed to efficiently connect the skin-tone pixels detected by ASSM into a more compact cluster. The procedures for the proposed fast 8-connected component labeling method are illustrated in Figure 6, where the white pixels represent the detected skin color pixels. The procedures are basically a series of scan-and-merge processes, as described below.

1. Scan the first two rows and merge the adjacent pixels in the same row into a block. Then label them using consecutive numbers, as shown in Figure 6(a).
2. Merge the adjacent row blocks into a bigger one and label the lower merged block with the same number as that for its adjacent upper one, as shown in Figure 6(b).
3. Continue to scan the next row and label the new block using the next number, as shown in Figure 6(c).
4. Merge the new block and its adjacent upper block into a bigger one. Then label the lower merged block using the same number as that of its adjacent upper one, as shown in Figure 6(d). However, if another block (i.e., block #2) is next to the new

one (i.e., block #1), then merge them into a bigger one and label block #2 using the same number as that of block #1, as shown in Figure 6(e).

- Repeat the steps of scan-and-merge until the whole image is completely processed. The final results of block labeling are shown in Figure 6(f).

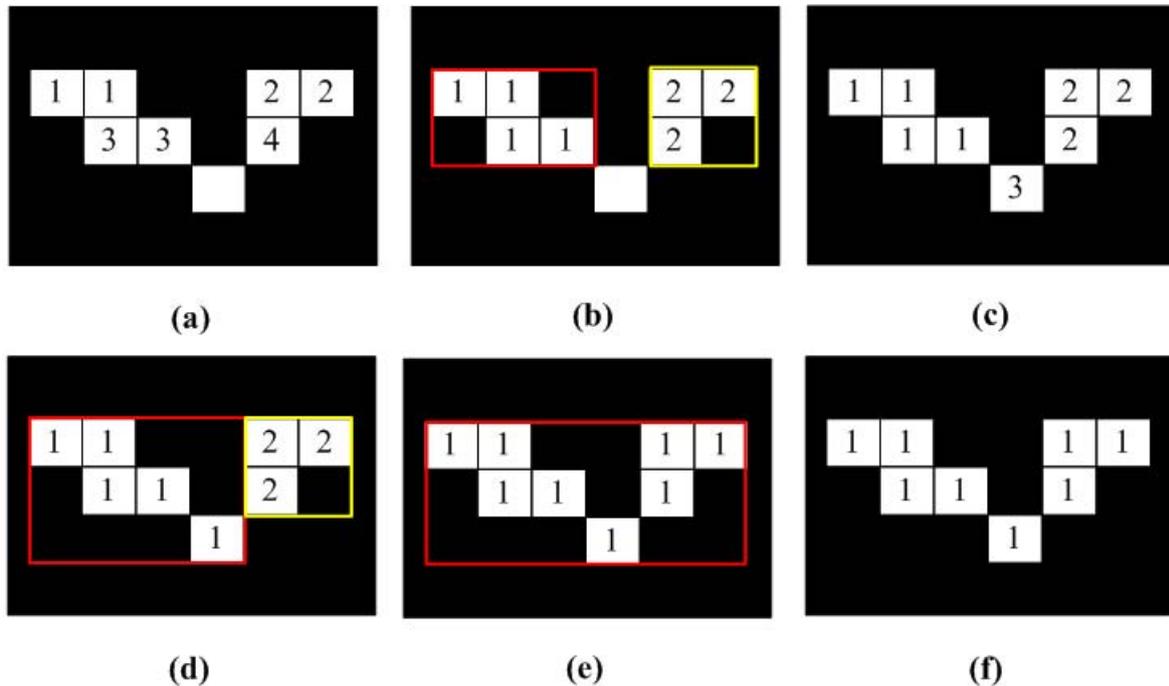


FIGURE 6. Procedures of the proposed fast 8-connected component labeling method

An artificial image created to illustrate the proposed method is shown in Figure 7(a). The results of block labeling for all connected clusters are shown in Figure 7(b). To evaluate the performance of the proposed method, four test images of size 320×240 pixels were used, as shown in Figure 8. A comparison of the proposed method with a conventional method was carried out. The runtimes required to complete 8-connected component labeling for the two methods are listed in Table 1. As indicated in the table, the proposed method is 37.5 to 73.2 times faster than the conventional one, which is highly dependent on the textures in the test images.

TABLE 1. Comparison of runtimes of 8-connected component labeling for the two methods

Test image	Runtime of conventional method (Method 1) (ms)	Runtime of proposed method (Method 2) (ms)	Method 1/Method 2
(a)	600	16	37.5
(b)	514	10	51.4
(c)	513	11	46.6
(d)	732	10	73.2

4. Experimental results and discussion. Experiments were performed on a computer with an Intel(R) Core(TM)2 Quad Q8200 2.33 GHz processor and 3.25GB of RAM. The

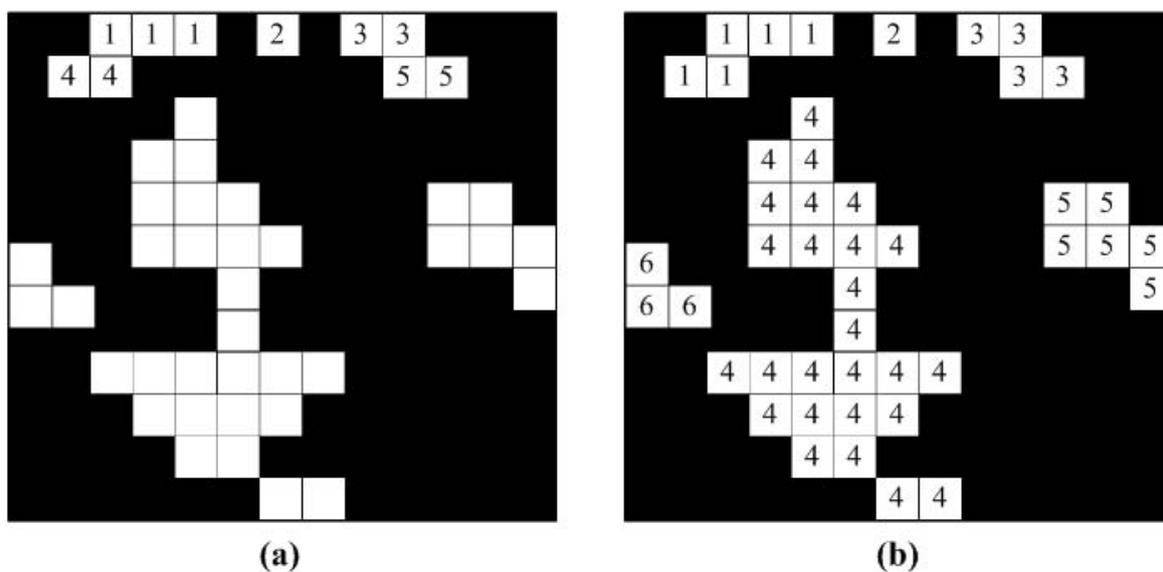


FIGURE 7. Illustration of the proposed fast 8-connected component labeling method. (a) Only the first two rows are scanned and labeled; (b) results of block labeling using the proposed method

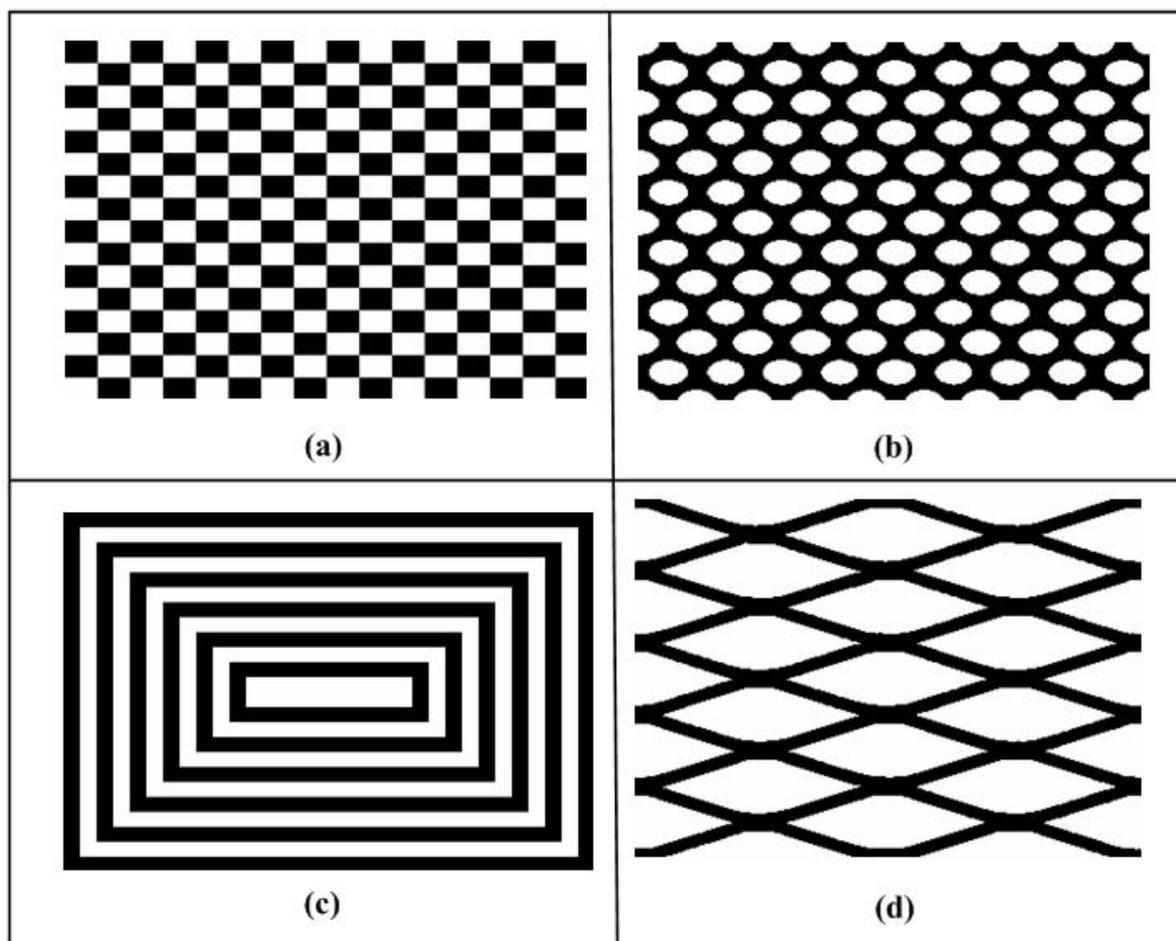


FIGURE 8. Four test images used to evaluate the performance of the proposed method

algorithm was implemented in BCB (Borland C++ Builder) 6.0. Figure 9 shows the typical results of detected skin clusters. Some of the most important features of a human face, i.e., eyes, may be lost due to severe lighting variations (see the red box with a size of $W \times H$). Therefore, to keep as many face features as possible, the width of the red box was increased by $0.5W$ on both sides, and the height of the red box is was increased by $0.5H$ on the top. The rule of thumb for choosing the box dimension is to keep most face features in the detected regions. Since the runtime of cascade AdaBoost is proportional to the size of detected regions, we chose $0.5W$ and $0.5H$, which is a compromise between performance and detected region size. Then, the detected possible face candidates (see the yellow box) are obtained and fed to the cascaded AdaBoost detector to further validate whether the possible candidate regions are human faces.

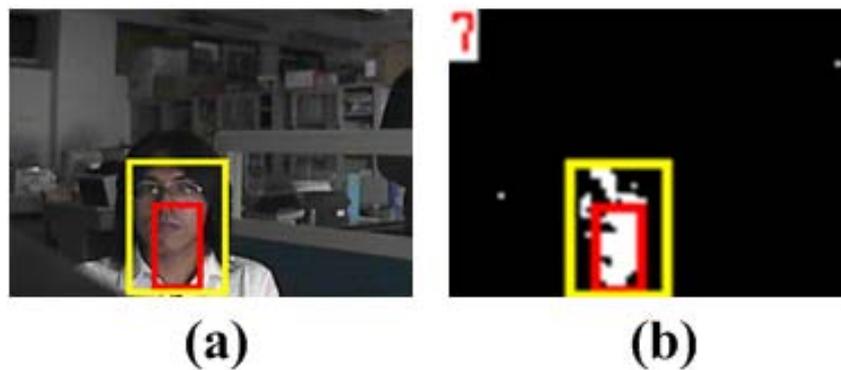


FIGURE 9. Results of face detection. (a) Original image; (b) results of detected skin cluster

In order to evaluate the robustness of the proposed method, a video sequence with varying illumination, including dim light, side light, and back light, was used, as shown in Figure 10. The results of quality measure for face detection with the application of ASSM are shown in Figure 11. The nine combinations of skin color models are denoted by M1 to M9, where the groups (M1, M2, M3), (M4, M5, M6), and (M7, M8, M9) use lighting compensation methods of reference white, modified reference white, and gray world assumption, respectively; the three models in each group use skin color models of YCbCr, the Gaussian mixture model, and Soriano's model, respectively. Since the values of quality measure for models M2, M3, M6, and M8 were too low, only the results of the remaining five models are shown in the figure. As shown in the figure, M1 and M7 were alternately used in the period of dim light, M7 dominates the period of side light (from left to right), and M5 is preferred in the back light situation. A single skin color model in this experiment thus could not accommodate all the lighting variations, from dim light to back light situations. In the period of dim light, M1, M4, and M7 were the three most frequently used models, with M4 being less important than M1 and M7. The selection of the lighting compensation method in this region is not very important because all three models use the same skin color model of YCbCr, which is luma-independent, as described earlier. This result illustrates that the influence of illumination in dim light can be greatly reduced by the introduction of the skin color model of YCbCr. A detection time of 60 ms for each frame was achieved with the aid of the ASSM method. A detection rate of 94.4% was obtained for the test video sequence.

5. Conclusions and future work. In this paper, an adaptive switching of skin color models (ASSM) with the cascaded AdaBoost method was proposed for face detection .



FIGURE 10. Face tracking sequence with varying illumination. (a) dim light; (b) side light; (c) back light

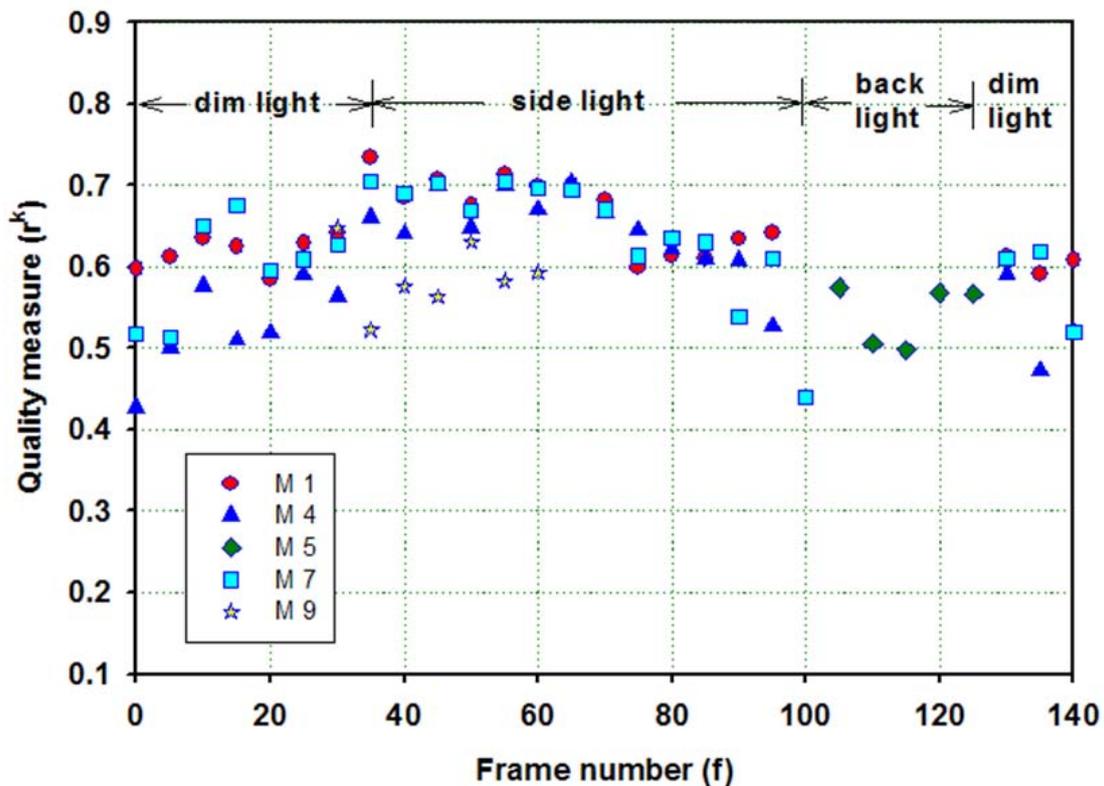


FIGURE 11. Variations of quality measure for face detection with the ASSM method

The proposed method can deal with large lighting variation, including dim light, side light, and back light scenarios. To increase the speed of clustering the skin-tone pixels detected by the ASSM method, a fast 8-connected component labeling method was proposed. The proposed method is 37.5 to 73.2 times faster than the conventional recursive method. A detection time of 60 ms for each frame was achieved with the aid of the ASSM method. A

detection rate of 94.4% was obtained for a test video sequence. However, the view angles of faces are constrained to within $\pm 30^\circ$ in the proposed system. Therefore, in a future study, we will focus on increasing the view angles.

Acknowledgment. This research was fully supported by the National Science Council of Taiwan under grant NSC-97-2221-E-212-035.

REFERENCES

- [1] S. Krinidis, and I. Pitas, Statistical Analysis of Human Facial Expressions, *Journal of Information Hiding and Multimedia Signal Processing (JIH-MSP)*, vol. 1, no. 3, pp. 241-260, 2010.
- [2] H. Sayoud, and S. Ouamour, Proposal of a New Confidence Parameter Estimating the Number of Speakers - An experimental investigation, *Journal of Information Hiding and Multimedia Signal Processing (JIH-MSP)*, vol. 1, no. 2, pp. 101-109, 2010.
- [3] R. Nishimura, S. Abe, N. Fujita, and Y. Suzuki, Reinforcement of VoIP Security with Multipath Routing and Secret Sharing Scheme, *Journal of Information Hiding and Multimedia Signal Processing (JIH-MSP)*, vol. 1, no. 3, pp. 204-219, 2010.
- [4] H. Ishii, M. Fukumi, and N. Akamatsu, Face detection based on skin color information in visual scenes by neural networks, *Proc. of the International Conference on Systems, Man, and Cybernetics*, Tokyo, Japan, vol. 5, pp. 350-355, 1999.
- [5] M. J. Jones and J. M. Rehg, Statistical color models with application to skin detection, *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, FortCollins, Colorado, USA, pp. 274-280, 1999.
- [6] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, Face detection in color image, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696-706, 2002.
- [7] 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, 1999.
- [8] D. Chai, and A. Bouzerdoun, A Bayesian approach to skin color classification in YCbCr color space, *Proc. of the IEEE Region 10 Conference (TENCON)*, Kuala Lumpur, Malaysia, vol. 2, pp. 421-424, 2000.
- [9] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, Skin detection in video under changing illumination conditions, *Proc. of the 15th International Conference on Pattern Recognition*, Barcelona, Spain, vol. 1, pp. 839-842, 2000.
- [10] L. M. Bergasa, M. Mazo, A. Gardel, M. A. Sotelo, and L. Boquete, Unsupervised and adaptive Gaussian skin-color model, *Image and Vision Computing*, vol. 18, no. 12, pp. 987-1003, 2000.
- [11] J. Yang and A. Waibel, A real-time face tracker, *Proc. of the IEEE Workshop on Applications of Computer Vision*, Sarasota, Florida, USA, pp. 142-147, 1996.
- [12] H. Stern, and B. Efron, Adaptive color space switching for tracking under varying illumination, *Image and Vision Computing*, vol. 23, no. 3, pp. 353-364, 2005.
- [13] C. Y. Chang and H. H. Chang, Adaptive color space switching based approach for face tracking, *Lecture Notes in Computer Science*, vol. 4233, pp.244-252, 2006.
- [14] J. Y. Xu, *Face detection and recognition technology research in complex background*, M.S. thesis, Shandong University of Technology, China, pp. 22-24, 2007.
- [15] E. Y. Lam, Combining gray world and Retinex theory for automatic white balance in digital photography, *Proc. of the Ninth International Symposium on Consumer Electronics*, Macau, China, pp.134-139, 2005.
- [16] P. Viola and M. J. Jones, Rapid object detection using a boosted Cascade of simple features, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, USA, vol. 1, pp. I511-I518, 2001.
- [17] W. Skarbek and A. Koschan, Color image segmentation - A survey, Technical report, Institute for Technical Informatics, Technical University of Berlin, October, 1994.
- [18] Center for Biological and Computational Learning at MIT (MIT-CBCL) Face Database #1. Available from: <http://cbcl.mit.edu/cbcl/software-datasets/FaceData2.html>