DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-176

Illumination Normalization and Skin Color Validation for Robust Face Detection

Sanghun Lee and Chulhee Lee, School of Electrical and Electronic Engineering, Yonsei Univ., Seoul, South Korea 03722

Abstract

The conventional Viola-Jones face detector may fail to detect faces under severe illumination conditions. The proposed method is based on the difference of Gaussian (DoG) that has been widely used to compensate for illumination effects. In the proposed method, we combine an original image and its DoG-filtered image, as a linear combination. This operation removed some illumination effects and improved face detection performance. Using the YCbCr color space, a skin color validation procedure was applied after face candidates were obtained using the proposed detector. Experiments using the Bao database showed that the proposed methods reduced over 50% of false positives.

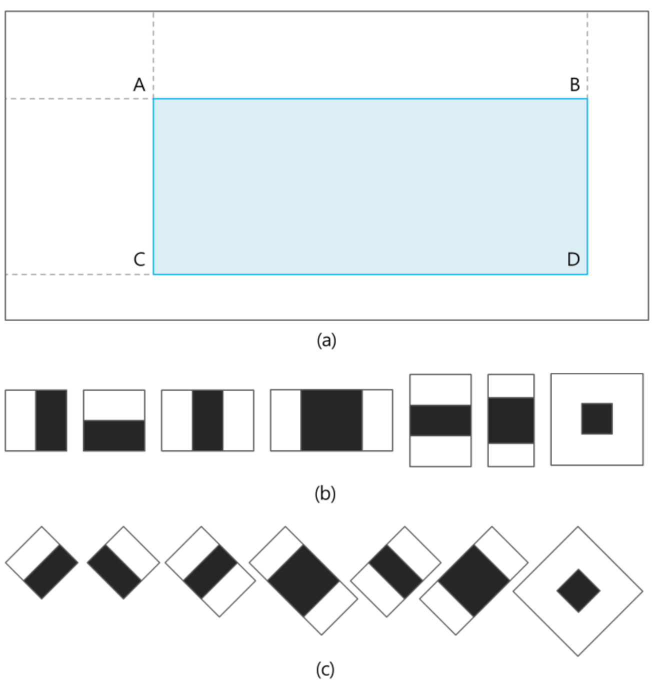

Figure 1. *(a) Sum of the intensities in the rectangle ABCD is calculated with the integral image. (b) Haar-like features with three different types: edge type, line type, and center-surround type. (c) Rotated Haar-like features.*

Introduction

Face detection is a fundamental technique [1, 2] for the systems based on recognizing human faces, such as face-aware focusing of digital cameras, automatic face tagging, surveillance systems, etc. However, face detection is not an easy task since face images often appear differently in terms of a head pose, illumination, expression, illumination conditions, occlusions, etc. One of the technical challenges for face detection is that variability in lighting can cause drastic changes and unpredictable illumination effects (e.g. shadow regions).

A widely-used method for face detection is the Viola-Jones face detector [3, 4] that utilizes integral image representation, Haar-like features, AdaBoost learning and attentive cascade structures. The integral image, which is also known as summed-area table, is a very efficient way to compute the sum of the intensities of a specific region, which was first introduced in [5]. The integral image can be computed as follows:

$$ii(x,y) = \sum_{x' \le x,\, y' \le y} I(x',y') \tag{1}$$

where $I(x,y)$ and $ii(x,y)$ denote an original image and its integral image, respectively. As shown in Fig. 1(a), the sum of the intensities of a specific region can be computed as follows:

$$\sum_{\substack{x_1 < x \le x_2 \\ y_1 < y \le y_2}} I(x,y) = ii(D) - ii(C) - ii(B) + ii(A) \tag{2}$$

where $A=(x_1,y_1)$, $B=(x_2,y_1)$, $C=(x_1,y_2)$ and $D=(x_2,y_2)$. Based on the Haar wavelet, Viola and Jones introduced Haar-like features that can be extracted from integral images (Fig. 1(b)). For adjacent sub-regions of specific image points, the average intensity value of each sub-region can be calculated. Then, the difference between the averages can be used as a feature of the point. Lienhart *et al*. extended the set of Haar-like features by introducing rotated Haar-like features [6] as shown in Fig. 1(c).

The AdaBoost (Adaptive Boosting) [7, 8] is a machine learning algorithm that selects only those features that improve the predictive power while reducing the dimensions of the feature vectors. Since there are over 160,000 Haar-like features in the Viola-Jones face detector, reducing the number of features is necessary.

Finally, the attentional cascade structure is an important component of the Viola-Jones face detector. The main idea of this structure is to preserve most of the positive samples when training weak classifiers while rejecting negative samples. Since most of the negative samples are rejected in early stages, face detection can be done in real-time.

However, the Viola-Jones face detector does not guarantee good performance when illumination conditions are bad. To tackle this problem, we combined illumination normalized images with the original image. Also, skin color validation was used to reject false positives.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-176

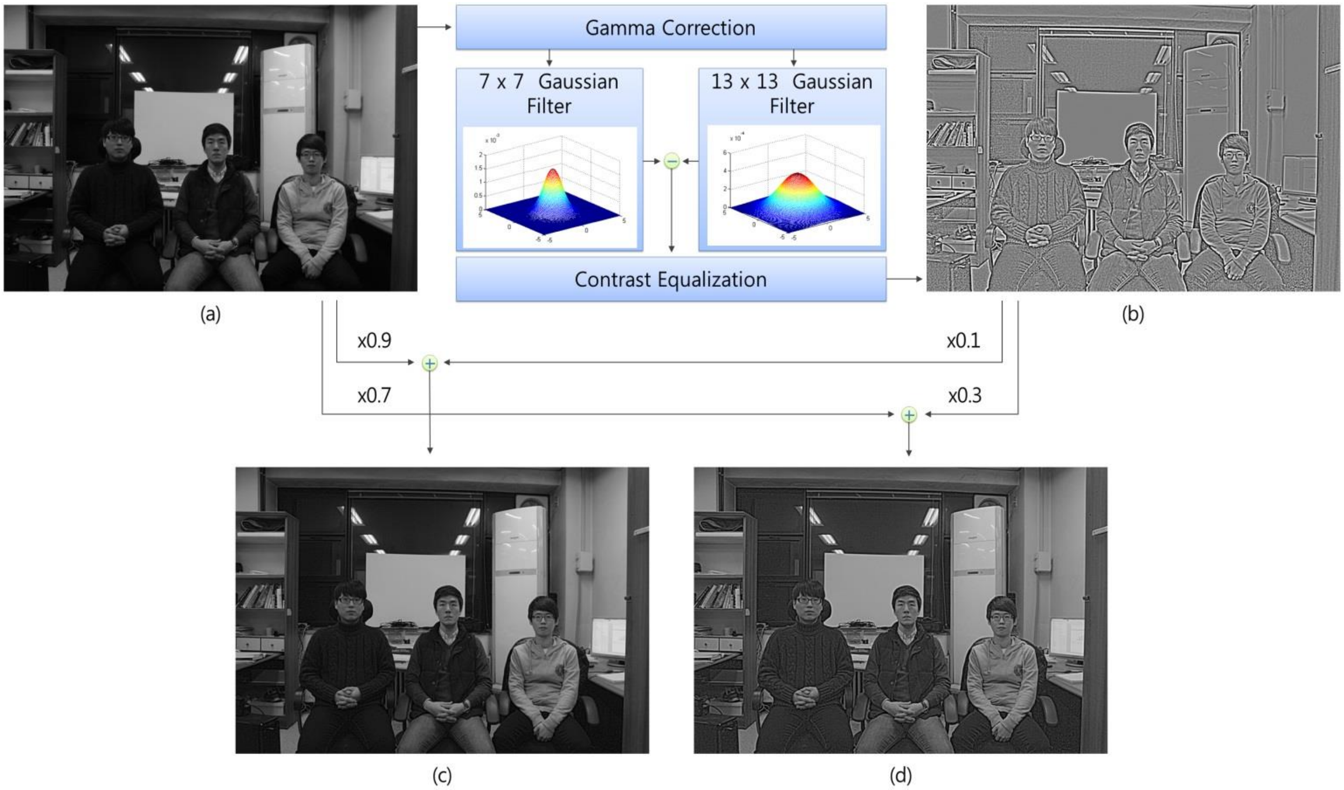

Figure 2. *(a) An input image. (b) Tan and Triggs' illumination normalization procedure comprises gamma correction, difference of Gaussian and contras equalization. (c,d) Examples of DoG blending.*

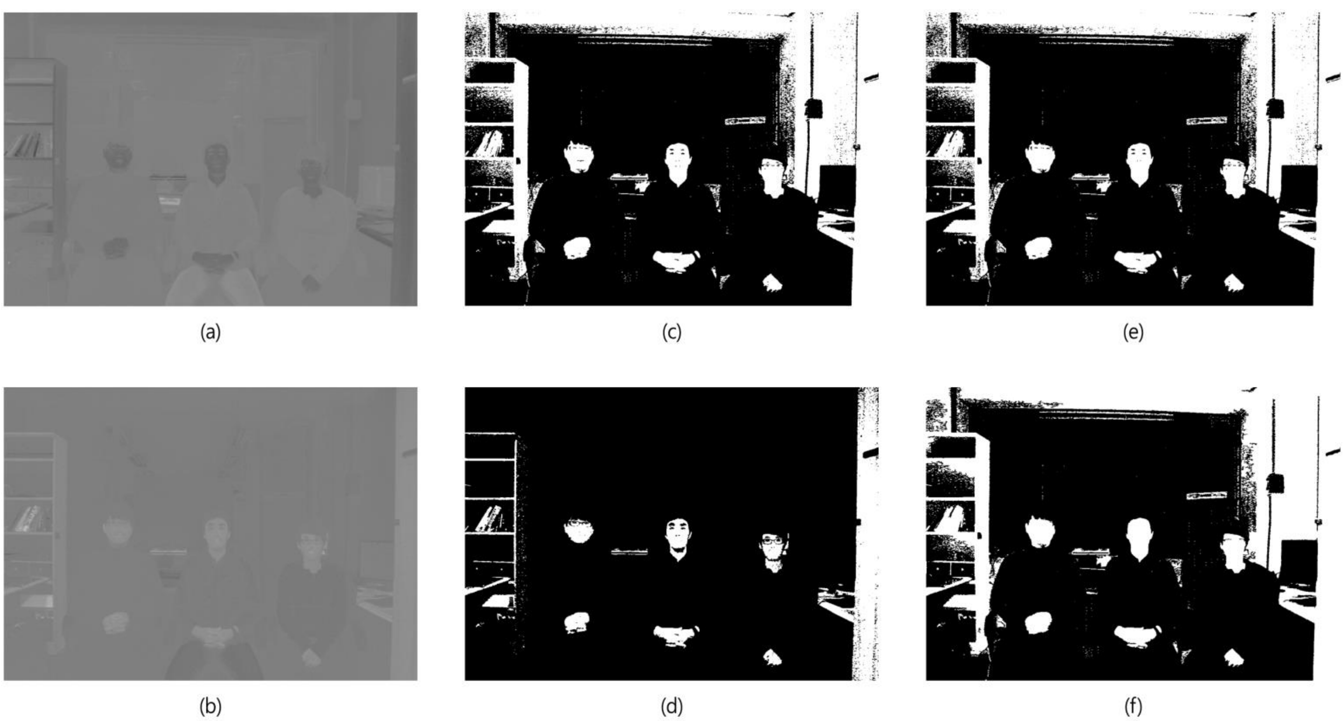

Figure 3. (*a) Cb image, (b) Cr image and (c,d) the threshold result of (a) and (b) according to (7), (e) skin color estimation using (c) and (d), (f) small holes in (e) can be removed using a morphological closing operation.*

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-176

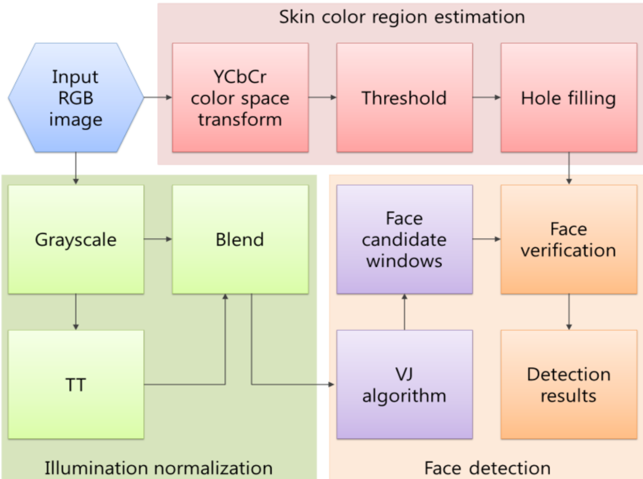

Figure 4. *Flow chart of the proposed face detection procedure.*

Proposed Method

Fig. 4 is a flowchart of the proposed face detection procedure. There are three main steps: illumination normalization, skin color region estimation and face detection.

First, we used the preprocessing chain of [9, 10] proposed by Tan and Triggs to reduce illumination effects, which is based on the difference of Gaussian (DoG). The preprocessing chain was used for illumination invariant face recognition in their original paper. We observed that a combination of an original image and illumination normalized images (Figs. 2(c,d)) improved face detection performance. Before the DoG was applied, an input image $I(x,y)$ was gamma-corrected in the preprocessing chain as follows:

$$GC(x,y) = I(x,y)^{\gamma} \tag{3}$$

where $I(x,y)$ represents $[0,1]$ and $GC(x,y)$ represents the gamma-corrected image. We used $\gamma = 0.2$ as recommended in [9, 10]. Then the DoG filter was applied to $GC(x,y)$ by using the 2D convolution with $\sigma_0 = 1.0$ and $\sigma_1 = 2.0$ for 7×7 and 13×13 Gaussian kernels, respectively. The DoG-filtered image $DoG(x,y)$ was further normalized using the following contrast equalization steps:

$$\begin{aligned} DoG'(x,y) &\leftarrow \frac{DoG(x,y)}{\operatorname{mean}\left(\left|DoG(x,y)\right|^{a}\right)^{1/a}} \\ DoG''(x,y) &\leftarrow \frac{DoG'(x,y)}{\operatorname{mean}\left(\min\left(\tau,\left|DoG'(x,y)\right|^{a}\right)\right)^{1/a}}. \end{aligned} \tag{4}$$

All the parameter settings in the preprocessing chain were the same as those used in [9, 10]. Finally, the illumination normalized image was blended with the input image as follows:

$$I(x,y) \leftarrow wI(x,y) + (1-w)DoG''(x,y) \tag{5}$$

where w represents a weight constant in $[0,1]$. Henceforth, we call this combination "DoG blending." Figs. 2(c,d) show some example images when w was 0.1 and 0.3, respectively. Some shadows cast on faces appear slightly flattened. After each image was processed with the preprocessing chain and the DoG blending approach, the OpenCV's Viola-Jones face detector was directly applied to $I(x,y)$ and face candidate regions were obtained. The OpenCV's Viola-Jones face detector provides three different pre-trained face detectors: *haarcascade frontalface alt*, *haarcascade frontalface alt2* and *haarcascade frontalface default*. In those detectors, the original Haar-like features and rotated Haar-like features were used. These features are described in [6, 11].

Second, false positives in face detection results can be reduced by the skin color estimation. There are several approaches that use skin color models based on the observation of skin color samples [12-15]. For skin color estimation, we used the YCbCr color space that was converted from a given RGB image as follows:

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & 74.203 & 112.0 \\ 112.0 & 93.786 & 18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{6}$$

The skin regions were estimated by using threshold operations on the C_b and C_r color channels [15] as follows:

$$\begin{aligned} r_{C_b}(x,y) &= \begin{cases} 1 & 82 < C_b(x,y) < 122 \\ 0 & otherwise \end{cases} \\ r_{C_r}(x,y) &= \begin{cases} 1 & 138 < C_r(x,y) < 168 \\ 0 & otherwise \end{cases} \\ \operatorname{skin}(x,y) &= \begin{cases} 1 & r_{C_b}(x,y) = 1 \ and \ r_{C_r}(x,y) = 1 \\ 0 & otherwise \end{cases} \end{aligned} \tag{7}$$

where $\operatorname{skin}(x,y)$ denotes the skin color region. The constants used in (7) were empirically obtained. Since the eyes and nostrils produced small holes in $\operatorname{skin}(x,y)$, we used morphological close operations for hole filling (Fig. 3). Then, for each face candidate, we made final classifications based on the proportions between the skin regions and the corresponding areas of face candidate regions V_i as follows:

$$\frac{C_i}{W_i \times H_i} \geq \mu \tag{8}$$

where

$$C_i = \sum_{(x,y)\in V_i} \operatorname{skin}(x,y), \tag{9}$$

W_i and H_i represent the width and the height of the i th candidate window, respectively [16]. Here we empirically set the constant μ to 0.2.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-176

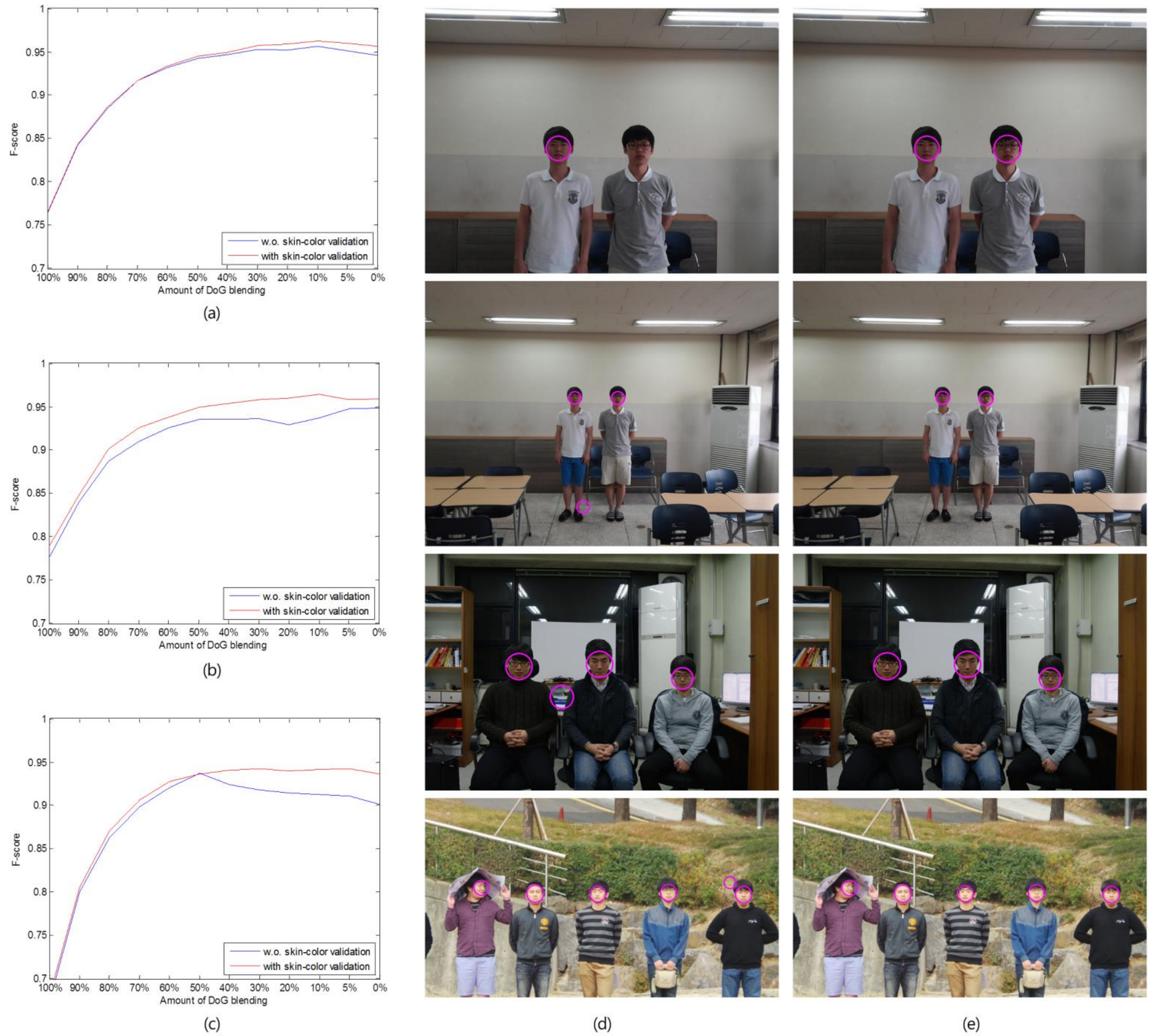

Figure 5. *(a-c) Face detection results (F-score) according to the amount of DoG blending and skin color validation when Opencv classifier #1 (alt), #2 (alt2) and #3 (default) were used as face detectors, respectively. (d) Some examples of face detection results with Opencv classifier #2 (alt2). (e) Some examples of face detection results with the proposed method. The false negatives (example images in first two rows) and the false positives (example images in last two rows) were correctly detected with the proposed method.*

Experiments

To evaluate the proposed method, we conducted experiments on the Bao color face database [17, 18]. The database comprises color images that contain multiple faces of various image sizes ($341 \times 501 \sim 1531 \times 913$). We used the first 100 images of the entire dataset containing 865 frontal faces.

We compared the proposed method with the original OpenCV's Viola-Jones face detectors (three different versions): *haarcascade frontalface alt*, *haarcascade frontalface alt2* and *haarcascade frontalface default*. In the proposed method, the input images were processed with the proposed DoG blending approach. Then, the OpenCV's Viola-Jones face detectors were applied to the pre-processed images. Finally, the detection results were evaluated with the skin color validation method [15, 16]. For a quantitative comparison, we used the F-score, which is the harmonic mean of precision and recall as follows:

$$F = \frac{2 \times \left(precision + recall \right)}{precision \cdot recall} = \frac{2TP}{2TP + FP + FN} \tag{9}$$

where TP, FP and FN denote the number of true positives, false positives and false negatives, respectively. The face detection results obtained by three different versions of Viola-Jones face

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-176

detectors are shown in Figs. 5(a-c). With skin color verification, the F-score is improved in most cases. Also, DoG blending slightly improved the F-score when $w = 0.05 \sim 0.1$, since illumination conditions were not so severe in the face images of the database. Note that the pre-trained classifiers were not based on the proposed DoG blended images.

Table 1. Face detection results with various classifiers (#1: haarcascade_frontalface_alt, #2: haarcascade_frontalface_alt2, #3: haarcascade_frontalface_default)

Classifier	Verif.	w	TP	FP	FN	F-score
#1	No	0	813	41	52	0.946
#1	No	0.1	815	24	50	0.957
#1	Yes	0	813	22	52	0.957
#1	Yes	0.1	815	13	50	0.963
#2	No	0	818	42	47	0.948
#2	No	0.1	790	31	75	0.937
#2	Yes	0	818	22	47	0.960
#2	Yes	0.1	821	16	44	0.964
#3	No	0	826	141	39	0.902
#3	No	0.05	828	125	37	0.911
#3	Yes	0	826	73	39	0.937
#3	Yes	0.05	828	64	37	0.943

Table 1 shows that using the skin color validation significantly reduces the false positives. Note that "Verif." indicates whether the skin color verification was used or not. Also, #1, #2 and #3 denote the pre-trained face detector *haarcascade frontalface alt*, *haarcascade frontalface alt2* and *haarcascade frontalface default*, respectively. Compared to the original method, about 50% of false positives were reduced in all the cases. Also, DoG blending further reduced the false positive rates and correctly found more faces. Figs. 5(d,e) show that some errors of the existing methods were correctly detected with the proposed method. Some examples of misclassified and missed results of the proposed method are shown in Fig. 6.

Conclusion

By using the combined results of original images and DoG-filtered images, we are able to improve recognition performance by using the Viola-Jones detection method. Also, the skin color validation procedure further eliminated non-facial candidates. Experiments using the Bao database showed that the proposed method showed improved performance, compared to the original Viola-Jones detection method. The proposed method can be applied to other color face images and Viola-Jones face detectors without any additional classifier training.

Figure 6. *Some examples of misclassified or missed results.*

References

[1] C. Zhang and Z. Zhang, "A survey of recent advances in face detection," Tech. rep., Microsoft Research2010.

[2] E. Hjelmås and B. K. Low, "Face detection: A survey," *Computer vision and image understanding,* vol. 83, pp. 236-274, 2001.

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. I-511-I-518 vol. 1.

[4] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision,* vol. 57, pp. 137-154, 2004.

[5] F. C. Crow, "Summed-area tables for texture mapping," *ACM SIGGRAPH computer graphics,* vol. 18, pp. 207-212, 1984.

[6] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, 2002, pp. I-900-I-903 vol. 1.

[7] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational learning theory*, 1995, pp. 23-37.

[8] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-176

boosting," *Journal of computer and system sciences,* vol. 55, pp. 119-139, 1997.
[9] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Analysis and Modeling of Faces and Gestures*, ed: Springer, 2007, pp. 168-182.
[10] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on,* vol. 19, pp. 1635-1650, 2010.
[11] A. Kuranov, R. Lienhart, and V. Pisarevsky, "An empirical analysis of boosting algorithms for rapid objects with an extended set of haar-like features," *MRLPTRP02P01,* 2002.
[12] N. Bojic and K. K. Pang, "Adaptive skin segmentation for head and shoulder video sequences," in *Visual Communications and Image Processing 2000*, 2000, pp. 704-711.
[13] D. Chai and K. N. Ngan, "Face segmentation using skin-color map in videophone applications," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 9, pp. 551-564, 1999.
[14] K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking," *Signal processing: Image communication,* vol. 12, pp. 263-281, 1998.
[15] S. L. Phung, A. Bouzerdoum, and D. Chai, "A novel skin color model in ycbcr color space and its application to human face detection," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, 2002, pp. I-289-I-292 vol. 1.
[16] C. Erdem, S. Ulukaya, A. Karaali, and A. T. Erdem, "Combining Haar feature and skin color based classifiers for face detection," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 1497-1500.
[17] R. Frischholz, "Bao face database at the face detection homepage," ed, 2012.
[18] X. Wang, H. Xu, H. Wang, and H. Li, "Robust real-time face detection with skin color detection and the modified census transform," in *Information and Automation, 2008. ICIA 2008. International Conference on*, 2008, pp. 590-595.

Author Biography

Sang-hun LEE received a B.S. degree in electrical and electronical engineering from Yonsei University, Seoul, South Korea, in 2013. He is currently working toward a Ph.D. degree in electronic engineering at Yonsei University. His current research interests include image processing and pattern recognition.

Chulhee Lee received the B.S. and M.S degrees in electronics engineering from Seoul National University in 1984 and 1986, respectively, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, Indiana, in 1992. From 1986 to 1987, he was a researcher in the Acoustic Laboratory at Technical University of Denmark (DTH). From 1993 to 1996, he worked with National Institutes of Health, Bethesda, Maryland. In 1996, he joined the faculty of the Department of Electrical and Computer Engineering, Yonsei University, Seoul, Korea. His research interests include image/signal processing, pattern recognition, and neural networks.