FACE DETECTION USING THE 3×3 BLOCK RANK PATTERNS OF GRADIENT MAGNITUDE IMAGES

Kang-Seo Park, Young-Gon Kim, and Rae-Hong Park

Department of Electronic Engineering, School of Engineering, Sogang University, 35 Baekbeom-ro (Sinsu-dong), Mapo-gu, Seoul 121-742, Korea

ABSTRACT

Face detection locates faces prior to various face-related applications. The objective of face detection is to determine whether or not there are any faces in an image and, if any, the location of each face is detected. Face detection in real images is challenging due to large variability of illumination and face appearances. This paper proposes a face detection algorithm using the 3×3 block rank patterns of gradient magnitude images and a geometrical face model. First, the illumination-corrected image of the face region is obtained using the brightness plane that is produced using the locally minimum brightness of each block. Next, the illumination-corrected image is histogram equalized, the face region is divided into nine (3×3) blocks, and two directional (horizontal and vertical) gradient magnitude images are computed, from which the 3×3 block rank patterns are obtained. For face detection, using the FERET and GT databases three types of the 3×3 block rank patterns are a priori determined as templates based on the distribution of the sum of the gradient magnitudes of each block in the face candidate region that is also composed of nine (3×3) blocks. The 3×3 block rank patterns roughly classify whether the detected face candidate region contains a face or not. Finally, facial features are detected and used to validate the face model. The face candidate is validated as a face if it is matched with the geometrical face model. The proposed algorithm is tested on the Caltech database images and real images. Experimental results with a number of test images show the effectiveness of the proposed algorithm.

1. INTRODUCTION

Detecting human faces from an image is a key problem in various face-related applications such as face recognition, facial expression recognition, face tracking, video surveillance, and human-computer interaction. Face detection is a preprocessing that locates faces prior to various face-related applications. The objective of face detection is to determine whether or not there are any faces in an image and, if any, the location of each face is detected. Face detection in real images is challenging due to large variability of illumination, face appearances, and complex background. The complex background leads to false detection in which a non-face region is incorrectly determined as a face or vice versa. The variability of face appearances is caused by the diversity of individuals, illumination, face pose, facial expression, occlusion, race, and so on [26, 19].

There are two approaches to face detection in an image. Most face detection algorithms are based on example-based learning that discriminates face or non-face images. Gaussian clusters are used to model face and non-face models according to the distance from the mean sample, and then a given window is classified as a face or non-face [21]. Rowley *et al*. proposed a neural network (NN)-based algorithm, in which a number of samples (face or non-face) are trained using an NN

[18]. For NN training and testing, different types of features (gradient, Gabor, and two-dimensional (2-D) wavelet features) were presented, instead of intensity-based features [6,7,8], which improved Rowley *et al.*'s algorithm for images with illumination effects. Kaushal and Raina also proposed the face detection method in which the feature vector based on Gabor filters was used as the input of classifier [10]. Viola and Jones proposed a strong classifier that uses the adaboost algorithm, in which a strong classifier consists of a number of weak classifiers that employ Haar-like features [22]. Haar-like features are efficiently computed with a rectangular integral image. They proposed a cascade structure to improve both the speed and detection rate. Also skin color information was used to extract face region with a Bayesian decision framework [13].

Other face detection approaches use template matching. A face in an image is determined by comparing the template model pattern and the window pattern obtained from a test image. If the window pattern is close to the template model stored in a system, then the window is determined as a face. Correlation templates were used, in which a set of different fixed-size windows was used as templates to detect facial features in an image [3]. Instead of using several fixed-size templates, the deformable (non-rigid) template is used [27], the size of which is changed to detect a face of different size. Jeng *et al*. proposed the geometrical face model that uses quantitative and relative information of facial components [9]. Their algorithm is improved by combining the geometrical face model with the skin color information, where the local patterns are used as templates for face detection [5]. Ahonen *et al*. proposed the local binary patterns that are insensitive to illumination effects [1]. Their algorithm was improved by the local successive mean quantized transform (SMQT), in which local SMQT patterns quantize the intensity level of local patterns and thus are insensitive to illumination effects [15]. Also, for robustness to illumination effects, a 3×3 block model using the wavelet coefficient values was used to detect a face [23]. Skin color is also used to detect face. To construct a template of accurate face region, Kherchaoui and Houacine proposed a template matching algorithm using a model of skin color with constraints [11].

Illumination is to be considered for robust face detection. There are two factors in illumination that determine face appearance and thus the success of face detection. First, illumination is often changed in real images by different lighting conditions such as indoor lighting, outdoor daylight, weather condition, and so on. Sometimes illumination during image acquisition is too bright or too dark. Due to different lighting conditions that produce dark shadows and highlights, face appearance varies. Second, face appearance varies according to illumination directions of light sources. Even with the same light sources, face appearance is changed by shadows (on the part of a face) that are caused by different directions of light sources, which prevents correct face detection. Finally, illumination may change the chromaticity, thus changing skin color to other colors. The performance of the face detection method using skin color is significantly degraded if the white balance condition is not satisfied.

This paper is an extended version of the paper [16], in which cascade structure based on block rank pattern is newly constructed and analysis of the detection rate in CMU PIE database [20] and color FERET database [17] is supplemented.

This paper proposes a face detection algorithm using a cascade structure, which is constructed from a large number of face and non-face samples and a geometrical face model [9]. First, all kinds of multiple local windows are classified as candidate face and non-face regions by the constructed cascade structure. The newly constructed cascade structure is generated from a lot of training face and non-face images. This processing gives the reference of classifying the face and non-face images. Next, the geometrical face model template is used to validate the correct face

model. The geometrical face model uses the information of the geometrical relationships between facial components, such as the distance and size. Face images classified by the 3×3 block rank patterns are matched with the geometrical face model template. The 3×3 block rank patterns roughly classify face images, and then, the geometrical face model finally validates them. As the geometrical face model validates the face images, the face images are correctly detected. Also, the proposed method uses the gradient map that represents the contours of an object. Contours are robust to illumination in the sense that they represent the shape of objects. Therefore, the proposed method is robust to illumination.

The rest of the paper is structured as follows. Section 2 proposes a face detection algorithm using the 3×3 block rank patterns of gradient images, in which scanning window, illumination correction, gradient magnitude images, block rank patterns, block pattern classification, and geometrical face model are presented. Section 3 describes experimental results and discussions. Finally, Section 4 gives conclusions.

2. PROPOSED FACE DETECTION ALGORITHM

Figure 1 shows the block diagram of the proposed face detection algorithm, which consists of five parts. At first, a whole input image is scanned by multiple local windows, because multiple faces of various sizes may exist at different positions in an input image. The scanning local window S is to be determined whether it contains a face or not. For more precise classification of the region covered by S as a face or non-face, the illumination correction part compensates for the non-uniform illumination which is caused by various factors such as the direction of light sources. The illumination-corrected image S' is divided into nine (3×3) blocks. Thus, in this paper, the size (height H and width W) of S is assumed to be a multiple of three. In the gradient magnitude image part, the scanning window S of the illumination-corrected image S' (face candidate region) is used to compute two gradient magnitude images depending on the direction of gradient computation. Next, in the block rank pattern part, the block rank patterns are generated from the gradient magnitude images G ($G=\{G^h, G^v\}$, where the superscripts h and v denote horizontal and vertical directions, respectively). The ranks in the block rank pattern are obtained based on the
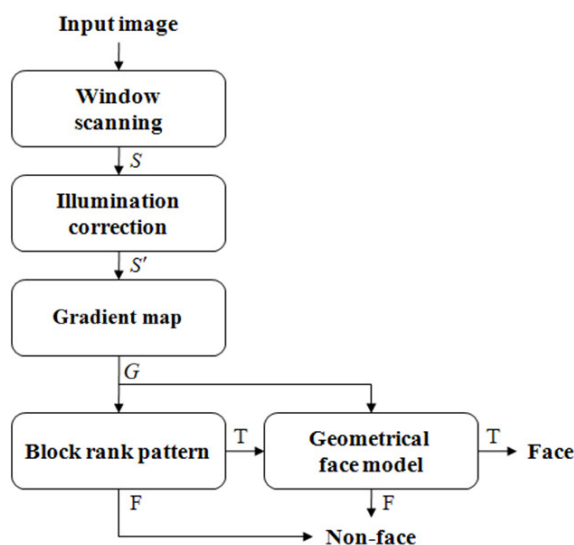

Figure 1. Block diagram of the proposed face detection algorithm.

distribution of the sum of gradient magnitudes in a block, in which the rank of a block is represented in descending order based on the sum of gradient magnitudes of all the pixels in a block (i.e., rank 1 for the largest sum of gradient magnitudes whereas rank 9 for the smallest). The rank of each block denotes the information corresponding to facial components that exist in the block. Ranks of nine blocks generate three types of 3×3 block rank pattern P ($P=\{P^h, P^v\}$). In the block pattern classification part, by comparing the block rank patterns P with the template block rank patterns T ($T=\{T^h, T^v\}$) the scanning window S with the block rank patterns P is classified as a face candidate or non-face. The template block rank patterns T are reference patterns which are obtained by training a large number of face image data. Finally, in the geometrical face model part, the face candidate is definitely validated as a face if it is matched with the geometrical face model. The faces are detected by the geometrical face model, in which the relative sizes of the facial components and the relative distances between face components are utilized [5].

2.1. Scanning Window

A whole input image is scanned by multiple local windows that have various sizes, and then, the sub-image covered by each of scanning windows is classified into a candidate face image or non-face image. There exist multiple faces in an input image. But the scanning process requires a high computational load, if the local window scans pixel by pixel and if the scale of a local window changes by one pixel unit.

To reduce a high computational load, many techniques were considered. Rowley *et al*. proposed a method in which a whole input image is scanned using hierarchical windows [18]. The multiple scaled pyramid images are generated from an input image, and then the multiple pyramid images are scanned using local windows of fixed scale. The genetic algorithm was used to find face location, by searching for the random locations with the reduced computational load [14]. However, it also reduces the detection rate. There is a trade-off between the detection rate and computational load.

2.2. Illumination Correction

Non-uniform illumination modifies face appearance in an image. It should be corrected for reliable face detection. We use Waring and Xiuwen's algorithm to correct the illumination effect [25]. The scanning window S, which is a local window scanned over an input image, is equally divided into nine (3×3) blocks. Then, the minimal brightness plane is estimated by minimum intensity values that are obtained from each of nine blocks. The minimal brightness plane is resized by bilinear interpolation to the image with the size equal to that of the scanning window S. For illumination correction, the minimal brightness plane is subtracted from the scanning window, which can reduce the distortion caused by the different direction of light sources.

Figure 2 shows five examples of illumination correction. Figure 2(a) shows five original face images (60×60) with illumination effect and Figure 2(b) shows the minimal brightness plane estimated from Figure 2(a). Figure 2(c) shows illumination-corrected images by intensity subtraction. Finally, the illumination-corrected images are enhanced by histogram equalization, as shown in Figure 2(d). The histogram equalization enhances the illumination of different lighting conditions. This illumination correction step compensates for the distortion due to changes in illumination and differences in camera responses.

Figure 2. Five examples of illumination correction. (a) original face images (60×60), (b) minimal brightness planes, (c) illumination-corrected images by subtracting (b) from (a), (d) images enhanced by histogram equalization.

2.3. Gradient Magnitude Images

Edge detection is used to extract facial features for reliable face detection. Edge features clearly represent the characterizing shapes that consist of contours. For frontal or nearly frontal face images, despite the variation of face, the face contour is approximately an oval shape and facial components such as eye, mouth, and nose are approximately described by horizontal and vertical edges. The intensity-based features do not explicitly represent the shape of a face as visually as edge features do [7]. Edges, which represent the shape as lines or contours, are more suitable for describing illumination changes than intensity-based features [21].

In the proposed algorithm, two directional (horizontal and vertical) gradient magnitude images are used, each of which well represents face contours and geometrical shape information. As the face and facial component (eye, eyebrow, nose, philtrum, mouth, and chin) shapes consist of straight lines as well as curved lines such as oval shape, the two directional gradient magnitude images are useful for efficient description of facial features.

The gradient images are computed using Sobel masks [4]. Sobel masks are computationally simple and widely used for computation of gradient images, and then edge detection. Figure 3 shows examples of gradient images. Figure 3(a) shows a 60×60 original face image. Figures 3(b) and 3(c) show the gradient magnitude images of horizontal and vertical directions, respectively. Intensity levels in Figures 3(b) and 3(c) are reversed for better display, i.e., dark area represents large gradient magnitudes. Figure 3(b) primarily shows eye, eyebrow, mouth, and nostrils that are described well by horizontal lines. Figure 3(c) mainly represents nose and philtrum that are described by vertical lines.
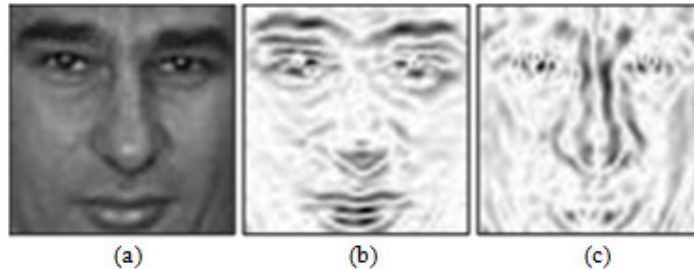

Figure 3. Original image and its gradient magnitude images (60×60, two directions). (a) original image. Gradient magnitude images of the (b) horizontal direction, and (c) vertical direction.

2.4. Block Rank Patterns and Block Pattern Classification

A face is roughly detected by three types of 3×3 block rank patterns generated from the gradient magnitude images *G*. For generating the block rank patterns, two gradient magnitude images are divided into nine (3×3) blocks. As mentioned, this paper is an extended version of previous algorithm [16] where the four gradient magnitude images corresponding to four directions (i.e., horizontal, vertical, diagonal, and anti-diagonal) are used. Although the diagonal and anti-diagonal directions are orthogonal to each other, sums of the diagonal and anti-diagonal gradient magnitudes are similar for each of facial components. Thus, sums of gradient magnitude images of both directions (combined diagonal/anti-diagonal) were combined by adding them [16]. However, in this paper, only two directions (i.e., horizontal and vertical) are used for block rank patterns to classify face or non-face candidate region. For classifying face and non-face candidate regions, we experiment with FERET and GT databases by varying the number of directions and rank patterns. It is observed through experiments that detecting three facial features using two directions (i.e., eyes and mouth features using horizontal block rank patterns whereas nose feature using vertical block rank pattern) is more effective than detecting facial features using four directions [16] (i.e., eyes+mouth features using horizontal block rank patterns and nose feature using vertical block rank pattern, and eyes+nose+mouth features using combined diagonal block rank pattern) for effectively eliminating false positives with true positive well preserved.

Each of nine blocks represents rough location of facial components in a face, where facial components are located in the specified blocks. Face candidate region is roughly separated into 3×3 blocks. As shown in Figure 4(a), left and right eyes are located in blocks (1, 1) and (1, 3), nose is located in blocks (1, 2) and (2, 2), and mouth is located in block (3, 2), where block (*i*, *j*), 1≤*i*, *j*≤3, denotes the block located in *i*th row and *j*th column of the 3×3 block pattern (face candidate region). Figures 4(b) and 4(c) show gradient magnitude images of the horizontal and vertical directions, respectively.
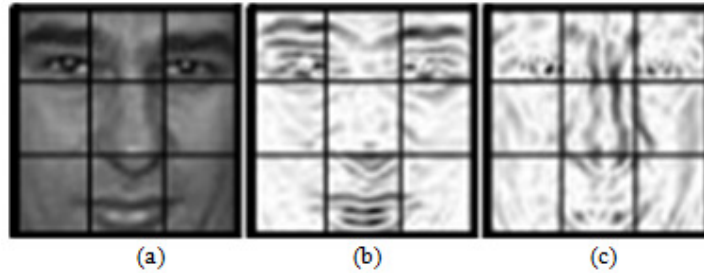
Figure 4. Block original image and its gradient magnitude images (60×60, two directions). (a) block original image. Gradient magnitude images of the (b) horizontal direction, and (c) vertical direction.

The nine blocks are used to find rough location of the facial components in the face candidate region. In the gradient magnitude images of the frontal face, the sum of gradient magnitudes is computed for all pixels in each block, because it is large in the block where the corresponding facial feature is located. For example, in the horizontal gradient magnitude image, the sum of gradient magnitudes is large in blocks (1, 1), (1, 3), and (3, 2) (eyes and mouth), the shapes of which mainly consist of horizontal lines. The sum of gradient magnitudes in the vertical gradient magnitude image is large in blocks (1, 2) and (2, 2) (nose), the shape of which consists of vertical lines.

Different block patterns such as 4×4 or 5×5 block patterns in a scanning window can be used, however each block in these block patterns does not correspond well to the specific facial features (such as eyes, mouth, and nose). Note that the appropriate number of blocks in scanning window is related to the sizes and locations of facial components, which is supported by experiments. Therefore, we use 3×3 block patterns. Three types of block rank patterns are generated, from two directional gradient magnitude images G. The two block sums of gradient magnitudes (horizontal block sum B^h and vertical block sum B^v) are described as, respectively,

$$
\begin{aligned}
B^h(i, j) &= \sum_{x=1+W(i-1)/3}^{Wi/3} \sum_{y=1+H(j-1)/3}^{Hj/3} \left| G^h(x, y) \right| \\
B^v(i, j) &= \sum_{x=1+W(i-1)/3}^{Wi/3} \sum_{y=1+H(j-1)/3}^{Hj/3} \left| G^v(x, y) \right|
\end{aligned}
\tag{1}
$$

where gradient magnitude images G^h and G^v correspond to horizontal and vertical directions, respectively. i and j represent the row and column indices of nine (3×3) blocks in a scanning window, $1 \le i, j \le 3$, respectively. W and H denote width and height of the scanning window S, respectively, and are assumed to be integer multiples of two. The block sums, B^h and B^v obtained by adding the sums of gradient magnitudes of all pixels in each block, are ranked in descending order.

For ranking the blocks, the block sums are converted to one-dimensional (1-D) arrays. The two types of 1-D arrays b_k^h and b_k^v corresponding to 2-D block sums, B^h and B^v, respectively, are expressed as

$$b_k^h = B^h(i, j), \quad b_k^v = B^v(i, j) \tag{2}$$

where k ($k=3(i-1)+j$, $1 \le k \le 9$) represents the index of 1-D arrays. Next, $temp_k^h$ and $temp_k^v$ are copied from the two types of 1-D arrays, b_k^h and b_k^v, respectively, and sorted in descending order. Then, 1-D array ranks, r_k^h and r_k^v are generated. Then, three types of the 3×3 block rank patterns in a face region, P_{eyes}^h, P_{mouth}^h, and P_{nose}^v are generated as

$$P_{eyes}^h(i, j), \; P_{mouth}^h(i, j) = r_k^h, \quad P_{nose}^v(i, j) = r_k^v \tag{3}$$

where $1 \le k \le 9$, $i = \lfloor k/3 \rfloor + 1$, with $\lfloor x \rfloor$ denoting the largest integer not greater than x, and $j = (k \bmod 3) + 1$. Three block rank patterns $P=\{ P_{eyes}^h, P_{mouth}^h, \text{and } P_{nose}^v \}$ are shown in Figures 5(a), 5(b), and 5(c), respectively, where elements in the 3×3 patterns are expressed in terms of r_k^h and r_k^v.
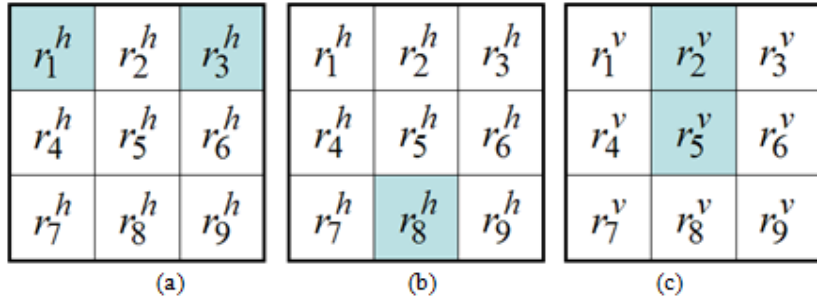

Figure 5. Three types of block rank patterns for a test face region ($P=\{ P_{eyes}^h, P_{mouth}^h, \text{and } P_{nose}^v \}$). (a) eyes in the horizontal direction (P_{eyes}^h), (b) mouth in the horizontal direction (P_{mouth}^h), and (c) nose in the vertical direction (P_{nose}^v).

Figure 6 shows an example of two block rank patterns obtained from gradient magnitude images in Figures 3(b) and 3(c). A test image in Figure 3(a) is an illumination-corrected image, as
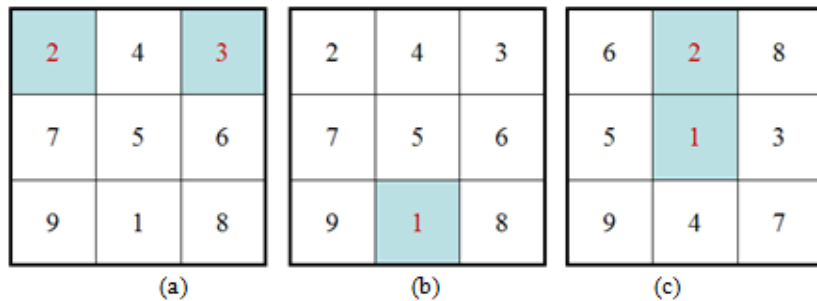

Figure 6. An example of three types of block rank patterns obtained from gradient magnitude images in Figures 3(b) and 3(c). (a) eyes in the horizontal direction (P_{eyes}^h), (b) mouth in the horizontal direction (P_{mouth}^h), and (c) nose in the vertical direction (P_{nose}^v).

described in Section 2.3. Figures 6(a) and 6(b) show the three block rank patterns: eyes in the horizontal direction (P_{eyes}^h), mouth in the horizontal direction (P_{mouth}^h), and nose in the vertical direction (P_{nose}^v), respectively. Figure 6 shows that rough location of facial components can be described by the block rank patterns, where the high ranks are located in the blocks that correspond to the facial components.

The rank of a block is determined based on the gradient magnitude distribution which is obtained from sums of gradient magnitudes of all pixels in each of nine (3×3) blocks. The sum of gradient magnitudes in a block reflects existence of facial components located in that block of the face candidate region as in Figure 4. Each of the three types of the block rank patterns represents different information of gradient magnitude distributions of facial features in a block depending on the direction. The face is detected if three types of the block rank patterns, shown in Figure 6 are the same as or similar to three reference block patterns, in which gray blocks are the highest-rank blocks that correspond to locations of facial features. However, the block rank patterns of all faces cannot be the same as those in Figure 6, since the face can have different block rank patterns that are changed by the variability of face appearances. Thus, a large number of test face images are used to construct the reference rank patterns, called templates. For construction of templates, experiments are performed by obtaining a large number of block rank patterns from each face in test images.

Figure 7 shows experimental results of rank averages for each of three block rank patterns. As rank patterns in Figure 7 are generated by a large number of block rank patterns of test face images, they can be used as reference patterns or templates. The test face images from FERET [17] and GT [24] databases give 1850 gradient magnitude images. Figures 7(a), 7(b), and 7(c) show the average horizontal T_{eyes}^h, T_{mouth}^h, and vertical T_{nose}^v block rank patterns, respectively, noting that the average ranks are real-valued. Figure 7 represents that the block rank patterns are not always the same as Figure 6, because the face images can generate different block rank patterns. Note that first we obtain block ranks of a given test face image and then average the block ranks of 1850 test images to get a set of templates T ($T = \{ T_{eyes}^h, T_{mouth}^h,$ and $T_{nose}^v \}$).

2.1	5.54	2.68
6.2	4.3	6.75
7.64	1.49	8.3

2.1	5.54	2.68
6.2	4.3	6.75
7.64	1.49	8.3

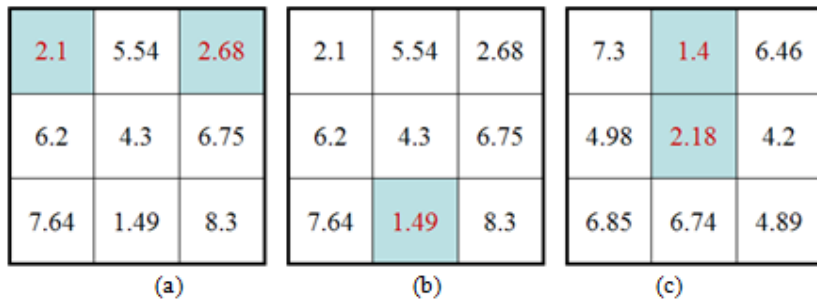| 7.3	1.4	6.46
4.98	2.18	4.2
6.85	6.74	4.89

(a)   (b)   (c)

Figure 7. Block rank averages for three types of template block rank patterns. Block rank averages for eyes in the horizontal direction (T_{eyes}^h), (b) mouth in the horizontal direction (T_{mouth}^h), and (c) nose in the vertical direction (T_{nose}^v).

Now we describe the procedure to define the significant blocks for two directions: horizontal and vertical. Significant blocks (i.e., highest-rank blocks) contain facial features that are well described in the gradient magnitude image for the specific direction. Figure 8 shows two distributions as a decreasing function of the block rank, calculated from 1850 gradient magnitude

images, in which the sum of gradient magnitudes is obtained by adding gradient magnitudes at all the pixels in a block. Figures 8(a) and 8(b) illustrate the horizontal gradient magnitude distribution as a decreasing function of the block rank. The first and second to third rank blocks are significant blocks distinguished from the remaining rank blocks because of the large difference of the sums of gradient magnitudes between two groups of rank blocks (the highest-rank blocks and the remaining rank blocks). The highest-rank blocks are blocks of eyes (1, 1) and (1, 3), and block of mouth (3, 2) that show large sums of gradient magnitudes in the horizontal gradient magnitude image. Figure 8(c) illustrates the vertical gradient magnitude distribution, in which the first and second rank blocks are significant blocks. The highest-rank blocks are blocks of nose (1, 2) and (2, 2) which show large sums of gradient magnitudes in the vertical gradient magnitude image. So, we use the two sets of block ranks in the horizontal gradient magnitude image (i.e., eyes and mouth) and one set of block ranks in the vertical gradient magnitude images (i.e., nose).

As previously mentioned, we can construct 4×4 or 5×5 block rank patterns, however in those cases two groups of rank blocks are not well defined.

Note that rank patterns are not always the same as previously mentioned, however, the highest-rank blocks (significant blocks indicated in gray) in Figure 5 are similar to those in Figure 7. In other words, there is not a large difference between rank values for the highest-rank blocks in Figure 8.
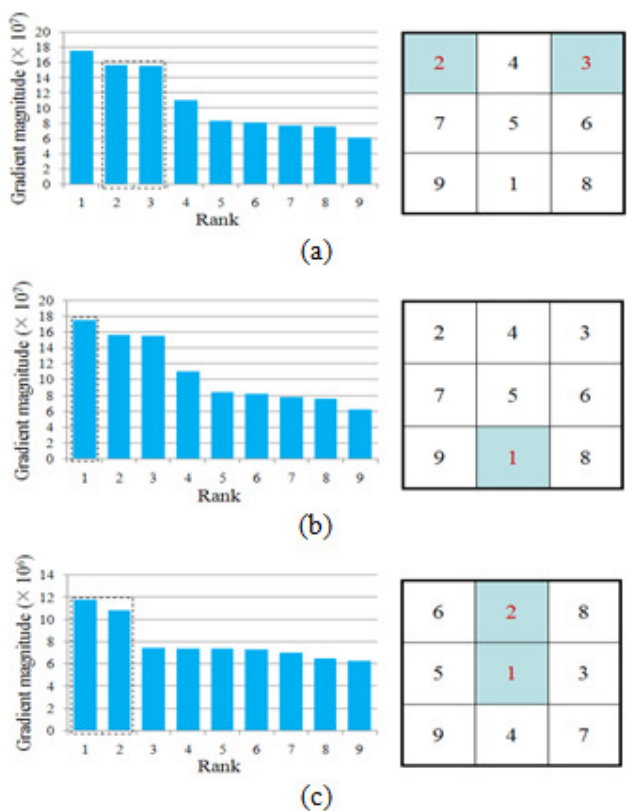

Figure 8. Distribution of the sum of gradient magnitudes, as a function of the block rank, obtained from 1850 face images in FERET and GT databases. (a) eyes in the horizontal direction, (b) mouth in the horizontal direction, and (c) nose in the vertical direction.

Figure 7 shows that the highest-rank blocks are not sensitive to the variability of face appearances, because rank patterns in Figure 7 are obtained from a large number of test face images (1850 images). The ranks of the highest-rank blocks may vary, however, limited by the large difference of the sums of gradient magnitudes between the highest-rank blocks and the remaining rank blocks, which are shown in Figure 8.

We define significant blocks in each of three block rank patterns, in which significant blocks well explain facial features of the face candidate region for three block rank patterns: $P_{eyes}^h, P_{eyes}^h,$ and P_{nose}^v . The three highest-rank blocks $HP_{eyes}^h, HP_{eyes}^h,$ and HP_{nose}^v can be defined as

$$HP_{eyes}^h = \left\{ r_k^h \mid 1 \leq r_k^h \leq 9, k = 1, 3 \right\}$$
$$HP_{mouth}^h = \left\{ r_k^h \mid 1 \leq r_k^h \leq 9, k = 8 \right\} \tag{4}$$
$$HP_{nose}^v = \left\{ r_k^h \mid 1 \leq r_k^h \leq 9, k = 2, 5 \right\}.$$

The elements of the highest-rank blocks are shown in Figure 6 as gray blocks. For example, three block rank patterns are generated from an input image, as shown in Figure 6. From the horizontal block rank pattern, the sum of ranks in the highest-rank blocks HP_{eyes}^h and HP_{mouth}^h (i.e., blocks of eyes and mouth (1, 1), (1, 3), and (3, 2)) is computed. Next, from the vertical block rank pattern, the sum of ranks in the highest-rank blocks HP_{nose}^v (i.e., blocks of nose (1, 2) and (2, 2)) is computed.

In templates (three types: $T_{eyes}^h, T_{eyes}^h,$ and T_{nose}^v), the highest-template blocks $HT_{eyes}^h, HT_{mouth}^h,$ and HT_{nose}^v can be described as

$$HT_{eyes}^h = \{ t_k^h \mid 1 \leq t_k^h \leq 9, k = 1, 3 \}$$
$$HT_{mouth}^h = \{ t_k^h \mid 1 \leq t_k^h \leq 9, k = 8 \} \tag{5}$$
$$HT_{nose}^v = \{ t_k^v \mid 1 \leq t_k^v \leq 9, k = 2, 5 \}.$$

Similarly in templates of block rank patterns for three types, the highest-template blocks are represented in gray in Figure 7.

In order to consider the variability of ranks of the highest-rank blocks, the ranks of the highest-rank blocks need to be considered together, instead of blockwise comparison. They are combined together and used for classification. According to the proposed algorithm, the face candidate region is classified as a face if the difference between the sum of ranks in the highest-rank blocks of a test image and that of a template is less than the threshold, which is described as

$$\left| \sum_{k=1,3} r_k^h - \left\lceil \sum_{k=1,3} t_k^h \right\rceil \right| \leq m_{eyes}^h, \quad \left| \sum_{k=8} r_k^h - \left\lceil \sum_{k=8} t_k^h \right\rceil \right| \leq m_{mouth}^h \tag{6}$$

$$\left| \sum_{k=2,5} r_k^v - \left\lceil \sum_{k=2,5} t_k^v \right\rceil \right| \leq m_{nose}^v$$

where m^h and m^v are the thresholds for three types of block rank patterns with the subscript denoting the facial features (eyes, mouth, or nose), $\lceil x \rceil$ represents the ceiling function that maps a

real number x to the smallest integer not less than x, and the selected index of summation k represents the highest-rank blocks. If all the two differences are less than each of the three thresholds (m_{eyes}^{h}, m_{mouth}^{h}, and m_{nose}^{v}), the face candidate region S is determined as a face. The thresholds denote the margins of the difference between the sum of ranks in the highest-rank blocks of a test image and that of a template. The ideal threshold is 0, however the variability of ranks causes the error which changes the ranks between the highest-rank blocks and the other blocks. If some occlusions or illumination effects occur in a face, the block ranks are easily changed in the block rank patterns. In these cases, the ideal threshold cannot correctly classify a face. Therefore, the margin of thresholds is considered for reliable performance. However, the margin can increase the false positive (FP) rate. The margin is a trade-off between the FP rate and detection rate. The two thresholds have different ranges. As the margins of thresholds affect the FP rate and detection rate, the two thresholds are determined experimentally as described in Section 3. In this paper, the two thresholds m^{h} and m^{v} are experimentally set to 0 and 3, respectively.

2.5. Geometrical Face Model

From the block pattern classification that roughly detects regions of faces by the 3×3 block rank patterns, the possible faces are validated using geometrical matching rules between facial features [9]. Each of roughly detected regions is preprocessed by the morphological operation to remove isolated noisy pixels and to fill up holes if any. The detected region is converted into a binary image by global thresholding to make essential facial features clearly visible.

This paper uses global threshold, which is experimentally determined to be twice the mean of gradient magnitudes. Next, the morphological operation is applied to the binary image, and the processed image is labeled to group the pixels into connected clusters without isolated small pixels. Figure 9 shows the processing of the morphological operation. Figure 9(a) shows a binary image by the global thresholding and Figure 9(b) shows the structure element of the morphological operation. Figure 9(c) shows candidate clusters of facial features.
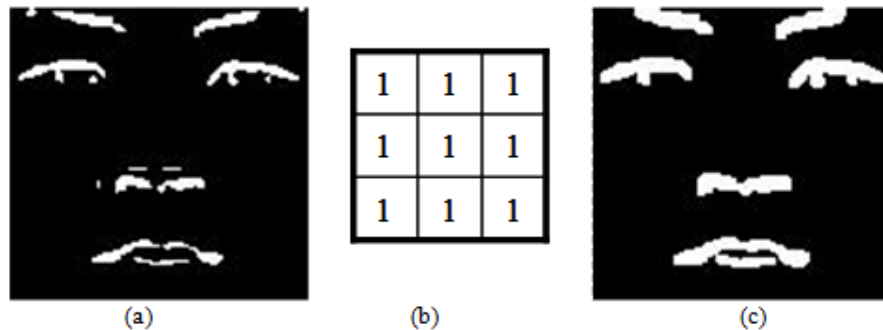

Figure 9. Processing results of the morphological operation. (a) binary image, (b) structure element (3×3), (c) results of the morphological operation.

Using the candidate clusters of facial features, each evaluation function corresponding to each feature (i.e., eyes E_{eyes}, nose E_{nose}, and mouth E_{mouth}) is calculated by quantitative and relative information of facial components [9]. Then, final evaluation function is defined by the weighted sum of respective evaluation functions, which is expressed as

$$E = 0.5E_{eyes} + 0.2E_{nose} + 0.3E_{mouth}. \tag{7}$$

E has the value from 0 to 1. The closer *E* is to 1, the closer the region is to a face. If the value of *E* is larger than threshold, i.e., 0.7, which is selected experimentally, the region is classified as a face.

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

The face images are collected from Caltech face database images [28] for performance evaluation of the proposed face detection method. The Caltech face data consists of 450 images with various illumination, location, and face expression. Image size is 896×592 with a single face per image. Images have color information with various illumination conditions. The database is suitable for performance evaluation of the proposed algorithm that is robust to illumination effects. We use FP and true positive (TP) as performance indicators of face detection. FP represents the area which tells us that there is an image of someone's face although there isn't. However, TP can distinguish someone's face images if there is more than one face. The two indicators have a tendency to trade-off. The performance of the proposed face detection is evaluated in terms of the numbers of FPs and TPs. Receiver operator characteristic (ROC) curve, which is a representative performance curve, shows the correlation between FP and TP by adjusting parameters of the face detection algorithm as shown in Figure 10.
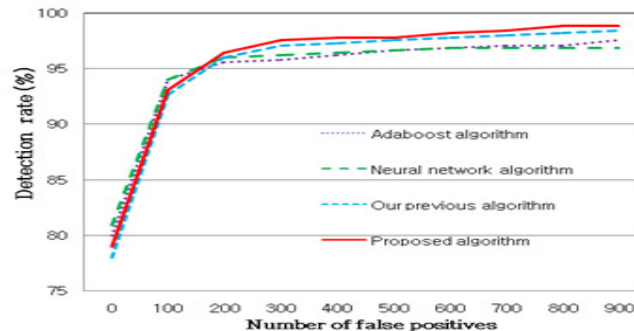

Figure 10. ROC comparison of existing and proposed algorithms.

This paper compares the performance of the proposed face detection algorithm and three existing algorithms: adaboost algorithm [22], NN algorithm [18], and our previous algorithm [16].

To compare the performance of face detection methods, parameter values that give TPs and FPs are varied. In the proposed algorithm, T^h, T^v, and T^d, which are used to obtain the block rank pattern of gradient magnitude images, are changed. In the adaboost algorithm, OpenCV [2] parameter of scale factor varies from 1.1 to 1.3. In the NN algorithm, the number of learning data from CMU database [21] varies from 40 to 69 for face images and from 40 to 59 for non-face images. Figure 10 shows the result of ROC curve. It is observed that the proposed face detection algorithm is more effective than two existing algorithms if the number of FPs is larger than 180. This result shows that the detection rate of the proposed algorithm is higher than those of two existing methods.

Although the proposed method has more FPs than the adaboost, NN, and our previous algorithms, the detection rate is higher. The proposed face detection algorithm is less sensitive to illumination than the adaboost and NN algorithm. It tells us that the face detection rate is higher for test images that are affected by illumination change because the proposed algorithm detects some faces that the adaboost and NN algorithms do not detect.

Also, Table 1 lists three performance measures [12], i.e., precision rate *P*, recall rate *R*, and *f*-measure *F*, with two different values of FPs. The precision rate *P* is defined by (number of TPs)/(number of faces) and the recall rate *R* is defined by (number of TPs)/(number of TPs + number of FPs). The *f*-measure *F* is defined as

$$f = \frac{2PR}{P+R}.$$ (8)

Table 1. Performance comparisons of four face detection algorithms

Number of FPs	Methods	*P*	*R*	*f*
100	Adaboost algorithm	0.8088	0.9400	0.8695
	NN algorithm	0.8088	0.9400	0.8695
	Our previous algorithm	0.8066	0.9267	0.8625
	Proposed algorithm	0.8073	0.9311	0.8648
200	Adaboost algorithm	0.6825	0.9556	0.7963
	NN algorithm	0.6835	0.9600	0.7985
	Our previous algorithm	0.6835	0.9600	0.7985
	Proposed algorithm	0.6845	0.9644	0.8007
900	Adaboost algorithm	0.3279	0.9756	0.4908
	NN algorithm	0.3263	0.9689	0.4882
	Our previous algorithm	0.3299	0.9844	0.4941
	Proposed algorithm	0.3304	0.9867	0.4950

When the number of FPs is larger than 100, *f*-measure of the proposed algorithm is larger than those of our previous algorithm, while larger than 200, *f*-measure of the proposed algorithm is larger than those of the adaboost, NN algorithm, and our previous algorithm, which is the same tendency observed from the ROC curve shown in Figure 10.

Figure 11 shows test images containing a single face in Caltech database images, in which the left image is affected by dark illumination whereas the right one by the direction of illumination source. Figure 12 shows simulation results using parameters in which 600 FPs are detected in 450 face images. If these images are used to detect face images by the adaboost and NN algorithm, faces are not detected well, as shown in Figures 12(a) and 12(b), respectively. However, the our previous algorithm [16] and proposed algorithm detect faces well, as shown in Figures 12(c) and 12(d), respectively, because of the illumination change correction.
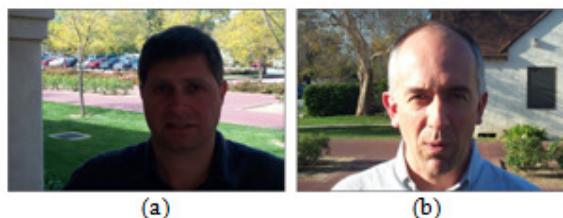

Figure 11. Caltech database images. Input images affected by (a) dark illumination,
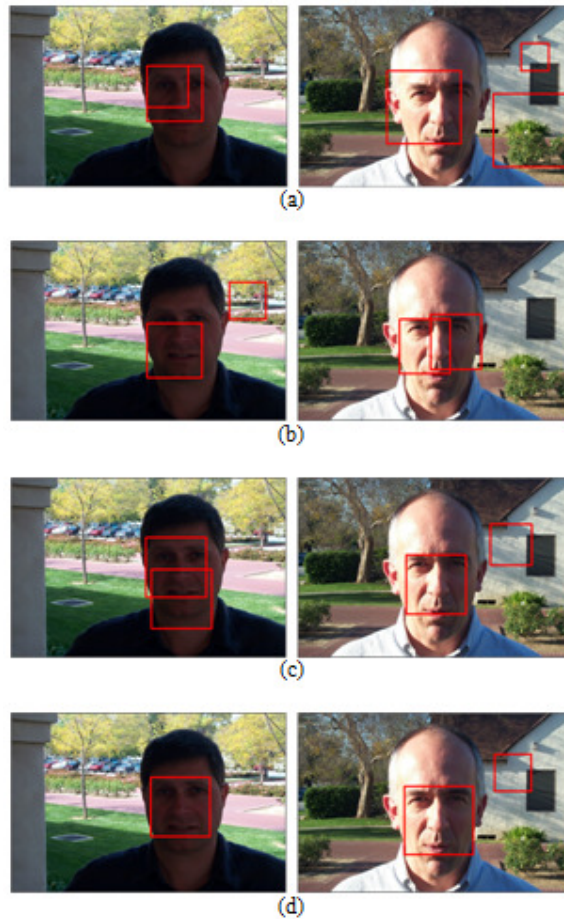(b) direction of illumination source.

Figure 12. Comparison of detection results in Caltech database images. Result images detected by the (a) adaboost algorithm, (b) NN algorithm, (c) our previous algorithm [16], and (d) the proposed algorithm.

Figure 13 shows test images containing multiple faces in real images, in which the left and right image are affected by the direction of illumination source in outdoor and indoor environment, respectively. Figure 14 shows simulation results using parameters in which 200 FPs are detected in 450 face images. The result shows that the adaboost, NN, and our previous algorithm [16] either do not detect faces with FPs or do detect face with FPs, as shown in Figures 14(a), 14(b), and 14(c), respectively. However, the proposed face detection algorithm does detect all faces without FP, as shown in Figure 14(d). However, the proposed face detection algorithm incorrectly detects some faces with mustache around the mouth, because mustache affects the gradient distribution.
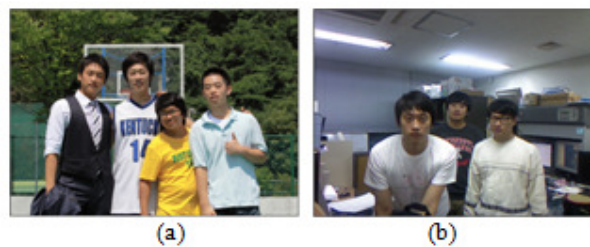

Figure 13. Real images containing multiple faces. Input images affected by varying illumination condition in (a) outdoor environment, and (b) indoor environment.
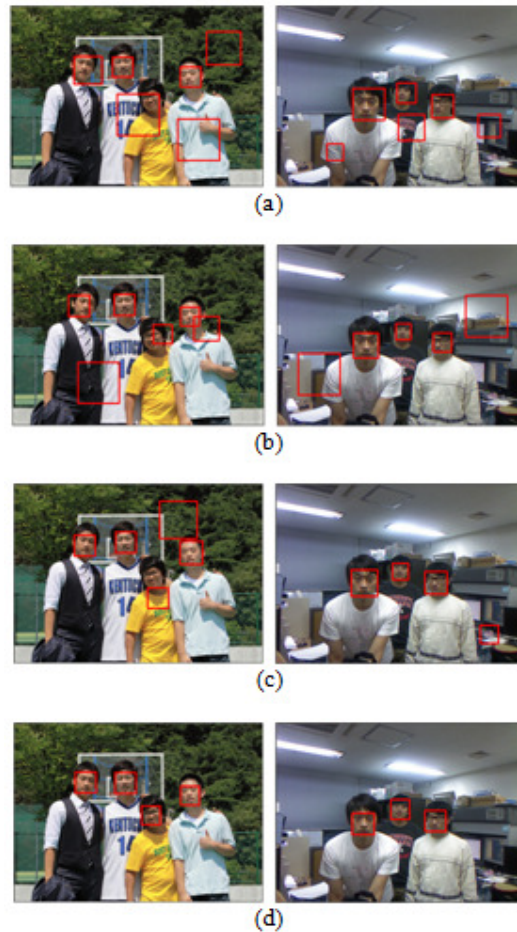
Figure 14. Comparison of detection results in real images. Result images detected by the (a) adaboost algorithm, (b) NN algorithm, (c) our previous algorithm, and the (d) proposed algorithm.

We experiment with varying illumination condition and various face expression for the proposed face detection algorithm. Experimental results with various test images show that the proposed algorithm is more robust to illumination than the adaboost and NN algorithms, and less affected by illumination change because of a preprocessing step that corrects distortion of illumination. However, some face images are incorrectly detected by some obstacles like mustache that affect the gradient distribution in a block detected.

4. CONCLUSIONS

In this paper, we propose a face detection algorithm using the 3×3 block rank patterns of the gradient magnitude images and a geometrical face model. Using the 3×3 block rank patterns in two directions, we obtain useful information, which represents well facial features of eyes, nose, or mouth. The proposed is relatively robust to illumination than pixel-based methods and less affected by illumination change because of a preprocessing step that corrects distortion of illumination. Experimental results show the effectiveness of the proposed algorithm, i.e., the proposed method is more robust to illumination change than the adaboost and NN algorithms, and the proposed algorithm detects faces with fewer FPs than our previous algorithm. Future research will focus on the performance improvement of the proposed algorithm for various complex images that contain occlusions.

ACKNOWLEDGEMENTS

REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition", *IEEE Trans. Pattern Anal. Machine Intell,*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[2] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*, Sebastopol, CA: O'Reilly Media Inc., 2008.

[3] R. Brunelli and T. Poggio, "Face recognition: Features versus templates", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 10, pp. 1042–1052, Feb. 1992.

[4] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, *Third edition*. Upper Saddle River, NJ: Pearson Education Inc., 2010.

[5] P. S. Hiremath and A. Danti, "Detection of multiple faces in an image using skin color information and lines-of-separability face model", *Int. J. Pattern Recognition and Artificial Intell.*, vol. 20, pp. 39–61, Jan. 2006.

[6] L.-L. Huang and A. Shimizu. "A multi-expert approach for robust face detection", *IEEE Trans. Pattern Recognition*, vol. 39, no. 9, pp. 1695–1703, Sep. 2006.

[7] L.-L. Huang, A. Shimizu, Y. Hagihara, and H. Kobatake, "Gradient feature extraction for classification-based face detection", *Pattern Recognition*, vol. 36, no. 11, pp. 2501–2511, Nov. 2003.

[8] L.-L. Huang, A. Shimizu, and H. Kobatake, "Robust face detection using Gabor filter features", *Pattern Recognition Letters*, vol. 26, no. 11, pp. 1641–1649, Aug. 2005.

[9] S.-H. Jeng, H.-Y. M. Liao, C.-C. Han, M.-Y. Chern, and Y. T. Liu, "Facial feature detection using geometrical face model: An efficient approach", *Pattern Recognition*, vol. 31, no. 3, pp. 273–282, Mar. 1998.

[10] A. Kaushal and J. P. S. Raina, "Face detection using neural network & Gabor wavelet transform," *Int. J. of Computer Science and Application*, vol. 1, no. 1, pp. 58–63, Sep. 2010.

[11] S. Kherchaoui and A. Houacine, "Face detection based on a model of the skin color with constraints and template matching," in *Proc. Int. Conf. Machine and Web Intelligence*, Algiers, Algeria, pp. 469–472, Oct. 2010.

[12] T. C. W. Landgrebe, P. Paclik, R. P. W. Duin, and A. P. Bradley, "Precision-recall operating characteristic (P-ROC) curves in imprecise environments", in *Proc. Int. Conf. Pattern Recognition*, vol. 4, pp. 123–127, Hong Kong, China, Aug. 2006.

[13] L. Liu, N. Sang, S. Yang, and R. Huang, "Real-time skin color detection under rapidly changing illumination conditions," *IEEE Trans. Consumer Electronics*, vol. 57, no. 3, pp. 1295–1302, Aug. 2011.

[14] Z. Liu, J. Zhou, G. Xiong, and Y. Liu, "An efficient approach for face detection based on genetic algorithm", in *Proc. IEEE Int. Conf. Communication, Circuits and Systems and West Sino Expositions*, pp. 1127–1131, Chengdu, China, June 2002.

[15] M. Nilsson, J. Nordberg, and I. Claesson, "Face detection using local SMQT features and split up snow classifier", in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 2, pp. 589–592, Honolulu, HI, Apr. 2006.

[16] K.-S. Park, R.-H. Park, and Y.-G. Kim, "Face detection using the 3x3 block rank patterns of gradient magnitude images and a geometric face model", in *2011 Digest of Technical Papers Int. Conf. Consumer Electronics*, pp. 818–819, Las Vegas, NV, Jan. 2011.

[17] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms", *Image and Vision Computing*, vol. 16, no. 27, pp. 295–306, Apr. 1998.

[18] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based on face detection", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 1, pp. 22–38, Jan. 1998.

[19] F. Schroff, T. Treibitz, D. Kriegman, and S. Belongie, "Pose, illumination and expression invariant pairwise face-similarity measure via Doppelgänger list comparison," in *Proc. IEEE Int. Conf. Computer Vision*, pp. 2494-2501, Nov. 2011.

[20] T. Sim, S. Baker, and M. Bast, "The CMU pose, illumination and expression database", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[21] K.-K. Sung and T. Poggio, "Example-based learning for view-based human face detection", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 1, pp. 39–51, Jan. 1998.

[22] P. Viola and M. Jones, "Robust real-time face detection", *Int. J. Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[23] J.-W. Wang, "Precise face segmentation for recognition", in *Proc. IEEE Int. Conf. Image Proc.*, pp. 2045–2048, Atlanta, GA, Oct. 2006.

[24] Y. Wang and K. N. Plataniotis, "Face based biometric authentication with changeable and privacy preservable templates", in *Proc. Biometrics Symposium*, pp. 1–6, Baltimore, MD, Sep. 2007.

[25] C. A. Waring and L. Xiuwen, "Face detection using spectral histograms and SVMs", *IEEE Trans. Systems, Man, and Cybernetics, Part B*, vol. 35, no. 3, pp. 467–476, June 2005.

[26] M.-H.Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 1, pp. 34–58, Jan. 2002.

[27] A. Yuille, P. Hallinan, and D. Cohen, "Feature extraction from faces using deformable templates", *Int. J. Computer Vision*, vol. 8, no. 2, pp. 99–111, Nov. 1992.

[28] W. Zhang, B. Yu, G. J. Zelinsky, and D. Samaras, "Object class recognition using multiple layer boosting with heterogeneous features", in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 323–330, San Diego, CA, June 2005.

AUTHORS

Kang-Seo Park received the B.S. and M.S. degrees in electronic engineering from Sogang University, Seoul, Korea in 2006 and 2008, respectively. His research interests include image processing, pattern recognition, and computer vision.

Young-Gon Kim received the B.S. degree in electronic engineering from Hallym University in 2010. Currently he is working toward the M.S. degree in electronic engineering at Sogang University. His current research interests are image processing and face recognition.

Rae-Hong Park received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1976 and 1979, respectively, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1981 and 1984, respectively. In 1984, he joined the faculty of the Department of Electronic Engineering, School of Engineering, Sogang University, Seoul, Korea, where he is currently a Professor. In 1990, he spent his sabbatical year as a Visiting Associate Professor with the Computer Vision Laboratory, Center for Automation Research, University of Maryland at College Park. In 2001 and 2004, he spent sabbatical semesters at Digital Media Research and Development Center (DTV image/video enhancement), Samsung Electronics Co., Ltd., Suwon, Korea. In 2012, he spent a sabbatical year in Digital Imaging Business (R&D Team) and Visual Display Business (R&D Office), Samsung Electronics Co., Ltd., Suwon, Korea. His current research interests are video communication, computer vision, and pattern recognition. He served as Editor for the Korea Institute of Telematics and Electronics (KITE) *Journal of Electronics Engineering* from 1995 to 1996. Dr. Park was the recipient of a 1990 Post-Doctoral Fellowship presented by the Korea Science and Engineering Foundation (KOSEF), the 1987 Academic Award presented by the KITE, the 2000 Haedong Paper Award presented by the Institute of Electronics Engineers of Korea (IEEK), the 1997 First Sogang Academic Award, and the 1999 Professor Achievement Excellence Award presented by Sogang University. He is a co-recipient of the Best Student Paper Award of the IEEE Int. Symp. Multimedia (ISM 2006) and IEEE Int. Symp. Consumer Electronics (ISCE 2011).