



Pixel selection based on discriminant features with application to face recognition

Sang-Il Choi^{a,*}, Chong-Ho Choi^b, Gu-Min Jeong^c, Nojun Kwak^d

^a Department of Applied Computer Engineering, Dankook University, 126, Jukjeon-dong, Suji-gu, Yongin-si, Gyeonggi-do 448-701, Republic of Korea

^b School of Electrical Engineering and Computer Science, Seoul National University, #047, San 56-1, Sillim-dong, Gwanak-gu, Seoul 151-744, Republic of Korea

^c Electrical Engineering, Kookmin University, 861-1, Jeongeung-dong, Songbuk-gu, Seoul 136-702, Republic of Korea

^d Division of Electrical and Computer Engineering, Ajou University, San 5, Woncheon-dong, Yeongtong-gu, Suwon 443-749, Republic of Korea

ARTICLE INFO

Article history:

Received 2 August 2010

Available online 13 January 2012

Communicated by S. Sarkar

Keywords:

Face recognition

Pixel selection

Illumination variation

Expression variation

Linear discriminant analysis

ABSTRACT

We propose a pixel selection method in a face image based on discriminant features for face recognition. By analyzing the relationship between the pixels in face images and features extracted from them, the pixels that contain a large amount of discriminative information are selected, while the pixels with less discriminative information are discarded. The experimental results obtained with various face databases show that the proposed pixel selection method results in improved recognition performance, especially in the presence of illumination or facial expression variations. Additionally, the proposed method greatly reduces the memory size and computational load in the face recognition process.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Numerous methods have been developed for face recognition in the last few decades (Jain et al., 2000; Tan et al., 2006; Zou et al., 2007). Some of the major difficulties in face recognition are due to a large number of pixels and heavy demand on computational resources. Most face recognition algorithms first project a face image into a lower dimensional space by using a dimension reduction method (Turk and Pentland, 1991; Belhumeur et al., 1997; Cevikalp et al., 2005; Kim and Choi, 2007; Jiang et al., 2008). In the methods based on holistic features, useful features for face recognition are extracted from a full face image. Holistic feature-based methods implicitly preserve all of the detailed texture and shape information that are useful for distinguishing faces, and can capture global aspects of faces. However, since a face image can change greatly depending on the pose, illumination, facial expression and partial occlusion, holistic features can be quite sensitive to these variations, which can be considered as noise. Moreover, not all of the pixels in a face image are helpful for extracting the discriminant features. For example, the pixels, which have small intensity variation in one class and large intensity variation in the other classes, contain discriminative information. On the contrary, the pixels which are not closely related to the class information are susceptible to noise in the discriminant analysis.

Unlike the holistic feature-based methods, some methods extract features from local images, which can be less sensitive to these variations, and therefore provide robust features against these variations. 2D Gabor-like filters were found to be suitable as local descriptors because of their robustness against translation, rotation, and scaling (Liu and Wechsler, 2002; Pang et al., 2004; Kokiopoulou and Frossard, 2006; Yang et al., 2004; Wang et al., 2011). In several papers, a face image was partitioned into a set of sub-images, and then local features were extracted from these sub-images (Gottumukkal and Asari, 2004; Tan and Chen, 2005). The information obtained from these local features was used with (Kim et al., 2005; Pentland et al., 1994) or without global features (Gokberk et al., 2007; Rajagopalan et al., 2007) to improve the accuracy and robustness in the face recognition process.

In this paper, we propose a pixel selection method based on discriminant features for face recognition. The pixels, whose intensity varies greatly due to variations such as illumination or facial expression variations, are likely to interfere in selecting good features for face recognition, and these pixels can be regarded as noise. Thus, we can expect to obtain better discriminant features by eliminating noisy pixels, which will lead to improve face recognition performance. Recently, we presented a preliminary result on the pixel selection method in a face image based on discriminant features for face recognition (Choi et al., 2008). Here, we make further improvements to this method, provide a more detailed analysis and extensive discussion, and present additional experimental results under various conditions.

In the proposed method, we first extract the holistic features from a full face image. Based on the discriminative power of each

* Corresponding author. Tel.: +82 31 8005 3657.

E-mail addresses: csichoisi@gmail.com (S.-I. Choi), chchoi@csl.snu.ac.kr (C.-H. Choi), gm1004@kookmin.ac.kr (G.-M. Jeong), nojunk@ajou.ac.kr (N. Kwak).

basis that constitutes the feature space, we select the pixels which play a more important role in extracting discriminant features, while discarding noisy pixels. The effect of noise reduction by the proposed method is demonstrated by using a toy example. Removing the noisy pixels makes the PSNR (peak signal-to-noise ratio) of the resultant image higher, which can improve the performance of the feature extracted by discriminant analysis (We call this image the ‘reduced image’).

The proposed method is different from other local feature-based methods that select the salient components (such as the eyes, nose and mouth) subjectively, because it objectively selects important pixels in a face image. The reduced image obtained by the proposed method is significantly smaller in size than the original image, resulting in a significant saving in data storage and computational effort. This leads to efficient data transmission and processing that are especially important in mobile applications and real-time systems. On the other hand, various input variable selection methods can be used to select meaningful pixels for face recognition. Information theoretic measures such as the posterior marginal probabilities of each face vertex (Ocegueda et al., 2011), distance discriminant (Liang et al., 2008), Fisher score (Yang et al., 2010), Laplacian score (He et al., 2006) and subspace based separability measure (Gunal and Edizkan, 2008) can be used to evaluate the usefulness of pixels. Unlike these methods, the proposed method is based on the projection vectors obtained by a feature extraction method (discriminant analysis algorithm). Each feature extraction method has its own characteristics, and thus an appropriate method must be used depending on the properties of the data and the problem to be solved. Since the proposed method can inherit the benefit of a properly selected feature extraction method, it can enhance the feature extraction method by eliminating the pixels that have only a small amount of discriminative information. Experimental results in Section 3 show that the proposed method gives better recognition rates than other methods.

The rest of this paper is organized as follows. Section 2 explains how to extract the discriminant features and select the pixels in a face image based on the extracted features. Section 3 presents the experimental results, followed by the discussion and conclusions in Section 4.

2. Pixel selection based on discriminant features

In order to select pixels in a face image which are useful for face recognition, we first extract discriminant features from full face images by using a feature extraction method. Pixels are selected based on these extracted features and the remaining pixels are discarded. This reduced image, consisting of only the selected pixels, is used in the final face recognition process. The overall architecture of face recognition based on the proposed method is shown in Fig. 1.

2.1. Relation between the feature space and the input space

Several methods have been proposed to extract discriminant features for face recognition (Turk and Pentland, 1991; Belhumeur et al., 1997; Cevikalp et al., 2005). Even though any good feature extraction method may be used, we use the discriminant common

vector (DCV) (Cevikalp et al., 2005) to extract discriminant features. This is because DCV, which is a variant of the null space LDA method (Chen et al., 2000), solves not only the “small sample size (SSS) problem” (Fukunaga, 1990), but also performs better for high-dimensional data (e.g. image data) compared to other feature extraction methods. Consider a set of N image samples $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, $\mathbf{x}_k \in \mathbb{R}^n$, each of which belongs to one of c classes. The i -th element of \mathbf{x}_k , x_{ki} , $i = 1, 2, \dots, n$, represents the gray intensity of the i -th pixel of image \mathbf{x}_k . The key idea of DCV is to make use of the observation that the null space of the within-class scatter matrix (S_w) contains a lot of discriminative information (Cevikalp et al., 2005). The columns of the projection matrix W obtained by DCV are the projection vectors $\{\mathbf{w}_l | l = 1, 2, \dots, n'\}$ and the image sample \mathbf{x}_k is represented as a low-dimensional feature vector $\mathbf{y}_k = W^T \mathbf{x}_k (\in \mathbb{R}^{n'}, n' \ll n)$.

The projection vector \mathbf{w}_l with a large eigenvalue λ_l gives the direction which is suitable for discriminant analysis. Thus, we estimate the significance of each pixel by measuring its contribution in constructing the projection vector that constitutes a discriminant feature space. Since \mathbf{w}_l and \mathbf{x}_k have the same dimension, \mathbf{w}_l can be represented as a vector in the image space. Let $\mathbf{w}_l \in \mathbb{R}^n$ be the projection vector corresponding to the l -th largest eigenvalue obtained by a feature extraction method, e.g. DCV, and let $\mathbf{e}_i \in \mathbb{R}^n$ be the i -th unit coordinate vector of the image space. Then, \mathbf{w}_l can be expressed by a linear combination of \mathbf{e}_i s as follows:

$$\mathbf{w}_l = [w_{l1}, w_{l2}, \dots, w_{ln}]^T = w_{l1}\mathbf{e}_1 + w_{l2}\mathbf{e}_2 + \dots + w_{ln}\mathbf{e}_n. \quad (1)$$

The magnitude of w_{li} indicates how much the i -th coordinate vector \mathbf{e}_i contributes to the projection vector \mathbf{w}_l . For example, when the dimension of input space is two ($n = 2$), the projection vector $\mathbf{w} \in \mathbb{R}^2$ can be represented in a two-dimensional input space. In Fig. 2, the components w_1 and w_2 of the projection vector \mathbf{w} are the weights to \mathbf{e}_1 and \mathbf{e}_2 , respectively. In this figure, since w_1 is larger than w_2 , the basis \mathbf{e}_1 plays a more important role than the basis \mathbf{e}_2 to construct the projection vector \mathbf{w} . If w_1 is sufficient larger than w_2 in magnitude, eliminating w_2 in representing the vector \mathbf{w} will make only small change in the direction of \mathbf{w} , which may be ignorable. By extending this to the n -dimensional image space, if we can approximate $\mathbf{y}_k = [y_{k1}, \dots, y_{kn}]^T = W^T \mathbf{x}_k$, ($n' \ll n$, by using fewer coordinate vectors of image space, the coordinate vector \mathbf{e}_i should be selected starting from the coordinate vectors corresponding to large $|w_{li}|$ to minimize its approximation error. Therefore, if w_{li} is larger in magnitude than w_{lj} in a projection vector \mathbf{w}_l , the coordinate vector \mathbf{e}_i (i.e., the i -th pixel) can be regarded as more important than \mathbf{e}_j (i.e., the j -th pixel).

In classification problems, feature extraction methods (such as Fisherface, DCV, ERLDA, etc.) produce a feature vector $\mathbf{y}_k = [y_{k1}, y_{k1}, \dots, y_{km}]^T$, which is used as an input to a classifier. For a given input image data sample (\mathbf{x}_k) and a projection vector (\mathbf{w}_l), each feature, y_{kl} , $l = 1, \dots, m$ is computed as a linear combination of pixels $\mathbf{x}_k = [x_{k1}, \dots, x_{kn}]^T$ and weights $\mathbf{w}_l = [w_{l1}, \dots, w_{ln}]^T$, i.e.,

$$y_{kl} = \mathbf{w}_l^T \mathbf{x}_k = w_{l1}x_{k1} + w_{l2}x_{k2} + \dots + w_{ln}x_{kn}. \quad (2)$$

This different weighting for each pixel differentiates the importance of each pixel, and it can be helpful in classification problems to

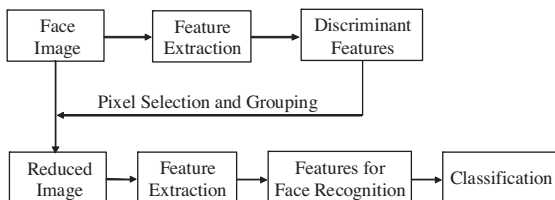


Fig. 1. Pixel selection process.

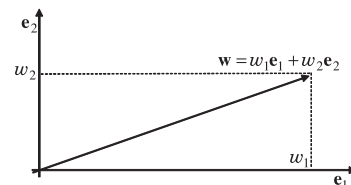


Fig. 2. Projection vector \mathbf{w} in two-dimensional input space.

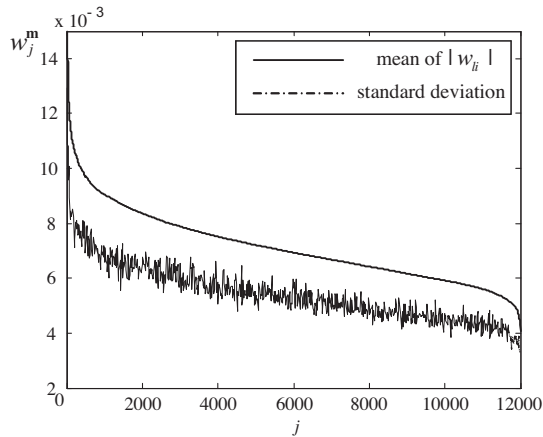


Fig. 3. The mean of $|w_{ji}|$, w_i^m ($= 1/n' \sum_{l=1}^{n'} |w_{li}|$), sorted in descending order and its one standard deviation.

consider only the pixels x_{ki} s associated with larger $|w_{ji}|$. Pixels x_{ki} s associated with smaller $|w_{ji}|$ do not influence much on the decision of the classifier, and consequently these pixels are less important for recognition purpose; rather they might be more susceptible to noise. Thus, eliminating such pixels based on the magnitudes of the projection vectors can help to produce better discriminative features.

We produce n' projection vectors with 400 images from the FERET database (Phillips et al., 2000) by using DCV (in the experiments of Section 3, we used full face images consisting of 120×100 pixels, i.e., $\mathbf{x}_k \in \mathbb{R}^{12,000}$ and set n' to 199). Fig. 3 shows the mean of $|w_{ji}|$, w_i^m ($= 1/n' \sum_{l=1}^{n'} |w_{li}|$), sorted in descending order and its one standard deviation. In the figure, note that the one standard deviation becomes smaller as w_i^m decreases. This implies that when the j -th component of \mathbf{w}_k ($k \neq l$), w_{kj} , is relatively small compared to the other components, the j -th component of the other projection vector \mathbf{w}_l ($k \neq l$), w_{lj} , is more likely to be small. Therefore, we can eliminate the coordinates which do not play a major role in building the feature space without much altering the direction of the projection vectors with large eigenvalues. By using this idea, we propose a pixel selection method based on the order of $|w_{ji}|$ for the projection vector \mathbf{w}_l .

2.2. Pixel selection based on the order vector

In order to effectively select pixels based on the magnitude of w_{ji} , we define an n -dimensional order vector $\mathbf{r}_l = [r_{l1}, r_{l2}, \dots, r_{ln}]^T$ for each projection vector. The i -th component of \mathbf{r}_l , r_{li} , represents the order of w_{ji} when the absolute value $|w_{ji}|$ is sorted in ascending order. For example, if $|w_{ji}|$ is the k -th largest value among $\{w_{ji} | i = 1, 2, \dots, n\}$, r_{li} is assigned the value $n - k + 1$. Then, we make a mask vector $\mathbf{m}'_l = [m'_{l1}, m'_{l2}, \dots, m'_{ln}]^T$ based on the order vector such as

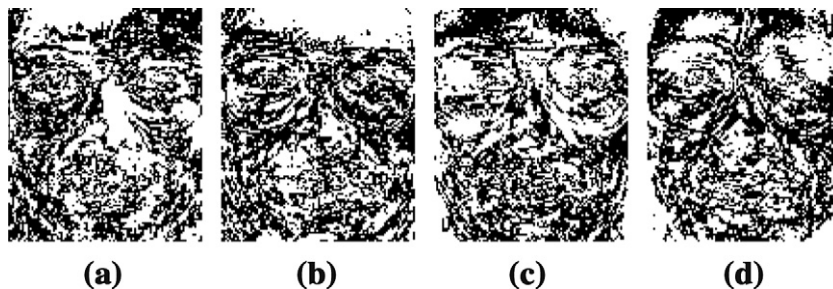


Fig. 4. Some masks \mathbf{m}'_l s obtained by (2); (a) 1st projection vector; (b) 2nd projection vector; (c) 3rd projection vector; (d) 4th projection vector.

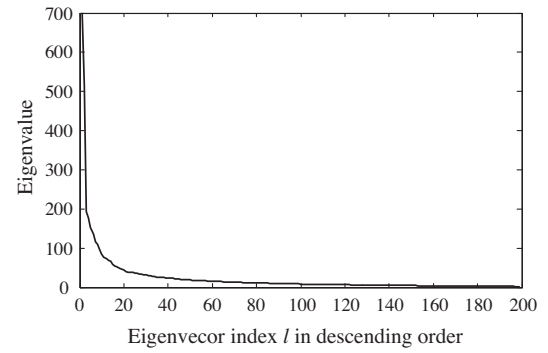


Fig. 5. Eigenvalues in descending order.

$$m'_{li} = \begin{cases} 1 & r_{li} > n - n_s \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where n_s is the total number of pixels to be selected. If $m'_{li} = 1$, it implies that the i -th pixel is considered to be valuable in the face recognition process.

Fig. 4 shows some masks \mathbf{m}'_l s, which are represented as two-dimensional binary images, obtained by (3). Here, the white and black pixels represent the selected and discarded pixels, respectively. As shown in Fig. 4, the mask vectors obtained from the projection vectors are usually different from each other. These mask vectors need to be into one mask vector based on the eigenvalue λ_l of each projection vector. Fig. 5 shows an example of the eigenvalues $\{\lambda_l | l = 1, 2, \dots, n'\}$ plotted against the index l after sorting them in descending order. As shown in Fig. 5, the eigenvalues decrease drastically and most of the sum of the eigenvalues are concentrated in the first 20–30 eigenvalues. Therefore, the order vector \mathbf{r}_l should be treated differently according to its eigenvalue. For this purpose, we define a new vector \mathbf{r}_s which is a sum of weighted order vectors.

$$\mathbf{r}_s = \sum_l \alpha_l \mathbf{r}_l, \quad \alpha_l = \frac{\lambda_l}{\sum_j \lambda_j}. \quad (4)$$

If a threshold T_s is set as the n_s -th largest value of the components of $\mathbf{r}_s = [r_{s1}, r_{s2}, \dots, r_{sn}]^T$, then the mask $\mathbf{m}'' = [m''_1, m''_2, \dots, m''_n]^T$ is obtained as follows:

$$m''_i = \begin{cases} 1 & r_{si} \geq T_s \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

It is noted that \mathbf{m}'' has n_s non-zero elements. Fig. 6(a) shows the masks \mathbf{m}'' obtained by (5), which are represented as two-dimensional binary images, as the total number of pixels to be selected (n_s) increases from 15% to 50% of the total number of pixels in the figure, a selected pixel is represented as a white pixel.

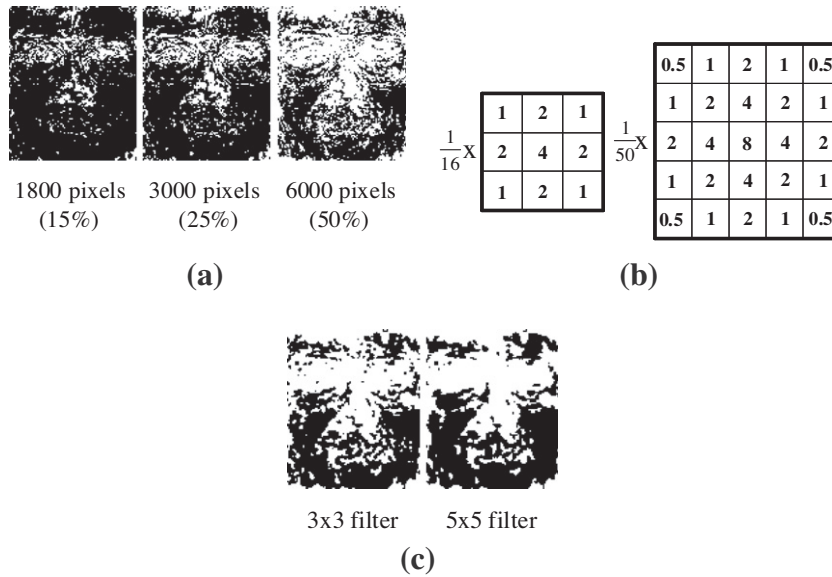


Fig. 6. Mask images; (a) masks \mathbf{m}^n for different numbers of selected pixels; (b) 3×3 and 5×5 spatial filters; (c) final masks \mathbf{m} after pixel grouping corresponding to \mathbf{m}^n for $n_s = 6000$.

2.3. Pixel grouping using a low pass filter

In Fig. 6(a), we see that the selected pixels constitute very complicated boundaries. In some regions, several selected pixels are disconnected from the other selected pixels, whereas in other regions most of the pixels are connected. However, it is unlikely that the discriminant power changes abruptly pixel by pixel. Moreover, since the detailed mask shape can change when different images are used in making masks or when there is an error in the face alignment (Kim and Choi, 2007), it is better to group the selected pixels to deal with such cases.

In order to group the selected pixels, we smooth the boundaries of the selected pixels of \mathbf{m}^n by applying a low pass filter (see Fig. 6(b)) to the mask of Fig. 6(a), and obtain a filtered mask $\mathbf{m}^F = [m_1^F, m_2^F, \dots, m_n^F]^T$. Then, we finally get the mask \mathbf{m} from \mathbf{m}^F with a threshold T_f as follows:

$$m_i = \begin{cases} 1 & m_i^F > T_f \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

The value for the threshold T_f is set so that the number (n_f) of final selected pixels is close to n_s . Fig. 6(c) shows the final masks \mathbf{m} resulting from the pixel grouping using the low pass filters to make a mask such as in Fig. 6(c). Here, the regions of the selected pixels are more densely distributed than regions in Fig. 6(a). This step resulted in a small improvement (0.0–0.8%) in the experiments in Section 3 depending on the database used. In order to investigate the effect of different filter sizes, we performed the pixel grouping process with two different filter sizes, 3×3 and 5×5 , and found that the 5×5 filter was better at alleviating the complicated boundaries in the mask \mathbf{m}^n . The masked images are obtained by multiplying the image \mathbf{x}_k with the mask \mathbf{m} pixel by pixel, and the reduced images $\mathbf{x}'_k \in \mathbb{R}^{n_f}$ are made by eliminating the pixels having zero value in the masked images. In pixel grouping, one may consider applying a low pass filter before the feature extraction step; however, we found that applying the filter after the feature extraction step produces better results.

The masks in Fig. 6(c) show that the selected pixels are distributed mainly around the eyes and nose. This is relatively consistent with the results in psychophysical reports that state that the eyebrows, eyes, nose, and mouth are the most useful components in

facial identification (Sinha et al., 2006; Ocegueda et al., 2011). However, as seen in Fig. 6(c), some pixels in the other region also contribute to facial identification. We will show later that the recognition performance can be improved by using the pixels selected in this way (I_{Re}^n) rather than the pixels selected intuitively (I_{Local}).

The effect of the mask \mathbf{m} on the distribution of w_{li} , $l = 1, 2$, can be seen in Fig. 7. Figs. 7(a) and (b) show the histograms of the values of w_{1i} and w_{2i} before and after applying the mask \mathbf{m} . Note that w_{1i} s and w_{2i} s are the i -th components of the projection vectors corresponding to the largest two eigenvalues produced by the feature extraction method in SubSection 2.1. The horizontal axis represents the values of w_{li} with the bin size of 0.001 and the vertical axis represents the number of w_{li} in a bin. When comparing Figs. 7(a) and (b), it can be seen that a great portion of smaller weights $|w_{li}|$ are eliminated while the larger weights $|w_{li}|$, which are regarded as important, are mostly kept after the pixel selection process.

The procedure of the proposed pixel selection method can be summarized as follows:

- **Step 1:** From face images, obtain projection vectors \mathbf{w}_l ($l = 1, \dots, n'$), which constitute a feature space by using a feature extraction method.
- **Step 2:** Make an order vector \mathbf{r}_l ($l = 1, \dots, n'$) representing the relative magnitude of each component for the projection vector \mathbf{w}_l , obtained at step 1. Based on the eigenvalues of the projection vectors, merge n' order vectors into a single order vector \mathbf{r}_s by using (4).
- **Step 3:** After obtaining a mask \mathbf{m}^n by using (5), produce the final mask \mathbf{m} through the pixel grouping process.
- **Step 4:** Obtain the reduced image consisting of the selected pixels by multiplying a face image with the final mask \mathbf{m} pixel by pixel.

2.4. Toy example

To show the effectiveness of the proposed method, a toy example is presented. Let us consider a set of 20 images ($\in \mathbb{R}^{10 \times 10}$) as shown in Fig. 8. Each image belongs to one of four classes, and its class can be identified by the position of the white pixels. The

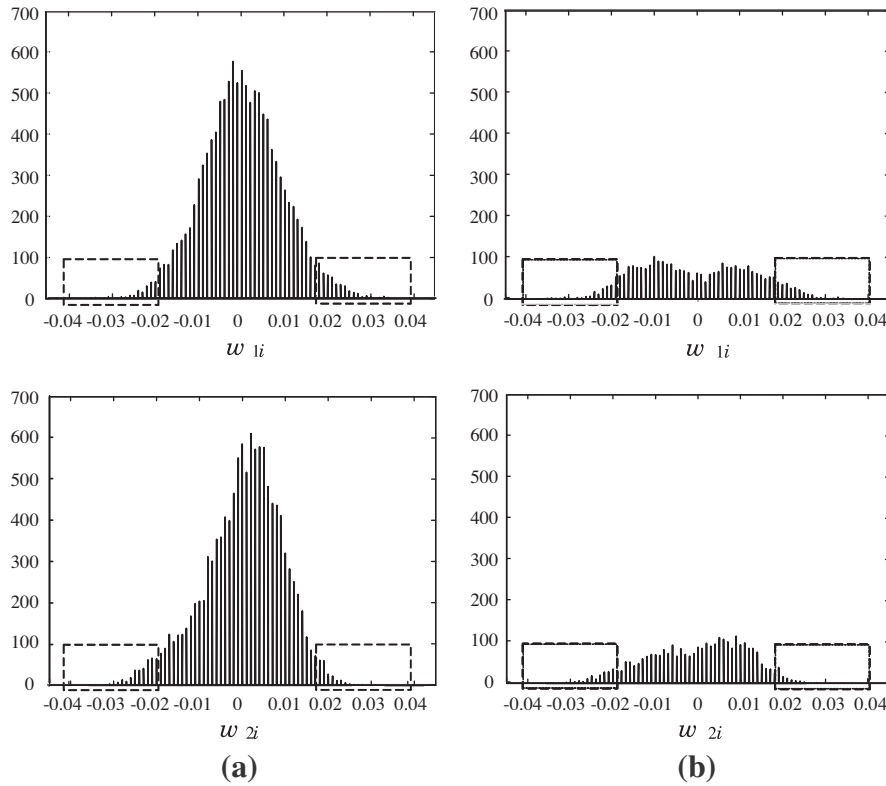


Fig. 7. Distribution of w_{1i} and w_{2i} : (a) before applying the mask; (b) after applying the mask.

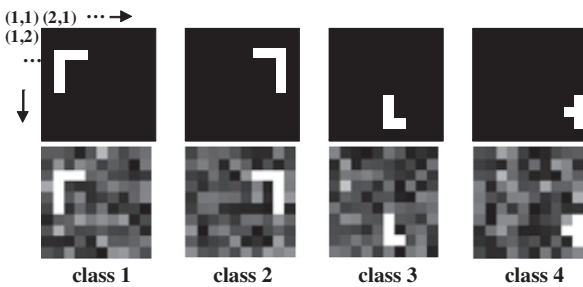


Fig. 8. A toy example. The images from four classes (the upper images: without noise, the lower images: with Gaussian random noise).

number of white pixels is 6, 6, 4 and 4 depending on the class (the images in the upper row of Fig. 8). Then, we added Gaussian random noise with standard deviation of 5 to each image so that the peak signal to noise ratio (PSNR) of an image is 10–13 dB for the

images in the lower row of Fig. 8. In the figure, each pixel located on the j -th row and the i -th column from the upper left corner is denoted by the index (i, j) . It is obvious that the pixels (total 20 pixels) corresponding to the white pixels have the most discriminative information because their variances in the same classes are zero, while those in the different classes are very large. We applied the proposed pixel selection method to this image set and observed which pixels were selected as n_s was increased from 20 to 100 (see Table 1).

For the images in the upper row of Fig. 8 (without noise), when selecting 20 pixels (20%) out of total 100 pixels by the proposed method, i.e., $n_s = 20$, only the 20 white pixels, which have discriminant information, were selected. However, in the experiments for the images in the lower row of Fig. 8 (with Gaussian random noise), 7 noisy pixels were selected along with 13 white pixels (Table 1). Selecting noisy pixels prior to the white pixels is due to the random noise added to images, which makes the projection vectors deviate from the optimal solution in the discriminant analysis. For the experiments with various values of n_s , the indices of missing

Table 1
Number of white pixels selected for various values of n_s .

n_s	20	30	40	50	60	70	80	90	100
No. of white pixels	13	17	18	20	20	20	20	20	20
PSNR ^a	19.1	17.3	15.9	14.8	13.8	13.3	12.6	12.2	11.6
Pixel index ^b	(9,3)	(9,8)	(10,8)	–	–	–	–	–	–
	(6,8)	(10,8)	(10,9)	–	–	–	–	–	–
	(6,9)	(10,9)	–	–	–	–	–	–	–
	(7,9)	–	–	–	–	–	–	–	–
	(9,8)	–	–	–	–	–	–	–	–
	(10,8)	–	–	–	–	–	–	–	–
	(10,9)	–	–	–	–	–	–	–	–

^a Average PSNR of the images which consist of n_s pixels.

^b Locations of white pixels that are eliminated in the selection process.

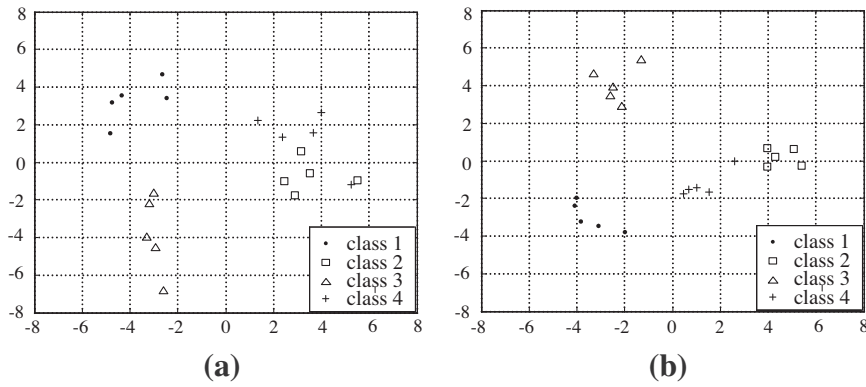


Fig. 9. Sample Distribution for various n_s in two principal component axes. (a) without pixel selection ($n_s = 100$); (b) $n_s = 55$.

white pixels are shown in Table 1. Note that for $n_s = 50$ all the white pixels were selected by the proposed method. This means that the proposed method effectively preserves discriminant pixels while removing the pixels that have less discriminative power.

We investigated the PSNR in an image after the pixel selection. As can be seen in Table 1, the average PSNR of the reduced image, which consists of n_s pixels, increases as n_s decreases. Therefore, by choosing n_s properly, we can obtain a reduced image which includes all the informative pixels and has a higher PSNR for extracting better discriminant features.

In order to compare before and after applying the proposed pixel selection method, we plot the original image samples (for the lower images in Fig. 8) and the reduced image samples for $n_s = 55$ in the subspace consisted of two principal axes. As shown in Fig. 9(a), the original image samples belonging to the same class are widely scattered, and one sample of class 2 and one sample of class 4 are located at nearly the same place. By removing noisy pixels, the samples of the same class are clustered more closely and there is no overlap between samples belonging to different classes (Fig. 9(b)).

3. Experimental results

In this section, we present the experimental results obtained with the FERET, CMU-PIE (Sim et al., 2003), Yale B (Georghiadis and Belhumeur, 2001), and AR (Martnez and Benevente, 1998) databases. Table 2 shows the characteristics of each database. In order to represent the degree of variation of each database, we selected an image taken under normal condition (no illumination, expression variation) for each subject as a reference image, and computed the PSNR of the other images of the subject. As shown in Table 2, the PSNR of the FERET database is higher than the other databases; thus, the images in the FERET database have relatively small variation. The center of each eye was manually detected and the eyes were rotated to be aligned horizontally as in (Choi et al., 2011). Each face image was cropped and rescaled so that the center of each eye is placed at a fixed point in an image of 120 (pixels) \times 100 (pixels). Then, the histogram equalization procedure (Kim et al., 2005) was applied to the downscaled image. Fig. 10 shows

Table 2
Characteristics of each database.

Database	FERET	CMU-PIE	Yale B	AR
No. of subjects	992	65	10	126
No. of images per subject	2	21	45	20
Illumination variation	None	Large	Large	Small
Expression variation	Small	None	None	Large
Occlusion	None	None	None	Partial
Degree of variation (avr. PSNR)	16.9	12.6	12.4	12.1

some examples of the cropped and rescaled face images obtained from the FERET, CMU-PIE, Yale B, and AR databases.

Fig. 11 shows the masks obtained from each database. As shown in Fig. 11, since each database has different characteristics (e.g., different variations), the mask shape may change slightly depending on the database used in making the mask. The reason is that it reflects not only general human facial information, but also specific characteristics of a database. Therefore, better recognition performance can be expected if the images used to make a mask and test images belong to the same database. However, since a test image can come from any database, it is important to prepare a mask that can be applied to a test image with unknown characteristics. The mask should not be biased to particular variation, and also represent the regions that are generally important in facial recognition and can effectively be applied to any face image not involved in the process of making the mask. Thus, we make a mask with the FERET database, which does not include large facial variation, that reflects general facial information well, and has a large number of subjects. Among the images of the 992 subjects whose images were in both 'fa' and 'fb' images of the FERET database, we use two images for each of 200 subjects, one each from the 'fa' and 'fb' images, to make a mask m . We applied the same mask to not only the images from the FERET database but also the images from other databases.

In order to know how many pixels should be selected for face recognition, we checked the recognition rates for the FERET, CMU-PIE and Yale B databases by increasing the number of selected pixels n_s from 10% (1200 pixels) to 100% (12,000 pixels) of the total number of pixels. The features for recognition were extracted by using DCV and the one nearest neighbor rule was used with the l_2 norm as a classifier. As can be seen in Fig. 12, the recognition rate does not always increase as the number of selected pixels increases. The best recognition rates are obtained at approximately 50–70% of the total number of pixels depending on database, and we can say that there are 30–50% redundant pixels in a face image. These may function as noise in extracting discriminant features, resulting in degradation of recognition performance. From these results, we set the number of selected pixels (n_f) to 6033 for all the experiments, which is approximately 50% of the full face image.

We evaluated the recognition performance depending on the type of variations; under small environmental variation, large illumination variation, and various kinds of variations. For each database, we compare the recognition rates of the reduced images I_{Re}^{lr} with several other images, which are the local images I_{Local} , the full face images I_F . The local images were produced to include only the salient facial components, which were the eyes, nose and mouth, commonly used in other local feature-based methods (Kim et al., 2005; Pentland et al., 1994; Brunelli and Poggio, 1993). The regions

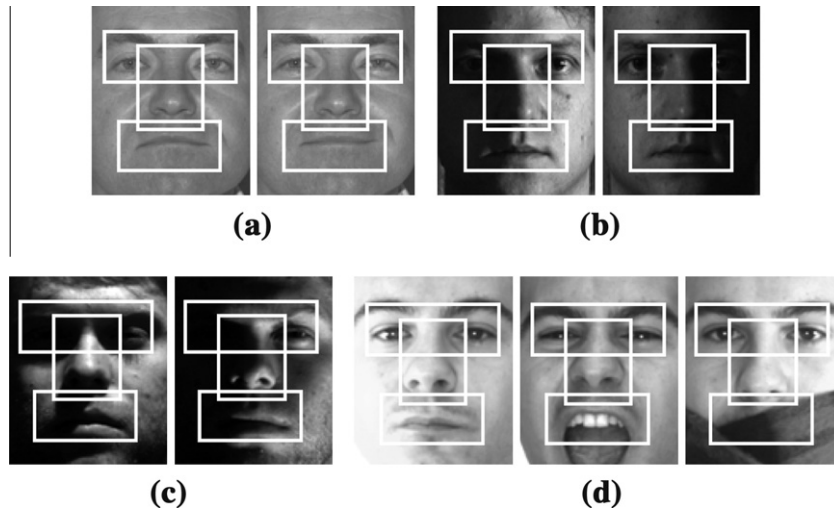


Fig. 10. Examples from various databases. The parts inside the white lines are local images corresponding to the eyes, nose and mouth; (a) FERET database; (b) CMU-PIE database; (c) Yale B database; (d) AR database.

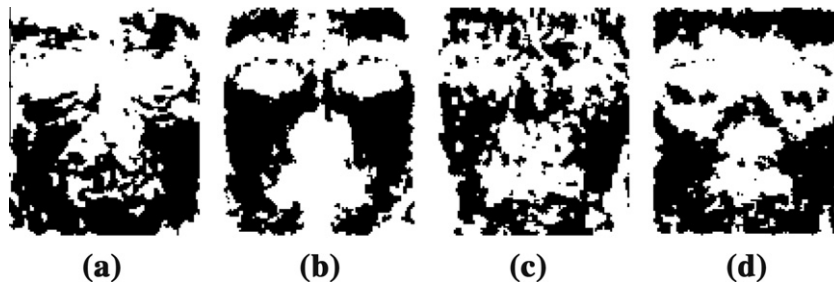


Fig. 11. Masks obtained from (a) FERET database; (b) CMU-PIE database; (c) Yale B database; (d) AR database.

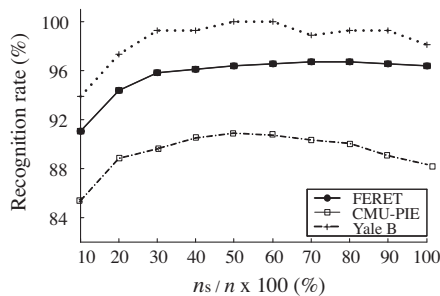


Fig. 12. Recognition performance for various n_s .

including the eyes, nose and mouth were cropped from the face images with reference to the midpoint between the eyes (Fig. 10).

In addition, the proposed method is evaluated with several input variable selection methods, which are the Uniform Sampling (US) (Gonzales and Woods, 2002), FSDD (Liang et al., 2008), SFS, SCS (Gunal and Edizkan, 2008), Laplacian score (LaS) (He et al., 2006) and Semi-Fisher score (SeFS) (Yang et al., 2010). The image I_{US} produced by US was obtained by down sampling the original image uniformly, and the number of pixels was set to nearly the same as n_f . The image resizing was implemented by the bilinear interpolation, which is widely used in image processing (Gonzales and Woods, 2002). The images I_{SFS} , I_{SCS} and I_{FSDD} , all of which consisted of 6000 pixels, were obtained by using SFS, SCS, FSDD, LaS and SeFS, respectively.

In evaluating the face recognition performance, the one nearest neighbor rule was used with the l_2 norm as a classifier, and the

features for the recognition were extracted by using the Fisherface (PCA + LDA), DCV, and Eigenvalue regularized LDA (ERLDA) methods (Jiang et al., 2008). Even though other feature extraction methods may be used, we limit the experiments to these three popular feature extraction methods.

Each database was partitioned into training and testing sets to evaluate recognition performance. For all the databases, training images are randomly selected and the other images not used as training images were tested for each round of experiments. Ten rounds of tests were conducted and the average recognition rate and its standard deviation were computed.

3.1. FERET database: under small environmental variation

For the experiments of face recognition under small environmental variation, we used the FERET database. The FERET database contains frontal images of 992 subjects from both 'fa' and 'fb'. With the exception of 200 subjects that were used to make the mask, 792 subjects were used to evaluate the recognition rates. Both 'fa' and 'fb' images are regular frontal images that do not have illumination variation and there is only a slight expression variation between them. Among 792 subjects, two images ('fa' and 'fb') of 100 subjects were randomly selected for training, and the remaining images of 692 subjects were used to test the recognition performance.

In Table 3, the recognition performance, which was evaluated by using various feature extraction methods (Fisherface, DCV and ERLDA), was compared for several types of images: full face image (I_f), local images (I_{Local}) and the images produced by using other input variable selection methods (I_{SFS}^{6000} , I_{US}^{6000} , I_{SCS}^{6000} , I_{FSDD}^{6000} , I_{LaS}^{6000} , I_{SeFS}^{6000}).

Table 3
Recognition rates for FERET database (%).

Image type	FE ^a		
	Fisher 45-dim.	DCV 69-dim.	ERLDA 69-dim.
I_F (Full face image)	93.0 (± 0.9)	96.2 (± 0.8)	95.8 (± 1.0)
I_{Local} (Local image)	88.6 (± 1.4)	92.5 (± 1.0)	91.9 (± 0.8)
I_{SFS}^{6000} (SFS)	93.3 (± 0.7)	96.0 (± 0.7)	95.7 (± 0.8)
I_{US}^{6000} (US)	93.4 (± 0.9)	96.5 (± 0.7)	96.1 (± 0.8)
I_{SCS}^{6000} (SCS)	93.0 (± 1.1)	96.2 (± 0.6)	95.6 (± 0.8)
I_{FSDD}^{6000} (FSDD)	94.1 (± 1.7)	96.4 (± 0.4)	96.1 (± 0.6)
I_{LaS}^{6000} (LaS)	92.0 (± 1.4)	95.0 (± 0.6)	94.7 (± 0.5)
I_{SeFS}^{6000} (SeFS)	93.9 (± 0.6)	96.4 (± 0.4)	96.2 (± 0.4)
I_{Re}^{6033} (Proposed method)	94.1 (± 0.6)	96.7 (± 0.4)	96.4 (± 0.5)

^a Feature extraction method.

The number in the parentheses represents the standard deviation for each case. The best recognition rates for Fisherface, DCV and ERLDA were obtained when the numbers of features were 45, 69 and 69, respectively.

By using the reduce image (I_{Re}^{6033}), we were always able to improve the recognition rate performance by 0.0–5.5% compared to the other types of images depending on the feature extraction method. The improvements gained by using the reduced images I_{Re}^{6033} are not too much compared to the other types of images, which may be explained by the small variation between the gallery and probe images and low noise (see Table 2) in the FERET database.

3.2. CMU-PIE and Yale B databases: under illumination variation

We applied the proposed method to the CMU-PIE and Yale B databases to evaluate its performance under illumination variation. The CMU-PIE database contains images of 68 subjects with 21 illumination variations. Among them, we selected the images of only 65 subjects, because the images of the other subjects had some defects or did not include all types of illumination variations. In order to confirm the robustness to illumination variation, three images of each subject were randomly selected from the images taken under small illumination variation ('27_06', '27_07', '27_08', '27_11', '27_20') to construct a feature space, while the other eighteen images were tested. Among the test images, one image under frontal illumination was used as a gallery image and the other images were used as the probe images. On the other hand, the Yale B database contains images of 10 subjects in nine poses and 64 illuminations per pose. We used 45 face images for each

individual in the frontal pose (YaleB/Pose00), which were further subdivided into four subsets (subset $i, i = 1, 2, 3, 4$) depending on the direction of light as in (Georghiades and Belhumeur, 2001). The index of the subset increases as the light source moves away from the front of the subject at the time of picture taking. To evaluate the recognition rates, eighteen images in subsets 1 and 2 were randomly selected for the training set and one image under frontal illumination was used as a gallery image. The other images not included in the training set were used as the probe images.

Table 4 shows the comparative recognition rates of several feature extraction methods for the various types of images for the CMU-PIE and Yale B databases. The numbers of features were set to 64 and 9 for the CMU-PIE and Yale B databases, respectively, so that each of the methods (Fisherface, DCV and ERLDA) gives the best recognition rate. For the CMU-PIE database, the reduced image gives 1.1–8.0% better recognition rates compared to the other types of images. Also, for the Yale B database, the proposed method also outperforms compared to other types of images. Although both databases have severe illumination variation, the overall recognition rates are higher in the Yale B database than in the CMU-PIE database. This is because there are more subjects in the CMU-PIE database than in the Yale B database (Table 2), which makes the recognition for the CMU-PIE database more difficult.

These experimental results on the CMU-PIE and Yale B databases show that the proposed method effectively selects the pixels and the resultant reduced images make the recognition performance more robust under illumination variation.

3.3. AR database: under various conditions

The AR database consists of over 4000 frontal images of 126 subjects. The images were taken in two different sessions fourteen days apart. Thirteen images were taken under controlled circumstances in each session. These images include facial variations such as illumination, expression and occlusion. In the experiment, we chose a subset of the database consisting of 64 male subjects and 54 female subjects that had images under all the types of variations. For each subject, 26 images with illumination variation, expression and partial occlusion were selected.

Five images of each subject were randomly selected for training, and the other images were used for testing ('test 1'). We performed another test with the occluded images ('test 2') to observe the effect of the proposed pixel selection method for partially occluded faces. As shown in Table 5, the overall recognition rates are lower than those in Tables 3 and 4, which is due to various types of variations in the AR database. The reduced image I_{Re}^{6033} gives 2.0–5.6% and

Table 4
Recognition rates for CMU-PIE and Yale B databases (%).

Image type	Database					
	CMU-PIE database			Yale B database		
	FE ^a			FE ^a		
	Fisher 64-dim.	DCV 64-dim.	ERLDA 64-dim.	Fisher 9-dim.	DCV 9-dim.	ERLDA 9-dim.
I_F (Full face image)	79.0 (± 2.0)	89.2 (± 0.8)	90.1 (± 1.1)	87.5 (± 1.7)	97.4 (± 1.1)	95.0 (± 1.3)
I_{Local} (Local image)	79.7 (± 2.4)	88.8 (± 1.3)	90.2 (± 1.2)	87.3 (± 1.4)	96.1 (± 1.5)	93.6 (± 1.3)
I_{SFS}^{6000} (SFS)	77.3 (± 2.4)	89.6 (± 1.0)	90.5 (± 1.1)	88.0 (± 2.2)	96.9 (± 1.2)	94.7 (± 0.9)
I_{US}^{6000} (US)	78.3 (± 2.5)	87.6 (± 1.1)	89.2 (± 1.1)	86.3 (± 2.0)	96.7 (± 1.6)	93.9 (± 1.7)
I_{SCS}^{6000} (SCS)	76.5 (± 2.2)	88.0 (± 1.0)	89.3 (± 1.1)	85.2 (± 2.3)	97.0 (± 1.1)	93.9 (± 1.3)
I_{FSDD}^{6000} (FSDD)	79.4 (± 1.8)	87.9 (± 1.5)	87.5 (± 1.4)	86.7 (± 2.0)	97.9 (± 0.9)	94.9 (± 1.2)
I_{LaS}^{6000} (LaS)	82.3 (± 2.2)	90.8 (± 1.5)	90.3 (± 1.6)	91.7 (± 1.7)	97.8 (± 0.9)	97.9 (± 1.1)
I_{SeFS}^{6000} (SeFS)	77.7 (± 1.6)	86.3 (± 1.2)	87.0 (± 1.1)	83.8 (± 2.4)	97.0 (± 1.3)	94.0 (± 1.7)
I_{Re}^{6033} (Proposed method)	84.5 (± 2.0)	91.9 (± 0.9)	92.1 (± 1.1)	91.1 (± 1.8)	99.2 (± 0.6)	96.9 (± 1.3)

^a Feature extraction method.

Table 5
Recognition rates for AR database (%).

Image type	Test					
	Test 1			Test 2		
	FE ^a			FE ^a		
	Fisher 64-dim.	DCV 64-dim.	ERLDA 64-dim.	Fisher 64-dim.	DCV 64-dim.	ERLDA 64-dim.
I_F (Full face image)	71.8 (±0.9)	70.6 (±0.9)	75.7 (±1.3)	66.2 (±1.8)	65.8 (±1.7)	70.1(±1.8)
I_{Local} (Local image)	70.0 (±1.2)	68.8 (±0.8)	75.3 (±0.9)	66.9 (±2.5)	66.1 (±2.2)	72.0(±1.9)
I_{SFS}^{6000} (SFS)	71.7 (±1.3)	70.5 (±1.0)	75.7 (±1.2)	66.6 (±1.6)	65.9 (±1.8)	70.5 (±1.5)
I_{US}^{6000} (US)	72.5 (±1.2)	72.1 (±0.9)	76.5 (±1.4)	66.8 (±2.3)	67.4 (±2.0)	71.2 (±2.1)
I_{SCS}^{6000} (SCS)	70.7 (±1.1)	70.0 (±1.4)	74.7 (±1.3)	65.1 (±2.0)	65.7 (±2.1)	69.6 (±1.6)
I_{FSDD}^{6000} (FSDD)	73.6 (±1.0)	71.5 (±1.5)	77.1 (±1.4)	71.0 (±2.6)	69.6 (±2.1)	74.7(±2.1)
I_{LaS}^{6000} (LaS)	71.1 (±1.0)	68.9 (±1.6)	74.9 (±1.3)	65.6 (±2.4)	65.5 (±2.1)	70.2 (±1.9)
I_{SeFS}^{6000} (SeFS)	72.6 (±1.2)	70.6 (±1.3)	76.2 (±1.3)	68.5 (±2.6)	67.1 (±2.3)	72.4(±2.3)
I_{Re}^{6033} (Proposed method)	75.6 (±1.4)	74.3 (±1.1)	79.2(±1.3)	73.4(±2.0)	73.0(±2.2)	77.7(±2.2)

^a Feature extraction method.

2.4–8.1% better recognition rates for ‘test 1’ and ‘test 2’, respectively, compared to the other types of images, regardless for the feature extraction methods.

The effect of the proposed pixel selection is more remarkable for the AR database than for the other databases. This is because that the PSNR of the AR database is lower than the other databases (see Table 2). From these results, we can see that the proposed pixel selection method can be well applied for face recognition under various types of variations.

4. Discussion and conclusion

Ocegueda et al. (2011), Wang et al. (2011) This paper proposes a pixel selection method for face recognition based on a quantitative measure using discriminant features unlike other local feature-based methods that select local regions heuristically. The elements of projection vectors, which constitute the feature space, can be used to estimate the discriminative power of each pixel. By analyzing the relationship between the input variables (image pixels) and projection vectors obtained by discriminant analysis, the pixels that contain a large amount of discriminative information are selected based on a quantitative measure, while the others are discarded.

The resultant mask of the proposed pixel selection shows that important information for face recognition is concentrated around the eyes, nose and mouth. This is relatively consistent with the result based on 3D face model or many psychophysical reports (Sinha et al., 2006; Ocegueda et al., 2011). The mask shape can change in detail when different images are used in making masks or when there is an error in face alignment. Therefore, in order to alleviate the effect of disturbance to some extent, it is better to group the selected pixels in making a mask as in Section 2.3. This step resulted in a small improvement (0.0–0.8%) in the experiments in Section 3 depending on the database used.

Through the toy example in Section 2.4, we showed that the proposed method can increase the PSNR of the reduced image and also maintain most of the discriminative information. Therefore, eliminating the pixels with less discriminative information can be helpful to extract better features than using all the pixels. In face recognition, the effect of the pixel selection was more pronounced in the CMU-PIE, Yale B and AR databases, which had large variations of various types, compared to the FERET database. This indicates that the proposed method effectively eliminates the pixels that may degrade the recognition performance due to environmental variations. By comparing the proposed method with several types of images and input variable selection methods, we

showed that the proposed method performed well for various types of variations. Moreover, the proposed method can be applied to any feature extraction method, and is expected to give additional performance improvement as demonstrated in the experiments.

The pixel selection method allows significant computational saving depending on the size of the reduced image. Reducing the computational complexity and image size have become more important in recent days when many applications with images are used in various mobile devices such as laptops or cellular phones.

The basic concept of the proposed method can be easily extended as an input variable selection method to other pattern classification problems, which will be the topic for future work.

Acknowledgment

This work was supported by Mid-career Researcher Program through NRF grant funded by the MEST (400-20100014).

References

- Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J., 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Machine Intell.* 19 (7), 711–720.
- Brunelli, R., Poggio, T., 1993. Face recognition: Features versus templates. *IEEE Trans. Pattern Anal. Machine Intell.* 15 (10), 1042–1052.
- Cevikalp, H., Neamtu, M., Wilkes, M., Barkana, A., 2005. Discriminative common vectors for face recognition. *IEEE Trans. Pattern Anal. Machine Intell.* 27 (1), 914–919.
- Chen, L.F., Liao, H.Y.M., Ko, M.T., Lin, J.C., Yu, G.J., 2000. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognit.* 33, 1713–1726.
- Choi, S.I., Choi, C.H., Jeong, G.M., 2008. Pixel selection in a face image based on discriminant features for face recognition. In: *Proc. Internat. Conf. on Automatic Face and Gesture Recognition*.
- Choi, S.I., Kim, C., Choi, C.H., 2011. Face recognition based on 2D images under illumination and pose variations. *Pattern Recognition Lett.* 32, 561–571.
- Fukunaga, K., 1990. *Introduction to Statistical Pattern Recognition*, 2nd ed. Academic Press.
- Georghiadis, A.S., Belhumeur, P.N., 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Machine Intell.* 23 (23), 643–660.
- Gokberk, B., Irfanoglu, M.O., Akarun, L., Alpaydin, E., 2007. Learning the best subset of local features for face recognition. *Pattern Recognit.* 40 (5), 1520–1532.
- Gonzales, R.C., Woods, R.E., 2002. *Digital Image Processing*, 2nd ed. Prentice Hall.
- Gottumukkall, R., Asari, V.K., 2004. An improved face recognition technique based on modular PCA approach. *Pattern Recognition Lett.* 25 (4), 429–436.
- Gunal, S., Edizkan, R., 2008. Subspace based feature selection for pattern recognition. *Inform. Sci.* 178 (19), 3716–3726.
- He, X., Cai, D., Niyogi, P., 2006. Laplacian score for feature selection. *Adv Neural Inform. Process. Syst.* 18, 507–514.

- Jain, A.K., Duin, R.P.W., Mao, J., 2000. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Machine Intell.* 22 (1), 4–37.
- Jiang, X., Mandal, B., Kot, A., 2008. Eigenfeature regularization and extraction in face recognition. *IEEE Trans. Pattern Anal. Machine Intell.* 30 (3), 1–12.
- Kim, C., Choi, C.H., 2007. Image covariance-based subspace method for face recognition. *Pattern Recognit.* 40 (5), 1592–1604.
- Kim, C., Oh, J., Choi, C.H., 2005. Combined subspace method using global and local features for face recognition, In: *Proc. Internat. Joint Conf. on Neural Networks*. pp. 2030–2035.
- Kokiopoulou, E., Frossard, P., 2006. Classification-specific feature sampling for face recognition, In: *Proc. IEEE Workshop on Multimedia Signal Processing*. pp. 20–23.
- Liang, J., Yang, S., Winstanley, A., 2008. Invariant optimal feature selection: A distance discriminant and feature ranking based solution. *Pattern Recognit.* 41 (5), 1429–1439.
- Liu, C., Wechsler, H., 2002. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. Image Process.* 11 (4), 467–476.
- Martnez, A.M., Benevente, R., 1998. The AR face database, CVC technical report #24, Computer Vision Center (CVC), Barcelona, Spain.
- Ocegueda, O., Shah, S.K., Karadiaris, I.A., 2011. Which parts of the face give out your identity? In: *Proc. Internat. Conf. on Computer Vision and Pattern Recognition*.
- Pang, Y., Zhang, L., Li, M., Liu, Z., Ma, W., 2004. A novel gabor-LDA based face recognition method. *Lecture Notes Comput. Sci.* 3331, 352–358.
- Pentland, A., Moghaddam, B., Starner, T., 1994. View-based and modular eigenspaces for face recognition, In: *Proc. Internat. Conf. on Computer Vision and Pattern Recognition*. pp. 84–91.
- Phillips, P., Moon, H., Rizvi, S., Rauss, P., 2000. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Machine Intell.* 22 (10), 1090–1104.
- Rajagopalan, A.N., Rao, K.S., Kumar, Y.A., 2007. Face recognition using multiple facial features. *Pattern Recognition Lett.* 28 (3), 335–341.
- Sim, T., Baker, S., Bsat, M., 2003. The CMU pose, illumination, and expression database. *IEEE Trans. Pattern Anal. Machine Intell.* 25 (12), 1615–1618.
- Sinha, P., Balas, B., Ostrovsky, Y., Russell, R., 2006. Face recognition by humans: Nineteen results all computer vision researchers should know about. In: *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948–1962.
- Tan, K., Chen, S., 2005. Adaptively weighted sub-pattern PCA for face recognition. *Neurocomputing* 64, 505–511.
- Tan, X., Chen, S., Zhou, Z.H., Zhang, F., 2006. Face recognition from a single image per person: A survey. *Pattern Recognit.* 39 (9), 1725–1745.
- Turk, M., Pentland, A., 1991. Eigenfaces for recognition. *J. Cognitive Neurosci.* 3 (1), 71–86.
- Wang, S.J., Zhou, C.G., Chen, Y.H., Peng, X.J., Chen, H.L., Wang, G., Liu, X., 2011. A novel face recognition method based on sub-pattern and tensor. *Neurocomputing* 74, 3553–3564.
- Yang, M., Chen, Y.J., Ji, G.L., 2010. Semi-fisher score: A semi-supervised method for feature selection. In: *Proc. Internat. Conf. on Machine Learning and Cybernetics*, pp. 527–532.
- Yang, P., Shan, S., Gao, W., Li, S.Z., Zhang, D., 2004. Face recognition using ada-boosted gabor features. In: *Proc. Internat. Conf. on Automatic Face and Gesture Recognition*, pp. 356–361.
- Zou, J., Ji, Q., Nagy, G., 2007. A comparative study of local matching approach for face recognition. *IEEE Trans. Image Process.* 16 (10), 2617–2628.