FACE RECOGNITION USING LOCAL STATISTICS OF GRADIENTS AND CORRELATIONS

Ying Ai Ju, Hyun Joo So, Nam Chul Kim, and Mi Hye Kim

School of Electronics Engineering, Kyungpook National University
1370 Sankyukdong, Bookgu, 702-701, Daegu, Korea
phone: + (82) 53 950 5530, fax: + (82) 53 950 5505, email: nckim@knu.ac.kr
web: vcl.knu.ac.kr

ABSTRACT

Most of face recognition methods often use a raw image itself for a feature vector. However, the feature vector directly formed from a raw image is seemed to be susceptible to variation of illumination and facial expression. In this paper, we propose a face recognition method using local statistics of gradients and correlations. BDIP (block difference of inverse probabilities) is chosen as a local statistics of gradients and two types of BVLC (block variation of local correlation coefficients) as local statistics of correlations. When a test image enters the system, it extracts the three types of feature vectors, fuses them, and classifies the image by using whitened PCA process and cosine distance. Experimental results for the three face DBs, Yale, Yale B, and Weizmann, show that the fused features of BDIP and BVLCs are more robust to variation of illumination and facial expression and so the proposed method yields good results.

1. INTRODUCTION

Over the past few decades, face recognition has received significant attention because of its wide applications in entertainment, information security, law enforcement, and surveillance, and so on [1]. One of the most simple and popular methods is eigenface technology [2], which is based on principal component analysis (PCA). It includes a linear core process that projects the high-dimensional data onto a lower dimensional space, based on second-order dependencies. Bartlett et al. further indicated that important information on face recognition may be contained in high-order relationships among facial pixels and hence presented two different independent component analysis (ICA) architectures, which are shown to outperform PCA [3]. However, Yang et al. claimed that the two ICA Architectures involve PCA process, whitening process, and pure ICA projection, and showed that pure ICA projection has only a little effect on the performance of face recognition [4]. Besides, Hsieh et al. proposed a hybrid approach based on sub-pattern technique and whitened PCA (WPCA) [5].

One of the most important factors to degrade the performance of face recognition is known to be the illumination variation problem. Many methods have been suggested to solve this problem. One of the effective approaches is trying to extract illumination invariant features. Based on a quotient image (QI) [6], which is designed for dealing with illumination variation, Wang et al. proposed a self-quotient image (SQI) model to extent the QI theory [7]. Recently, Zhang et al. proposed a gradientfaces method as a preprocessing technique for face recognition under varying lighting. They concluded that traditional PCA methods ignore the underlying relationship between neighboring pixels but the gradient domain considers the relationship between neighboring pixels and can reveal underlying inherent structure of image data [8].

Related to the illumination variation problem, it is also worthy of notice that BDIP (block difference of inverse probabilities) and BVLC (block variation of local correlation coefficients) operators, which have been applied to image retrieval [9], [10], face detection [11], ROI determination [12], and texture classification [13], and yielded very good results. Both of the operators are bounded and well locally normalized to be robust to illumination variation. BDIP is a kind of nonlinear operator normalized by local maximum, which is known to effectively measure local bright variations. BVLC is a maximal difference between local correlations according to orientations normalized by local variance, which is known to measure texture smoothness well [13].

In this paper, we apply the two operators to extracting three types of facial features. The fusion of the three features is first transformed by WPCA and then classified by the nearest neighbour classifier with cosine distance. The results show that the proposed method yields quite good results. The rest of this paper is organized as follows. Section 2 will give a simple description of WPCA and explain some facial features. The proposed method is described in section 3 and the experimental results in section 4. Finally, the conclusion is shown in section 5.

2. FACE RECOGNITION USING WPCA AND FACIAL FEATURES

In this section, we will describe a typical face recognition method using WPCA and explain some facial features utilized in face recognition and image retrieval areas.

2.1 Overview of face recognition using WPCA

Figure 1 shows the block diagram of a typical face recognition using WPCA. In the training phase, the feature vectors are first extracted from training images in a DB and their mean vector and WPCA matrix are computed.

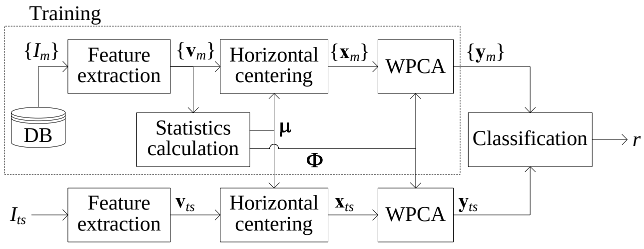


Figure 1 – Block diagram of a typical face recognition using whitened PCA.

The training feature vectors are next horizontally centered and finally transformed by WPCA to get transformed feature vectors. In the testing phase, a test feature vector is extracted from a test image, horizontally centered, transformed by WPCA, and compared with the transformed training feature vectors to obtain a classification result.

Suppose that there are M training feature vectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\cdots$, $\mathbf{v}_M$ which are extracted from training images $I_1, I_2, \cdots, I_M$. The mean vector and covariance matrix are written as

$$\boldsymbol{\mu} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{v}_m \tag{1}$$

and

$$\mathbf{S} = \frac{1}{M}\sum_{m=1}^{M}(\mathbf{v}_m - \boldsymbol{\mu})(\mathbf{v}_m - \boldsymbol{\mu})^T = \frac{1}{M}\mathbf{X}\mathbf{X}^T \tag{2}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M]$ stands for matrix consisting of horizontally centered vectors $\mathbf{x}_m = \mathbf{v}_m - \boldsymbol{\mu}$ for $m = 1, 2, \cdots, M$.

Let us assume we obtain the set of M largest $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$ of $\mathbf{S}$ and their eigenvector $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \cdots, \boldsymbol{\psi}_M$ [2]. Selecting n largest eigenvalues where $n \leq M$, the horizontally centered vectors are then transformed by the WPCA as follows:

$$\mathbf{y}_m = \boldsymbol{\Phi}^T\mathbf{x}_m = \left[\frac{\boldsymbol{\psi}_1}{\sqrt{\lambda_1}}, \frac{\boldsymbol{\psi}_2}{\sqrt{\lambda_2}}, ..., \frac{\boldsymbol{\psi}_n}{\sqrt{\lambda_n}}\right]^T \mathbf{x}_m \tag{3}$$

for $m = 1, 2, \cdots, M$.

In the testing phase, for the vector $\mathbf{v}_{ts}$ extracted from a test image I_{ts}, the transformed vector $\mathbf{y}_{ts}$ is calculated by

$$\mathbf{y}_{ts} = \boldsymbol{\Phi}^T\mathbf{x}_{ts} = \boldsymbol{\Phi}^T(\mathbf{v}_{ts} - \boldsymbol{\mu}). \tag{4}$$

The cosine distance between the mth transformed training and test vectors is then computed by

$$c_m = \frac{\mathbf{y}_{ts}\cdot\mathbf{y}_m}{\|\mathbf{y}_{ts}\|\,\|\mathbf{y}_m\|}, \quad m = 1, 2, \cdots, M \tag{5}$$

Finally, the test image is classified to the class of the rth training image which gives the maximum distance as

$$r = \arg\max_{m\in\{1,2,...,M\}} c_m. \tag{6}$$

2.2 Facial features

Up to now various face recognition methods have been suggested, most of which use a raw image itself as a feature vector. However, a raw image is seemed to be susceptible to variation of illumination and facial expression. In this section, we thus introduce more robust features useful for face recognition.

2.2.1 Gradientface

A gradientface is defined as [8]

$$F = \tan^{-1}\left(\frac{I * \nabla_y G_\sigma(x, y)}{I * \nabla_x G_\sigma(x, y)}\right) \tag{7}$$

where $\nabla_y G_\sigma(x, y)$ and $\nabla_x G_\sigma(x, y)$ are the derivatives of Gaussian kernel function in the x and y direction, respectively. The values of a gradientface should be in a period of angles.

2.2.2 BDIP

BDIP for an image I is defined as

$$D_p = \frac{<\hat{I}_p - I_{p+q}>_R}{\hat{I}_p} = 1 - \frac{\bar{I}_p}{\hat{I}_p} \tag{8}$$

where I_p denotes intensity at a pixel p of an image I, $<\cdot>_R$ the averaged value over the pixels q's in a moving window R, and $\hat{I}_p$ and $\bar{I}_p$ stand for the maximum value and mean value over the window whose center is at p, respectively. Since the quantity within $<\cdot>_R$ means the gradient of a pixel, D_p implies the mean of normalized gradients over the local region whose center is at p. As it is normalized by the local maximum, it is expected to be robust to variation of illumination. For stabilization, the denominator in (8) is clipped as $\hat{I}_p = \max(\hat{I}_p, \delta_D)$.

2.2.3 BVLC

BVLC for an image I is defined as

$$C_p = \max_{d\in O}\rho_p(d) - \min_{d\in O}\rho_p(d) \tag{9}$$

where $\rho_p(d)$ is the local correlation coefficient along a direction d at a pixel p. It is defined as

$$\rho_p(d) = \frac{<I_{p+d+q}I_{p+q} - \bar{I}_{p+d}\bar{I}_p>_R}{\sqrt{Var(I_{p+d})Var(I_p)}} \quad \text{for } d \in O \tag{10}$$

where $\bar{I}_p$ and $Var(I_p)$ stand for the mean and variance over

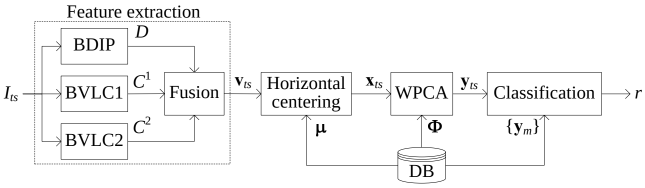


Figure 2 – Block diagram of the proposed face recognition using whitened PCA.

the window R whose center is at p, respectively. $\bar{I}_{p+d}$ and $Var(I_{p+d})$ denote the mean and variance of the moving window R whose center is at the pixel $p+d$, respectively. The symbol O denotes a set of orientations, which may be chosen as $O = \{(-k, 0), (k, 0), (0, -k), (0, k)\}$. Since $\rho_p(d)$ means the correlation coefficient along a direction d, C_p in (11) implies the maximum deviation of correlation coefficients over the local region whose center is at p. As it is normalized by local standard deviations, it is also expected to be robust to variation of illumination. For stabilization, the variances in the denominator of (10) are clipped with δ_V.

3. PROPOSED METHOD

In this section, we will describe our face recognition method which is shown in Figure 2. When a test image I_{ts} enters the system, it first extracts three types of features and fuses the three into a feature vector $\mathbf{v}_{ts}$. Next, it obtains the horizontally centered vector $\mathbf{x}_{ts}$ and the transformed feature vector $\mathbf{y}_{ts}$ by performing horizontal centering and WPCA process. The system finally classifies the test feature vector by comparing it with the training feature vectors in a database (DB).

One of three types of features used in the proposed method is the BDIP defined in (8) and the others are the two types of BVLC. In order to distinguish them from each other, we redefine the BVLC in terms of the distance k as follows:

$$C_p^k = \max_{d \in O_k} \rho_p(d) - \min_{d \in O_k} \rho_p(d) \qquad (11)$$

where O_k denotes a set of four orientations according to k. For simplicity, we call them BVLC1and BVLC2 in case of $k = 1$ and $k = 2$, respectively.

Examples of BDIP, BVLC1, and BVLC2 feature images are illustrated in Figure 3. The first column (a) consists of four original images taken from the Yale B and Weizmann [5]. The former is chosen for lighting variant experiment and the latter for lighting plus expression variant experiment. The first two original images come from Subset 1 and Subset 4 of Yale B DB and the last two from the training set and Subset 3 of Weizmann DB. The second column (b) corresponds to the BDIP images, the third (c) to the BVLC1 images, the fourth (d) to the BVLC2 images.

We can see from Figure 3 that BDIP, BVLC1, and BVLC2 images are shown to be different features from raw images. It is shown that BDIP can extract sketch-like feature images, where edges and valleys around the eyes and lips are more emphasized both for the normal image and the shadowy image. We also see that BVLC1 and BVLC2 can extract features around eyes, noses, and lips region well. Since the texture features BDIP, BVLC1, and BVLC2 are normalized well, all of them seem helpful to overcome variation of illumination. In addition, even though facial expressions change, the property of facial textures does not change so much, so that all of them look less sensitive to variation of expression.

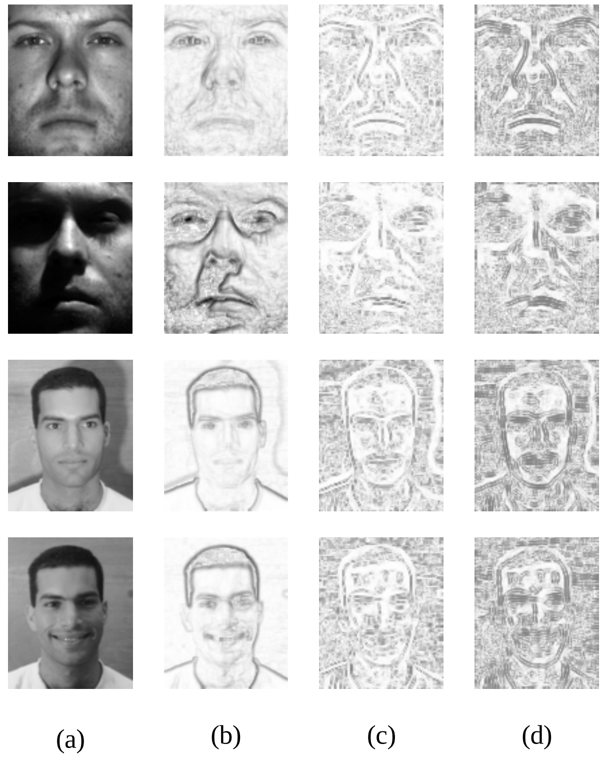


Figure 3 – Examples of BDIP, BVLC1, and BVLC2 feature images. (a) Original images, (b) BDIP images, (c) BVLC1 images, (d) BVLC2 images.

Extracting BDIP, BVLC1, and BVLC2 images from a test image, the system forms the test feature vector $\mathbf{v}_{ts}$ by fusing the three feature images. That is, it is written as

$$\mathbf{v}_{ts} = [\mathbf{v}_D^T, \mathbf{v}_{C^1}^T, \mathbf{v}_{C^2}^T]^T \qquad (12)$$

where $\mathbf{v}_A$ denotes the feature vector formed from a feature image $A \in \{D, C^1, C^2\}$. As a result, the dimension of the fused feature vector becomes three times larger than that of a single feature vector.

4. EXPERIMENTAL RESULTS

In this section, the performance of the proposed approach is evaluated with three face DBs: Yale, Yale B, and Weizmann. The facial parts of images in Yale and Yale B are cropped and resized to images of 112×92 pixels and those in Weizmann are resized to images of 112×92 pixels without cropping. The further detailed description can be found in [5].

For Yale, two types of experiments are performed. In the expression variant test, a normal expression for each of 15 persons is used for a training image and five expressions (happy, sad, sleepy, surprised, and winking) for each person are chosen for test images. In the leave out one test, every one of 11 expression images for each of 15 persons is selected for a test image and the other 10 images for training images.

For Yale B, a lighting variant experiment is executed. A training set consists of 190 images of 10 persons in Subset 1 and Subset 2, and the test set for the experiment contains 140 images in Subset 4. For Weizmann DB, an expression plus lighting variant experiment is performed. A training set consists of 130 images of 26 persons, and the test set for experiment contains 520 images in Subset 3.

For performance comparison, we implement not only our method (BCIP+BVLC1+BVLC2) but also other methods using raw image, gradientface, BDIP, BVLC1, and BVLC2 respectively. As for gradientface, the parameter σ in Gaussian kernel is set to 0.5. As for BDIP and BVLCs, the clipping threshold δ_D is set to 2 and δ_V to 0.001, respectively. Their moving window whose size is chosen as 3×3 is slid pixel by pixel. As a result, BDIP, BVLC1, BVLC2 images have the same size as the raw image.

The performance of face recognition is measured as the averaged recognition rate, which is defined as the ratio of the number of test images classified correctly to the number of all the test images. Figure 4 shows the recognition rates according to the number of selected eigenvectors and Table 1 lists the top recognition rates.

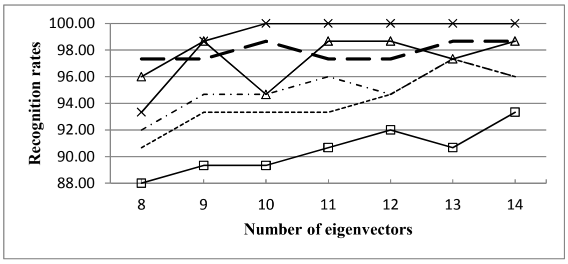
We can see from Table 1 that the performance of raw image feature is the best in the expression variant experiment of Yale but the worst in all the other experiments. The performance of all single features except raw image feature are better than that of raw image feature for all the experiments except for expression variant experiment of Yale. Among single features, BDIP and BVLC2 are the best for the leave out one experiment and BVLC1 is the best for expression variant tests of Weizmann. As a result, the fusion of BDIP and BVLCs are shown to be the best for all the experiments except for expression variant test of Yale. It also gives the gain of maximum 14.04% over raw image feature and the gain of maximum 5.34% over gradientface feature.

5. CONCLUSION

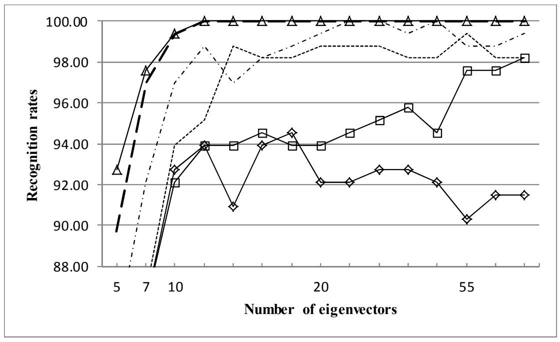
In this paper, we have proposed a face recognition method using the fusion of BDIP, BVLC1, and BVLC2. The fused feature vector for a test image was transformed by horizontal centering and WPCA and classified by the nearest neighbour classifier with cosine distance. Experimental results for three face DBs showed that our proposed method yielded the best results among eight methods and was robust to variation of illumination and facial expression.

6. ACKNOWLEDGEMENT

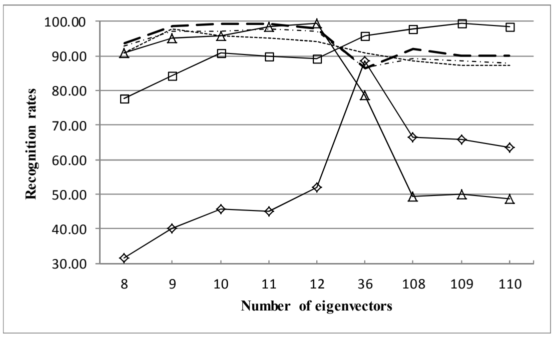
This work was supported by the Brain Korea 21 Project.



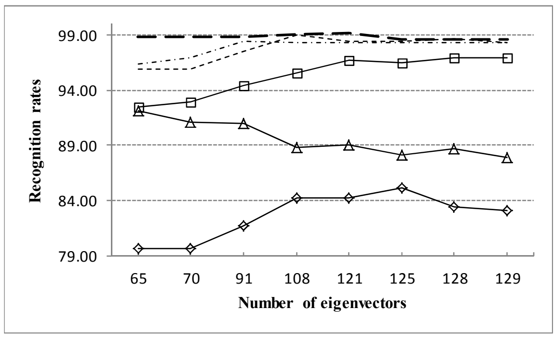
(a)



(b)



(c)



(d)

Figure 4 – Recognition rates according to the number of selected eigenvectors: (a) expression variant test of Yale DB, (b) leave one out test of Yale DB, (c) lighting variant test of Yale B DB, (d) expression plus lighting variant test of Weizmann DB.

Table 1 – Top recognition rates [%] according to various features for the three face DBs.

Feature \ DB	Yale		Yale B	Weizmann
	Expression variant	Leave out one	Lighting variant	Expression plus lighting variant
Raw	100.00	94.55	88.57	85.19
Gradientface	93.33	98.79	99.29	96.92
BDIP	98.67	100.00	99.29	92.12
BVLC1	97.33	99.39	97.86	99.04
BVLC2	97.33	100.00	97.86	98.46
BDIP+BVLC1+BVLC2	98.67	100.00	99.29	99.23

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, Dec. 2003.

[2] M. Turk and A. Pentland, “Face recognition using eigenfaces,” *Computer Vision and Pattern Recognition*, pp. 586-591, Jun. 1991.

[3] M. S. Bartlett, J. R. Movellean, and T. J. Sejnowski, “Face recognition by independent component analysis,” *IEEE Trans. Neural Networks*, vol. 13, no. 6, Nov. 2002.

[4] J. Yang, D. Zhang, and J. Y. Yang, “Is ICA significantly better than PCA for face recognition?” in *Proc. IEEE Int. Conf. Computer Vision*, Beijing, China, Oct. 17-21. 2005, vol. 1, pp. 198-203.

[5] P. C. Hsieh and P. C. Tung, “A novel hybrid approach based on sub-pattern technique and whitened PCA for face recognition,” *Pattern Recognition*, vol. 42, no. 5, pp. 978-984, May 2009.

[6] A. Shashua and T. Riklin-RaVIV, “The quotient image: Class-based rendering and recognition with varying illumination,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 129-139, Feb. 2001.

[7] H. Wang, S. Z. Li, Y. Wang, and J. Zhang, “Self quotient image for face recognition,” in *Proc. IEEE Int. Conf. Image Processing*, Singapore, Oct. 24-27. 2004, vol. 2, pp. 1397-1400.

[8] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, “Face recognition under illumination using gradientfaces,” *IEEE Trans. Image Processing*, vol. 18, no. 11, pp. 2599-2606, Nov. 2009.

[9] Y. D. Chun, N. C. Kim, and I. H. Jang, “Content-based image retrieval using multiresolution color and texture features,” *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1073-1084, Oct. 2008.

[10] Y. D. Chun, S. Y. Seo, and N. C. Kim, “Image retrieval using BDIP and BVLC moments,” *IEEE Trans. Circuits Syst. for Video Technology*, vol. 13, no. 9, pp. 951-957, Sep. 2003.

[11] H. J. So, M. H. Kim, Y. S. Chung, and N. C. Kim, “Face detection using sketch operators and vertical symmetry,” *FQAS 2006*, *Lecture Notes in Artificial Intelligence*, vol. 4027, pp. 541-551, Jun. 2006.

[12] T. D. Nguyen, S. H. Kim, and N. C. Kim, "An automatic body ROI determination for 3D visualization of a fetal ultrasound volume," *KES 2005*, *Lecture Notes in Artificial Intelligence*, vol. 3682, no. 2, pp. 145-153, Sep. 2005.

[13] H. J. So, M. H. Kim, and N. C. Kim, “Texture classification using wavelet-domain BDIP and BVLC features,” in *Proc. EUSIPCO2009*, Glasgow, Scotland, Aug. 24-28. 2009, pp. 1117-1120.