PAPER

Two-Stage Block-Based Whitened Principal Component Analysis with Application to Single Sample Face Recognition

Biao WANG[†a)], Wenming YANG[†b)], *Nonmembers*, Weifeng LI[†c)], *Member*, *and* Qingmin LIAO[†*d)], *Nonmember*

SUMMARY In the task of face recognition, a challenging issue is the one sample problem, namely, there is only one training sample per person. Principal component analysis (PCA) seeks a low-dimensional representation that maximizes the global scatter of the training samples, and thus is suitable for one sample problem. However, standard PCA is sensitive to the outliers and emphasizes more on the relatively distant sample pairs, which implies that the close samples belonging to different classes tend to be merged together. In this paper, we propose two-stage block-based whitened PCA (TS-BWPCA) to address this problem. For a specific probe image, in the first stage, we seek the K-Nearest Neighbors (K-NNs) in the whitened PCA space and thus exclude most of samples which are distant to the probe. In the second stage, we maximize the "local" scatter by performing whitened PCA on the K nearest samples, which could explore the most discriminative information for similar classes. Moreover, block-based scheme is incorporated to address the small sample problem. This two-stage process is actually a coarse-to-fine scheme that can maximize both global and local scatter, and thus overcomes the aforementioned shortcomings of PCA. Experimental results on FERET face database show that our proposed algorithm is better than several representative approaches.

key words: *face recognition, one sample problem, principal component analysis, whitening transform, K-Nearest Neighbors*

1. Introduction

Automatic face recognition has remained an extensively studied topic during the last several decades. Lots of effective approaches have been proposed to date [1]. Provided with large representative training samples, machine vision can surpass human vision on the recognition of frontal faces [2]. However, one challenging problem for machine based face recognition technology is the poor generalization ability from one training sample per person, which is the well-known *one sample problem* [3]. The lack of adequate training samples often results in misclassification for probe images with aging, illumination and pose variations. However, due to the laborious effort to collect multiple samples and the heavy cost to store and process them, it's common that only a single sample for each person is stored in many face recognition applications, e.g., surveillance identification, forensic identification and access control. Thus, one sample problem has emerged as an active topic in the face recognition community.

In recent years, a variety of approaches have been proposed to address the one sample problem. They can be roughly divided into two categories. The first category takes advantage of a generic face database consisting of multiple training samples per person, which are different from the gallery subjects [4], [5]. However, the underlying assumption that feature transformation learned from the generic face database can be used to extract the discriminative features of variable unseen probes is arguable [6].

Without resorting to the generic face database, the second category takes advantage of the existing training set which consists of a single sample for each subject, either to directly extract robust features as matching templates [7], [8] or to learn a low-dimensional feature space [5], [9]. Eigenfaces (PCA with L2 norm as the similarity measure) tries to seek a low-dimensional representation of data such that the global scatter is maximized [9]. It is the first effective technology for modern face recognition and is usually set as a baseline for performance evaluation. Various extensions have been proposed to further improve the performance of standard PCA. For example, Wu et al. [10] proposed projection-combined PCA ($(\mathrm{PC})^2\mathrm{A}$) by performing PCA on a combination of original image and its first-order projection map. Zhang et al. [11] proposed Enhanced $(\mathrm{PC})^2\mathrm{A}$ by including a second-order projection map. Recently, whitened PCA (WPCA) with the cosine distance as the similarity measure has proved to be significantly better than Eigenfaces [5], [12], [13]. The whitening transform treats variances along all principal component axes as equally significant and scales each principal direction according to the corresponding eigenvalue to uniform the spread of the data and is arguably appropriate for single sample problem [13]. Two-dimensional PCA (2DPCA) uses 2D image matrices rather than 1D vectors for covariance matrix estimation [14]. However, 2DPCA is not suitable for the local descriptors, such as Gabor wavelets and LBP, which have been shown to be more discriminative than pixel intensity [7], [8]. In [15], the authors argue that by dividing an image into several non-overlapped blocks and performing whitened PCA blockwisely, one will get significant performance improvement. The underlying reason may be that block-based PCA is suitable for small sample size problem. Moreover, 2DPCA has proved to be a special case of block-based PCA [16].

In this paper, we propose a novel global-to-local, coarse-to-fine scheme to address the one sample problem

Manuscript received June 23, 2011.
Manuscript revised September 30, 2011.
[†]The authors are with the Visual Information Processing Laboratory, Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University, Beijing 100084, China.
[*]Correspondence author.
a) E-mail: wangbiao08@mails.tsinghua.edu.cn
b) E-mail: yang.wenming@sz.tsinghua.edu.cn
c) E-mail: li.weifeng@sz.tsinghua.edu.cn
d) E-mail: liaoqm@tsinghua.edu.cn
DOI: 10.1587/transinf.E95.D.853

for face recognition. Our work is inspired by the fact that to maximize global scatter, PCA emphasizes more on the relatively distant sample pairs and tends to merge the similar classes together, which is known as local confusion limitation [5]. To further improve the recognition performance, one should focus on the local concentrations consisting of the similar classes and map the close sample pairs to be distant. Therefore, a two-stage scheme is proposed. In the first stage, whitened transform is utilized to partially resolve the local confusion limitation of PCA. For a specific probe image, we search K-NNs in the whitened PCA space and exclude the distant samples that are unlikely to be the genuine subject. In the second stage, we perform WPCA on the local concentration formed by the K-NNs to uniform the pairwise distances. Furthermore, block-based scheme is incorporated to address the small sample size problem. Experiments demonstrated that the proposed TS-BWPCA algorithm outperforms Eigenfaces by large margins ranging form 12.2% to 34.0%.

2. Motivation and the Proposed Method

2.1 The Local Confusion Limitation of PCA

Given a set of N training samples $x_1, x_2, \cdots, x_N$, where x_i is a column vector in D-dimensional space, PCA tries to seek a low-dimensional representation $y_1, y_2, \cdots, y_N$ by projecting each x_i onto the projection axis w such that the global scatter is maximized, which could be formulated as follows:

$$\max_{\|w\|=1} w^T \Sigma w, \tag{1}$$

where Σ is the covariance matrix and defined as follows:

$$\begin{aligned}\Sigma &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})(x_i - \overline{x})^T \\ &= \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(x_i - x_j)(x_i - x_j)^T,\end{aligned} \tag{2}$$

As can be seen from Eq. (2), the covariance matrix describes the pairwise distance between any two samples. For face recognition with single sample per person, the PCA subspace can roughly maximize the pairwise distance between all subjects. From this respect, the PCA subspace is suitable for one sample problem, which agrees with the conclusion in [4] that when complex image variations are presented, many previous proposed methods perform no better than the simplest Eigenfaces method.

However, there is a fundamental limitation for PCA based single sample face recognition. The objective function illustrated in Eq. (1) emphasizes more on the relatively distant sample pairs, while the close sample pairs belonging to different classes tend to be merged together. This problem is first pointed out by Deng et al., and is known as "local confusion limitation" [5]. Figure 1 gives a simple intuitive illustration of the aforementioned phenomenon.

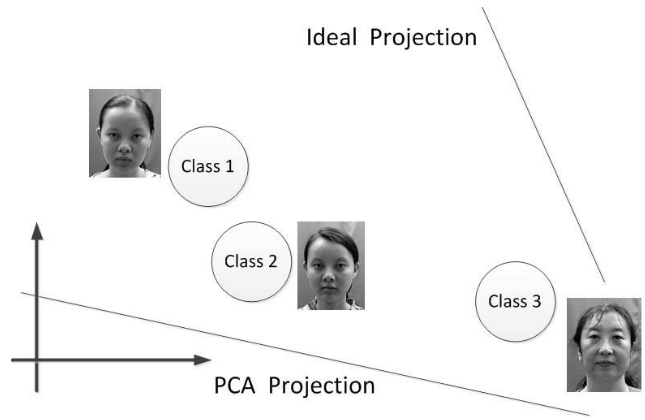


Fig. 1 The PCA subspace tends to merge similar classes together.

2.2 Whitened PCA

Despite of the local confusion limitation mentioned above, it's shown that the leading eigenvectors mainly encode the illumination and expression variation rather than discriminative information [17]. The whitening transform scales each principal direction with weights inversely proportional to the corresponding eigenvalue.

$$W_{wpca} = W_{pca}\tilde{D}^{-\frac{1}{2}}, \tag{3}$$

where the PCA projection matrix W_{pca} is the solution of Eq. (1), and usually formed by reserving P ($P < N - 1$) eigenvectors of Σ which correspond to the P largest eigenvalues. $\tilde{D}$ denotes the diagonal matrix of P eigenvalues. Consequently, the data in whitened PCA space tends to be uniform and the negative impact of the leading eigenvectors are reduced, while the discriminative information encoded in the trailing eigenvectors is emphasized.

For the purpose of classification, different choices of similarity measure have a significant impact on the performance. It has shown that compared with popular similarity measures like L1 distance, L2 distance, Mahalanobis distance, cosine similarity measure performs best in the whitened PCA space [12]. The WPCA algorithm for face recognition is summarized in Table 1.

2.3 Local Binary Pattern

Instead of using pixel intensity directly, current face recognition techniques often take advantage of low-level features such as local binary pattern (LBP) [5], [7], [18] or Gabor wavelet [8], [13], [19], which have been shown to be robust to the variations due to illumination and expression. The LBP operator was originally defined by encoding each pixel with 8 bit code, each of which is determined by thresholding the 3×3 neighborhood with the center pixel [20]. Formally, it could be described as follows:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} 2^n s(I_n - I_c), \tag{4}$$

Table 1 Description of whitened PCA algorithm.

Training Stage:
Input:
$X = [x_1, x_2, \cdots, x_N] \in R^{D\times N}$: Data matrix.
P: the number of largest eigenvectors to be reserved.
Output:
$\bar{x} \in R^{D\times 1}$: Data Mean.
$W_{wpca} \in R^{D\times P}$: the Transformation Matrix of Whitened PCA.
$\tilde{X} \in R^{P\times N}$: the low-dimensional representation of X.

1. Data Centering: $\hat{X} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_N]$,
   where $\hat{x}_i = x_i - \bar{x}$, and $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$.
2. Whitened PCA:
   Decompose $\hat{X}$ by singular value decomposition as follows:
   $$\hat{X} = UDV^T,$$
   Let $\tilde{D}$ be the diagonal matrix with P largest singular values,
   $\tilde{U}$ is the matrix of corresponding left singular vectors.
   $$W_{wpca} = \tilde{U}\tilde{D}^{-1},$$
   Whitened data matrix:
   $$\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_N] \in R^{P\times N}, \text{ in which } \tilde{x}_i = W_{wpca}^T \hat{x}_i.$$

Testing Stage:
Input:
$\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_N] \in R^{P\times N}$.
$\bar{x} \in R^{D\times 1}$: Data Mean.
$y \in R^{D\times 1}$: A probe feature vector.
Output:
label: the ID of the gallery the nearest to the probe.

1. Seek low-dimensional representation :
   $$\tilde{y} = (W_{wpca})^T (y - \bar{x}),$$
2. Classification:
   $$label = \arg\min_j similarity(\tilde{x}_j, \tilde{y}),$$
   in which $similarity(x, y) = -\frac{x^T y}{\|x\|\|y\|}$.

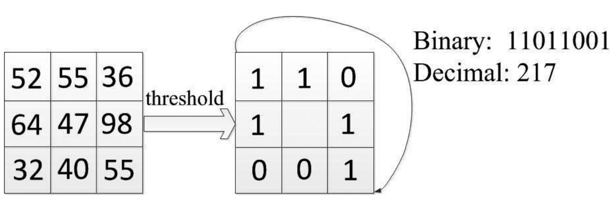


Fig. 2 The illustration of LBP encoding process.

in which (x_c, y_c) is the location of the center pixel, I_c and I_n are the intensity of the central pixel and its n-th neighbor, and $s(u)$ is 1 for $u \geq 0$ and 0 otherwise. The encoding process of LBP is illustrated in Fig. 2.

Two important extensions were proposed by Ojala *et al.* [21]. The first one extends LBP to multiscale by defining neighborhood of different sizes. The second defines the so-called *uniform patterns*: a LBP code is 'uniform' if it contains no more than two $0-1/1-0$ transitions. For example, the LBP code in Fig. 2 is non-uniform. LBP was first successfully applied to face recognition by Ahonen *et al.* in 2004 [7]. To encode both texture and structure information for human face, the LBP map (see Fig. 3) of a face image

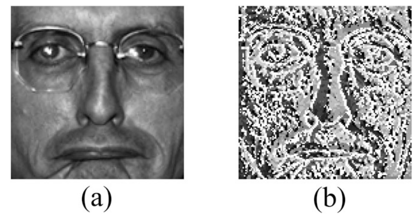


Fig. 3 (a) A sample face image and (b) its corresponding LBP map (uniform pattern).

is divided into several nonoverlapping blocks and histogram computed in each block is concatenated together to form the final representation. In this paper, we use LBP pattern by thresholding 8 neighboring pixels in a circle of radius 2 and extracted the histograms in 8×8 blocks with 59 bins, each corresponding to a uniform pattern. For more details, please refer to [7].

2.4 Two-Stage Block-based Whitened PCA

As pointed out in the above subsection, PCA is suitable for the application of single sample face recognition, but still suffers from the local confusion limitation which tends to merge the similar classes together. This is because in the process of seeking the maximum global scatter, PCA focuses more on the relatively distant samples. To obtain more discriminative information for the similar classes, one should focus more on the local concentrations consisting of similar samples and try to maximize the pairwise distances in the local concentrations. Therefore, we propose a two-stage whitened PCA (TS-WPCA) algorithm. In the first stage, for a specific probe, we perform WPCA to seek K-NNs as candidates and discard the distant ones (outliers), which would significantly reduce the side impact of local confusion limitation of PCA. The K nearest samples have high probability to include the genuine subject and are similar to each other, and thus form a local concentration. In the second stage, we perform WPCA with the K samples, which tries to maximize the pairwise distance among them and thus makes it easier to distinguish the similar subjects.

A simple intuitive illustration of the proposed method is given in Fig. 4. It is much like a "Coarse-to-Fine" scheme. It first maximizes global scatter and makes it easy to exclude distant subjects. Then it maximizes "local" scatter on the local concentrations formed by the similar classes. Another advantage of our proposed method is that the determination of local concentration is probe dependent. This is reasonable because for different probes, even if they belong to the same class, the most K confusable classes are different due to different illumination, expression, aging variations.

However, small sample size (SSS) problem, i.e., the extracted feature dimensionality is much larger than the number of training samples, still exists in the proposed method. This problem is even serious in the second stage WPCA because we discard most of the samples which are relatively distant to the probe. However, by dividing the original im-

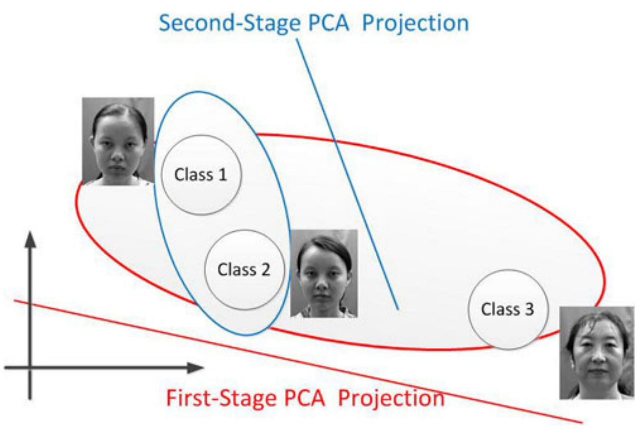


Fig. 4 A intuitive illustration of our proposed algorithm.

Table 2 Training stage of the proposed TS-BWPCA algorithm.

Input:
$T = \{I_1, I_2, \cdots, I_N\}$: the training set with N face images.
S: number of nonoverlapping blocks.
P: the number of largest eigenvectors to be reserved.
Output: $(i = 1, 2, \cdots, S)$
$\overline{h^i} \in R^{D\times 1}$: mean for the LBP histograms of i-th blocks.
$W^i_{wpca} \in R^{D\times P}$: WPCA transformation matrix for the i-th blocks.
$\tilde{H}^i \in R^{P\times N}$: low-dimensional projection for the i-th blocks.

1. For $I_j \in T(j = 1, 2, \cdots, N)$, compute its block-based LBP representation according to Steps 1.1 − 1.2.
1.1 Compute its LBP maps L_j.
1.2 Divide L_j into S nonoverlapping blocks $L^i_j(i = 1, 2, \cdots, S)$, for each block, further partition each block into several sub-blocks, each with size 8×8. For each sub-block, calculate the histogram and concatenate the histograms of all sub-blocks. one gets LBP representation for the ith block of the jth sample and denote it as h^i_j.
2. For each block i $(i = 1, 2, \cdots, S)$,
Data Centering: $\hat{H}^i = [\hat{h}^i_1, \hat{h}^i_2, \cdots, \hat{h}^i_N]$,
where $\hat{h}^i_j = h^i_j - \overline{h^i}$, and $\overline{h^i} = \frac{1}{N}\sum_{j=1}^{N} h^i_j$
3. First-Stage Whitened PCA:
Decompose $\hat{H}^i_j$ by singular value decomposition as follows:
$$\hat{H}^i = UDV^T,$$
Let $\tilde{D}$ be the diagonal matrix with P largest singular values,
$\tilde{U}$ is the matrix of corresponding left singular vectors.
$$W^i_{wpca1} = \tilde{U}\tilde{D}^{-1},$$
Whitened data matrix:
$$\tilde{H}^i = [\tilde{h}^i_1, \tilde{h}^i_2, \cdots, \tilde{h}^i_N] \in R^{P\times N},$$
in which $\tilde{h}^i_j = (W^i_{wpca1})^T \hat{h}^i_j$.

age into several nonoverlapping blocks and performing subspace analysis block-wisely, the SSS problem is much alleviated. This “divide and conquer” strategy is also used in [15], [19]. Therefore, we incorporate the block-based strategy into our algorithm and finally get the two-stage block-based whitened PCA scheme (TS-BWPCA). The detailed description of the training stage and testing stage for our proposed TS-BWPCA algorithm is summarized in Table 2 and Table 3. The block diagram is illustrated in Fig. 5.

Table 3 Testing stage of the proposed TS-BWPCA algorithm.

Input:
S: number of nonoverlapping blocks.
$\bar{h}^i \in R^{D\times 1}$: mean for the LBP histograms of i-th blocks.
$W^i_{wpca} \in R^{D\times P}$: WPCA transformation matrix for the i-th blocks.
$\tilde{H}^i \in R^{P\times N}$: low-dimensional projection for the i-th blocks.
V: A probe face image.
K: the number of nearest neighbors to search for each probe.
Output:
label: the ID of gallery which is the nearest to the probe.

1. For V, compute its block-based LBP representation y^i $(i = 1, 2, \cdots, S)$. This process is the same as in Step 1 for training stage.
Perform WPCA on y^i: $\tilde{y}^i = (W^i_{wpca1})^T(y^i - \bar{h}^i)$.
2. Seek the K-NNs of probe image V.
The distance of V to each gallery $I_j(j = 1, 2, \cdots, N)$, is determined by the summation of block-wise similarity:
$$dist(V, I_j) = \sum_{i=1}^{S} similarity(\tilde{y}^i, \tilde{h}^i_j),$$
in which $similarity(x, y) = -\frac{x^T y}{\|x\|\|y\|}$.
Denote $Q = [t_1, t_2, \cdots, t_K]$ as the labels of the K nearest neighbors for probe image V.
3. Second-Stage Whitened PCA:
For each $i(i = 1, 2, \cdots, S)$, denote $G^i = [\tilde{h}^i_{t_1}, \tilde{h}^i_{t_2}, \cdots, \tilde{h}^i_{t_K}]$.
$\tilde{G}^i = [\tilde{h}^i_{t_1} - \bar{g}^i, \tilde{h}^i_{t_2} - \bar{g}^i, \cdots, \tilde{h}^i_{t_K} - \bar{g}^i]$ and $\bar{g}^i = \sum_{k=1}^{K} \tilde{h}^i_{t_k}$.
Decompose $\tilde{G}^i$ by singular value decomposition as follows:
$$\tilde{G}^i = UDV^T,$$
Let $\tilde{D}$ be the diagonal matrix with $K - 1$ largest singular values,
$\tilde{U}$ is the matrix of corresponding left singular vectors.
$$W^i_{wpca2} = \tilde{U}\tilde{D}^{-1},$$
Whitened transformation of $\tilde{G}^i$:
$$\check{G}^i = [\check{h}^i_{t_1}, \check{h}^i_{t_2}, \cdots, \check{h}^i_{t_K}] \in R^{(K-1)\times K},$$
in which $\check{h}^i_{t_k} = (W^i_{wpca2})^T(\tilde{h}^i_{t_k} - \bar{g}^i)$.
Whitened transformation of $\tilde{y}^i$: $\check{y}^i = (W^i_{wpca2})^T(\tilde{y}^i - \bar{g}^i)$.
4. Classification:
$$label = \arg\min_{t_k} \sum_{i=1}^{S} similarity(\check{y}^i, \check{h}^i_{t_k}),$$
in which $similarity(x, y) = -\frac{x^T y}{\|x\|\|y\|}$.

3. Experimental Results

3.1 Experiment Setting

In this section, experiments are conducted on one publicly available large-scale face database, namely, FERET database to illustrate the effectiveness of our proposed method. All face images are properly aligned, cropped and resized to 128×128 with the centers of the eyes fixed at (29,34) and (99,34). No further preprocessing is performed.

We use the standard FERET protocol to conduct our experiments. The gallery set Fa consists of 1,196 images of 1,196 subjects. There are four probe sets: Fb (different expressions with gallery, 1,195 images of 1,196 subjects), Fc (different illumination conditions with gallery, 194 images of 194 subjects), Dup I (images taken later in time, 722 images of 243 subjects), Dup II (images taken at least 18 months after the corresponding gallery, 234 images of 75 subjects). Figure 6 shows samples of the same person from

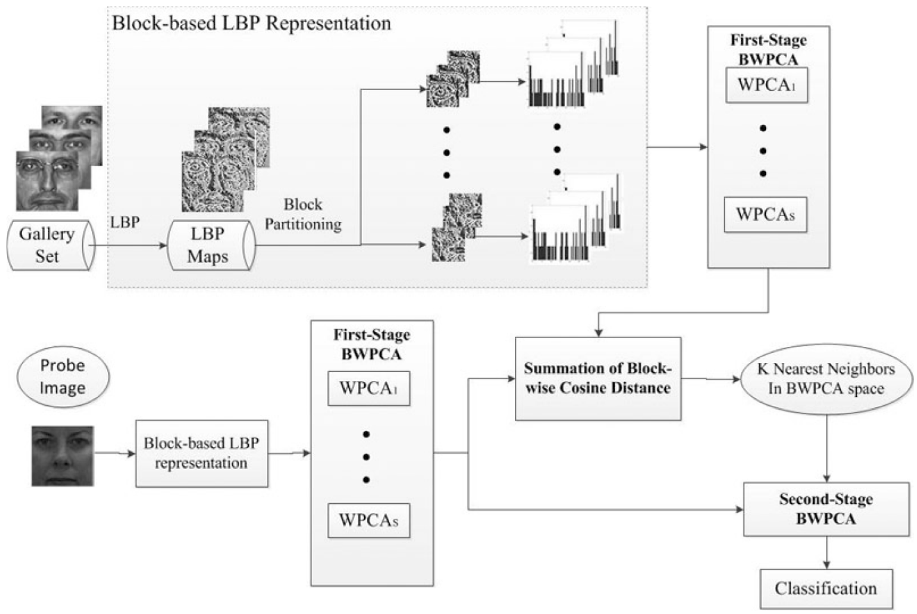


Fig. 5 Block diagram of our proposed TS-BWPCA algorithm.

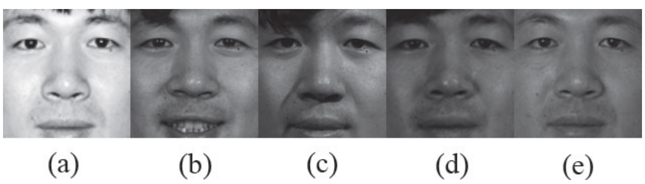


Fig. 6 Sample face images from the FERET database. (a) Fa (b) Fb (c) Fc (d) Dup I (e) Dup II.

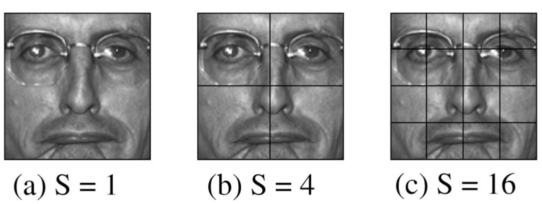


Fig. 7 Partition face image into S nonoverlapping blocks.

the five sets.

3.2 Performance Evaluation

As illustrated in Table 2 and Table 3, there are basically four parameters in our proposed method:

(a) S: number of nonoverlapping blocks. We will try three different block partitioning schemes: $S = \{1, 4, 16\}$, as illustrated in Fig. 7.

(b) P: dimensionality of first-stage WPCA. We will try $P = \{200, 300, 400, 500, 600\}$.

(c) K: number of nearest neighbors to search in the WPCA space. We will try $K = \{10, 20, 30, 40, 50\}$.

(d) Dimensionality of second-stage WPCA: it's set to be $K - 1$ as illustrated in Table 3.

We conducted face recognition experiments on FERET database with 75 ($3 \times 5 \times 5$) different configurations of parameters. We found that for $S = 1$ and $S = 4$, our proposed TS-BWPCA consistently improves the performance of block-based WPCA. However, for $S = 16$, TS-BWPCA performs even worse than block-based WPCA. The reason may be that by dividing face image into such small blocks (as shown in Fig. 7 (c)), we lose much structure information and thus it's meaningless to perform second-stage WPCA for each block. In our experiment, the best choice of P for $S = 4$ is $P_{opt} = 400$. The choice of proper K is critical in our proposed method. If this value is too small, there are too few train samples and the small sample size problem will be serious in the second-stage WPCA. If it's too large, relatively distant samples will be included and the local confusion limitation takes effect. Figure 8 shows the recognition rates with various choices of K. As can be seen, $K_{opt} = 20$ yields the highest recognition rate. However, the optimal value for K is actually expected to be dependent on the data set and it's not easy to develop a method to determine it theoretically. We empirically check this on CAS-PEAL database [22], which is another widely used face database, and find that K with a range from 15 to 40 always provides better performance than BWPCA, similar to that has illustrated in Fig. 8.

The experimental results are illustrated in Table 4. To demonstrate advantages of the proposed method, we compare it with WPCA [12] and block-based WPCA (BWPCA) [15]. Moreover, we also present the results of Eignefaces [9] and $LBP + \chi^2$ distance [7] as baselines.

From Table 4, we could draw the following conclusions:

(a) In the whitened PCA space, cosine similarity measure

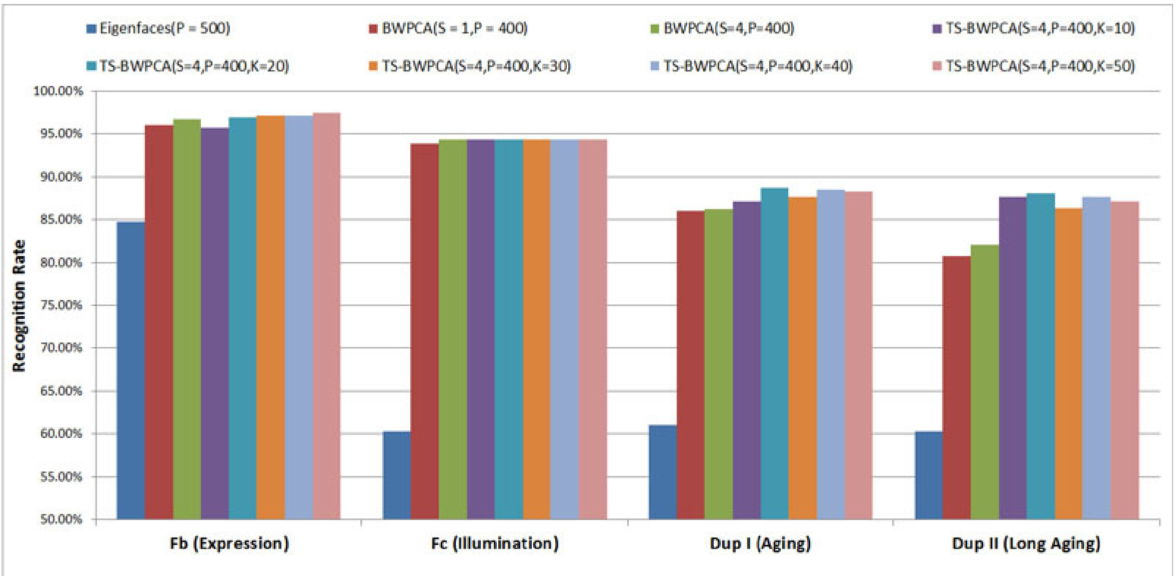


Fig. 8 Recognition rates vs. different choices of K.

Table 4 Recognition rates (%) of several basic approaches on FERET database.

Methods	Similarity Measure	Dimensionality	Fb	Fc	Dup I	Dup II
LBP	χ^2	15104	92.9	81.4	72.6	72.2
LBP + Eigenfaces	L2	500	84.9	61.3	61.5	60.7
LBP + PCA	Cos	500	85.6	46.4	66.5	55.6
LBP + WPCA	L2	500	76.4	59.8	51.9	54.3
LBP + WPCA	Cos	500	96.0	93.8	86.0	80.8
LBP + BWPCA	Cos	1600	96.7	94.3	86.2	82.1
LBP+TS-BWPCA	**Cos**	**1600**	**97.0**	**94.3**	**88.6**	**88.0**

performs much better than L2 distance. This agrees with the conclusion in [12].

(b) Among the four PCA related methods (i.e., TS-BWPCA, BWPCA, WPCA and PCA), TS-BWPCA achieves the best performance. Here we assume cosine similarity measure is adopted for each method. Whitening transform uniforms the data spread by scaling the principal axis, and thus could partially resolve the local confusion limitations of PCA. BWPCA weakens the small sample size problem by dividing images into several nonoverlapping blocks. Our proposed TS-BWPCA algorithm inherits their merits and further proposes a global-to-local, coarse-to-fine scheme to maximize both the global and local scatter.

(c) It seems that TS-BWPCA could improve the performance of aging face recognition. Compared with BW-PCA, TS-BWPCA improves the recognition rate of Dup I and Dup II from 86.2%, 82.1% to 88.6%, 88.0% respectively. It's known that both shape and appearance of human face will change as time elapses. By maximizing the local scatter in the second-stage, the variations due to aging will be covered.

Note that for Fb and Fc, we see little priority of TS-BWPCA to BWPCA, while significant improvement exists in Dup I and Dup II. This could be understood as follows. Variations corresponding to illumination and expression mostly change the prototypical representation of face images and can be regarded as low frequencies, while the fine details which are crucial for face recognition mainly correspond to high frequencies. The local limitation confusion actually implies the Mean-Square-Error principal underlying PCA preferentially weights low frequencies [23] and makes the discriminative information contained in the high frequency components contribute little for face recognition. WPCA resolve this problem by re-scaling each principal direction according to the corresponding eigenvalue and thus can effectively utilize the discriminative information in high frequencies. Gradual aging changes the fine details of face (e.g., wrinkles appear in forehead and the corner of the eyes) and mainly corresponds to high frequencies. Such variations corresponds to high frequencies may be overemphasized in the first-stage WPCA, which is harmful for robust aging-face recognition. Our proposed second-stage WPCA can select the most discriminative features among the local concentration formed by the K-NNs and remove these variations. We design another experiment to testify this assumption. We take a noise-polluted ('salt and pepper' noise, mainly corresponding to high frequencies) version of Fa as the probe set. Several sample images and their corresponding images with noise are illustrated in Fig. 9. The performance of the above four approaches are illustrated in Table 5, from which we can see that TS-BWPCA

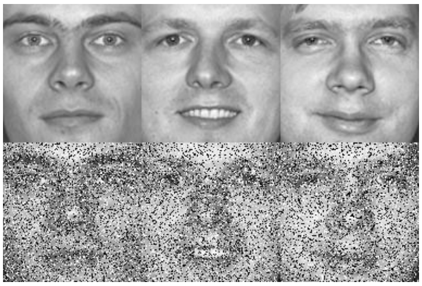


Fig. 9 Sample faces (the first row) and the corresponding images with 'salt & pepper' noise (the second row).

Table 5 Recognition rates (%) of several basic approaches on the noise-polluted probe set.

PCA	WPCA	BWPCA	TS-BWPCA
0.59	44.06	63.04	**70.48**

Table 6 Performance comparison of TS-BWPCA with several state-of-the-art approaches on FERET database.

Methods	Dim	Fb	Fc	Dup I	Dup II
FERET97 Best [25]	N/A	96.0	82.0	59.0	52.0
LBP [7]	2891	97.0	79.0	66.0	64.0
LGBPHS [8]	519200	98.0	97.0	74.0	71.0
HGPP [26]	737280	97.6	98.5	77.7	76.1
LGBP + WPCA [13]	< 1000	98.1	98.9	83.8	81.6
LBP + UP [5]	400	96.5	94.8	88.5	86.3
LBP+TS-BWPCA	**1600**	**97.0**	**94.3**	**88.6**	**88.0**

is indeed robust to high frequency variations and provides the best performance. More specifically, we improve the recognition rate of BWPCA from 63.04% to 70.48%, i.e., 89 people that were originally misclassified by BWPCA are now correctly recognized by our algorithm.

To further demonstrate the effectiveness of TS-BWPCA, we compare it with other state-of-the-art methods reported in literatures. The results are summarized in Table 6. As can be seen, TS-BWPCA outperforms all the others on the Dup I and Dup II probe set with relatively low dimensionality. Gabor (or Gabor + LBP) based method performs better than LBP on Fb and Fc probe set due to multi-scale representation. However, our TS-BWPCA is a general framework and could be easily incorprated with Gabor features. Moreover, for the face recognition in various illumination conditions, several effective illumination normalization methods have been proposed [18], [24]. But these are all beyond the concern of this paper.

However, a disadvantage of the proposed TS-BWPCA is the extra computation cost. For PCA, the most time-demanding operation is the eigenvalue decomposition. Typically, for an $N \times N$ matrix, this complexity is $O(N^3)$. When applied to rank-1 face recognition, the projection matrix learning of PCA, WPCA and BWPCA can be performed off-line and the time complexity is $O(NP)$ for nearest neighbor searching, in which N represents the number of samples and P is the projected dimension, respectively. However, for TS-BWPCA, the complexity for K-NNs searching and the second-stage WPCA are $O(NPK)$ and $O(K^3)$, respectively. Although we typically set $K << N$, the extra computation cost maybe restricts TS-BWPCA to the applications that are sensitive to the time issue.

4. Conclusion

The proposed TS-BWPCA method is designed to address the one sample problem for face recognition. Standard PCA maximizes the global scatter and focuses more on the distant sample pairs, which results in the so-called local confusion limitation. Whitening transform can partially reduce this side effect but it's not enough. We propose a two-stage scheme to address these limitations. For a specific probe, in the first stage, K nearest neighbors are searched in the whitened PCA space, thus excluding most of the distant classes. In the second stage, a second WPCA is performed to maximize the "local" scatter. Moreover, the block-based scheme is incorporated to address the small sample size problem. Experimental results demonstrated the advantages of our proposed method on high frequency variations. Although in this paper TS-BWPCA is proposed to address the one sample problem in face recognition. This global-to-local, coarse-to-fine scheme expects successfully applications to finger, palm recognition [27] and image retrieval [28].

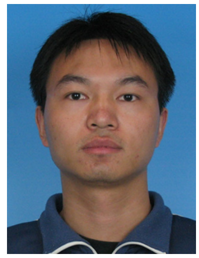
Acknowledgements

The authors would like to thank the anonymous reviewers for their critical and constructive comments and suggestions, which are helpful to improve both technical and the literary quality of this paper. We also thanks the FERET Technical Agent, the U.S. National Institute of Standards and Technology for providing the FERET database. Special thanks are given to Dr. W. Deng for his help to complete the implementation of the uniform pursuit algorithm. This work was supported by the Shenzhen-Hongkong Innovation Circle Project under grant No. ZYB200907070030A.

References

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face recognition: A literature survey," ACM Comput. Surv., pp.399–458, 2003.
[2] A.J. O'Toole, P.J. Phillips, F. Jiang, J. Ayyad, N. Penard, and H. Abdi, "Face recognition algorithms surpass humans matching faces over changes in illumination," IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.9, pp.1642–1646, 2007.
[3] X. Tan, S. Chen, Z.H. Zhou, and F. Zhang, "Face recognition from a single image per person: a survey," Pattern Recognit., vol.39, no.9, pp.1746–1762, 2006.
[4] J. Wang, J. Lua, K.N. Plataniotisa, and A.N. Venetsanopoulosa, "On solving the face recognition problem with one training sample per subject," Pattern Recognit., vol.39, no.9, pp.1746–1762, 2006.
[5] W. Deng, J. Hu, J. Guo, W. Cai, and D. Feng, "Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach," Pattern Recognit., vol.43, no.5, pp.1748–1762, 2010.
[6] J. Zou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," IEEE Trans. Image Process., vol.16, no.10, pp.2617–2628, 2007.

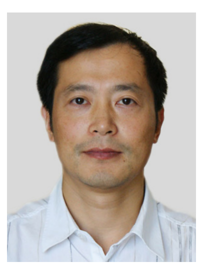
[7] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol.28, no.12, pp.2037–2041, 2006.
[8] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition," ICCV, 2005, pp.786–791, 2005.
[9] M.A. Turkand and A.P. Pentland, "Eigenfaces for recognition," J. Cognitive Neuroscience, vol.3, no.1, pp.71–86, 1991.
[10] J. Wu and Z.H. Zhou, "Face recognition with one training image per person," Pattern Recognit. Lett., vol.23, no.14, pp.1711–1719, 2002.
[11] D. Zhang, S. Chen, and Z.H. Zhou, "Enhanced $(pc)^2a$ for face recognition with one training image per person," Pattern Recognit. Lett., vol.25, no.10, pp.1173–1181, 2004.
[12] W. Deng, J. Hu, and J. Guo, "Gabor-eigen-whiten-cosine: A robust scheme for face recognition," AMFG, 2005, pp.336–349, 2005.
[13] Nguyen, V. Hieu, B. Li, and L. Shen, "Local gabor binary pattern whitened pca: A novel approach for face recognition from single image per person," Proc. Third International Conference on Advances in Biometrics, 2009.
[14] J. Yang and D. Zhang, "Two dimensional pca: a new approach to appearance-based face representation and recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol.26, no.1, pp.136–137, 2004.
[15] P. Hsieh and P. Tung, "A novel hybrid approach based on sub-pattern technique and whitened pca for face recognition," Pattern Recognit., vol.42, no.5, pp.978–984, 2009.
[16] L. Wang, X. Wang, X. Zhang, and J. Feng, "The equivalence of two-dimensional pca to line-based pca," Pattern Recognit. Lett., vol.26, no.1, pp.57–60, 2005.
[17] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," IEEE Trans. Pattern Anal. Mach. Intell., vol.19, no.7, pp.721–732, July 1997.
[18] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," IEEE Trans. Image Process., vol.19, no.6, pp.1635–1650, 2010.
[19] S. Xie, S. Shan, X. Chen, and J. Chen, "Fusing local patterns of gabor magnitude and phase for face recognition," IEEE Trans. Image Process., vol.19, no.5, pp.1349–1361, 2010.
[20] T. Ojala, Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," Pattern Recognit., vol.29, no.1, pp.51–59, 1996.
[21] T. Ojala, Pietikäinen, and T. Maenpaa, "Multiresolution gray-scale and rotation invarianat texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no.7, pp.971–987, 2002.
[22] W. Gao, B. Cao, and S. Shan, "The cas-peal large-scale face database and evaluation protocols," Tech. Rep. JDL, CAS, 2004.
[23] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," IEEE Trans. Pattern Anal. Mach. Intell., vol.19, no.7, pp.696–710, 2004.
[24] B. Wang, W. Li, W. Yang, and Q. Liao, "Illumination normalization based on weber's law with application to face recognition," IEEE Signal Process. Lett., vol.18, no.8, pp.462–465, 2011.
[25] P.J. Phillips, H. Moon, P. Rizvi, and P. Rauss, "The feret evaluation method for face recognition algorithms," IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.10, pp.1090–1104, 2000.
[26] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of gabor phase patterns (hgpp): a novel object representation approach for face recognition," IEEE Trans. Image Process., vol.16, no.1, pp.57–68, 2006.
[27] A.K. Jain and A. Ross, "Introduction to biometrics," in A. Handbook of Biometrics. pp.1–22, Springer, 2008.
[28] M.S. Lew, N. Sebe, C. Djeraba, and Ramesh. Jain, "Content-based multimedia information retrieval: State of the art and challenges," ACM Trans. Multimedia Comput. Commun. Appl., vol.2, pp.1–19, 2006.



Biao Wang was born in Huanggang, HuBei, China, in 1986. He received the B.E. degree from Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2008. He is currently working towards the Ph.D. degree in electronics engineering at Tsinghua University, Beijing, China. His research interests include applications of image processing and pattern recognition in biometrics.

Wenming Yang received the B.S. and M.S. in material science from Harbin University of Science & Technology, Harbin, China in 2000 and 2003, respectively. In 2006, he received the Ph.D. in electronic engineering from Zhejiang University, Hangzhou, China. From 2007 to 2009, He was a Postdoctoral Fellow in Department of Electronic Engineering, Tsinghua University. Since 2010, He has been a lecturer in Department of Electronic Engineering at Graduate School at Shenzhen, Tsinghua University. His research interests include image processing and pattern recognition, computer vision, biometrics, video surveillance, image super-resolution and non-destructive testing based on artificial vision.

Weifeng Li received the M.E. and Ph.D. degrees in Information Electronics at Nagoya University, Japan in 2003 and 2006, under the guidance of Prof. Fumitada Itakura and Prof. Kazuya Takeda. He joined the Idiap Research Institute, Switzerland in 2006, and in 2008 he move to Swiss Federal Institute of Technology, Lausanne (EPFL), Switzerland as a research scientist. Since 2010 he has been an associate professor in the Department of Electronic Engineering, Graduate School at Shenzhen, Tsinghua University, China. His research interests are in the areas of audio and visual signal processing, Biometrics, Human-Computer Interactions (HCI), and machine learning techniques. He is a member of the IEEE and IEICE.



Qingmin Liao was born in 1963. He received the B.S. degree in radio technology from the University of Electronic Science and Technology of China, Chengdu, China, in 1984, and the M.S. and Ph.D. degrees in signal processing and telecommunications from the University of Rennes 1, Rennes, France, in 1990 and 1994, respectively. Since 1995, he has been joining with Tsinghua university, Beijing, China. He became Professor in the Department of Electronic Engineering of Tsinghua University, in 2002. From 2001 to 2003, he served as the Invited Professor with a tri-year contract at the University of Caen, France. Since 2010, he has been Director of the Division of Information Science and Technology in the Graduate School at Shenzhen, Tsinghua University, Shenzhen, China. His research interests include image/video processing, transmission and analysis; biometrics; and their applications to teledetection, medicine, industry, and sports. He has published over 90 papers internationally.