



Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach

Weihong Deng^{a,*}, Jiani Hu^a, Jun Guo^a, Weidong Cai^b, Dagan Feng^b

^a School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, 100876 Beijing, China

^b School of Information Technologies, J12, The University of Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:

Received 20 July 2009

Received in revised form

25 October 2009

Accepted 5 December 2009

Keywords:

Face recognition

One sample problem

Principal component analysis

Whitening transformation

Uniform pursuit

ABSTRACT

Current face recognition techniques rely heavily on the large size and representativeness of the training sets, and most methods suffer degraded performance or fail to work if there is only one training sample per person available. This so-called “one sample problem” is a challenging issue in face recognition. In this paper, we propose a novel feature extraction method named uniform pursuit to address the one sample problem. The underlying idea is that most recognition errors are due to the confusions between faces that look very similar, and thus one can reduce the risk of recognition error by mapping the close class prototypes to be distant, i.e., uniforming the pairwise distances between different class prototypes. Specifically, the UP method pursues, in the whitened PCA space, the low dimensional projections that reduce the local confusion between the similar faces. The resulting low dimensional transformed features are robust against the complex image variations such as those caused by lighting and aging. A standardized procedure on the large-scale FERET and FRGC databases is applied to evaluate the one sample problem. Experimental results show that the robustness, accuracy and efficiency of the proposed UP method compare favorably to the state-of-the-art one sample based methods.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Human faces are arguably the most extensively studied object in image-based recognition. This is partly due to the remarkable face recognition capability of the human visual system and partly due to numerous important applications for face recognition technology. During the last two decades, there has been a significant effort to develop recognition algorithms from frontal face images, with a lot of encouraging results, even 100% accuracy [1,2], reported in the literature. A recent study showed that computer algorithms are capable of outperforming people on recognition of frontal face images [3], when large and representative training data set is available. These algorithms represent faces in the derived face feature space before matching by identity, and they won human may because of information they can exploit from the representative training images about the variability of individuals across changes in illumination [3].

However, machine based algorithms are still limited in the number of image variations they can generalize across. One of the most obvious differences between human and machine is the ability to learn from very few, even single, examples. After a close look at a certain face, people can memorize and recognize that

face in many unseen situations, such as new pose, lighting, and ages. In contrast, with limited training examples, state-of-the-art face recognition algorithms can only handle simple expression or occlusion changes, lacking the generalization to new complex situations. Recently, we further pointed out “automatic face recognition is a complex pattern-recognition problem involved with early processing, perceptual coding, and cue-fusion mechanisms. Although countless solid contributions have been made, 100% accuracy in automatic face recognition in real-world settings remains an ambitious goal [2]”.

This paper focuses on the one sample per person problem (or, one sample problem for short) in face recognition, which is defined as follows. Given a store database of faces with only one image per person, the goal is to identify a person from the database later in time in any different and unpredictable expression, lighting, and aging, etc. from the individual image [4]. Due to its technical challenge and wide-range of applications, one sample problem has rapidly emerged as an active research field in the face recognition community. Many methods have been proposed to attempt to address this problem, such as several extensions of the principal component analysis (PCA), the synthesization of the virtual samples, and the high-dimensional local feature based methods. A comprehensive categorization and comparison of these method can be found in [4]. Many methods have reported promising results on the one sample problem. However, a recent study pointed out that many previously

* Corresponding author. Tel.: +86 10 62283059; fax: +86 10 62285019.

E-mail addresses: whdeng@bupt.edu.cn, whdeng@it.usyd.edu.au (W. Deng).

proposed methods cannot improve the simplest Eigenfaces (PCA) method when the complex image variations are presented [5]. In the one sample situation, the robust and accurate recognition in the complex situations still requires further researches.

In this paper we propose a novel statistical feature extraction algorithm, called uniform pursuit (UP), to address the one sample problem in face recognition. The underlying idea of the proposed method is that most recognition errors are due to the confusions between faces that look similar, and thus a recognition-oriented face encoding should map the close class prototypes to be distant, i.e., uniforming the pairwise distances between different class prototypes. Specifically, in the uniform pursuit algorithm, PCA first projects the face feature vectors into a lower dimensional space, followed by the whitening transformation, which counteracts the fact that the “sum of pairwise distance maximization” principle underlying PCA ignores the local confusion between the similar faces. Lower dimensional space for enhanced recognition performance is then driven by a locality dispersing projection. The locality dispersing criterion in search of the optimal face basis by maximizing the pairwise distance between the faces that are easily confused (see Fig. 4). By encoding the faces in a manner that the pairwise distances between face prototypes tend to be uniform, the UP algorithm aims to simultaneously improve the robustness, accuracy and efficiency of the face recognition algorithm.

The feasibility of the UP algorithm is successfully tested on the large-scale experiments on the FERET database, which involves 1196 persons with challenging image variations across expressions, lighting, aging and long-term aging. The results show that, based on single training sample per person, the UP based face feature code leads to a robust, accurate and efficient recognition of the face image in new situations. In particular, for the different types of probe images, the UP algorithm outperforms the Eigenfaces method by large margins ranging from 9.8% to 36.2%, using the same input feature vector. The accuracy of the UP algorithm reaches an outstanding level when combined with the Gabor wavelet or LBP based face feature vector: its recognition rate reaches to 96.7%, 99.0%, 88.9%, and 87.6% for the fb, fc, duplicate I and duplicate II probe images, respectively. Moreover, compared with the state-of-the-art method [6] with closest accuracy level, the UP algorithm uses three order of magnitude lower feature dimension, largely reducing the memory and computation requirements of the recognition stage.

The remainder of this paper is organized as follows. Section 2 presents the background and the motivations of the new method. Section 3 details the rationale and algorithmic procedure of the UP method. Section 4 suggests a standard procedure to evaluate the one sample problem. Section 5 assesses the feasibility of the UP algorithm using the new evaluation procedure. Section 6 draws the conclusions.

2. Background and motivations

PCA seeks a low-dimensional representation of the data to retain as much as the variance in the data as possible. Given a set of data points $\{x_i\}_{i=1}^N$, where $x_i \in \mathbb{R}^M$ is an M -dimensional column vector, we expect to get their low-dimensional representations $\{y_i\}_{i=1}^N$ by projecting each x_i onto the direction vector $w \in \mathbb{R}^M$. The objective function of PCA is defined as follows:

$$\max_{\|w\|=1} \sum_{i=1}^N \sum_{j=1}^N (y_i - \bar{y})(y_j - \bar{y}) \quad (1)$$

where $y_i = w^T x_i$, and \bar{y} is the centroid of $\{y_i\}_{i=1}^N$. The objection function therefore can be reformulated as a more familiar form as

follows:

$$\max_{\|w\|=1} w^T \Sigma w \quad (2)$$

where $\Sigma = (1/N) \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T$ is the sample covariance matrix. The eigenvectors of Σ corresponding to the largest $d < M$ eigenvalues span the optimal subspace of the PCA. In face recognition, x_i represents a face image, and the eigenvectors are the so-called Eigenfaces [7].

This section first reviews the PCA extensions which attempt to address the one sample problem, and then point out a fundamental limitation of the PCA technique on the one sample problem, which has not been addressed by previous methods.

2.1. Previous PCA extensions

Based on the standard Eigenfaces technique, researchers have developed various extended algorithms during the last decades, such as probabilistic-based Eigenfaces [8], linear discriminative analysis (LDA) based subspace algorithms [9], evolution pursuit [10], and Laplacianfaces [11], etc. All of these approaches claimed to be superior to Eigenfaces. However, on the one sample problem, most of them will either reduce to the basic Eigenfaces approach or simply fail to work in that case [4]. Therefore, the Eigenfaces method becomes a standard technique for the one sample problem. Several PCA extension algorithms have been proposed to address the one sample problem.

In Ref. [12], an extension of the standard PCA was proposed, denoted as projection-combined PCA ($(PC)^2A$). The proposal introduced a new pre-processing scheme by combining the original image with its first-order projection map followed by a standard PCA. In order to enhance performance, following the introduction of $(PC)^2A$ framework, Chen et al. [13] proposed an Enhanced $(PC)^2A$ solution by including a second-order projection map while Zhang et al. introduced a SVD perturbation in the pre-processing step [14]. All these pre-processing based methods have reported slightly better performance (1–2%) compared to that of the standard PCA on the data set with expression variation. Wang et al. [15] elaborately selected the PCA basis for feature extraction according to its discriminant power on the generic database with multiple samples per person. However, the feasibility of these methods on complex image variations is still under question. A recent study even showed that the Enhanced $(PC)^2A$ and SVD perturbation procedures deteriorate the performance of standard PCA method when the complex image variations are presented [5].

Two-dimensional PCA (2DPCA) is a significant extension of PCA [16]. It uses straightforward 2D image matrices rather than 1D vectors for covariance matrix estimation, thus claimed to be more computationally cheap and more suitable for small sample size problem. However, it has been proven that 2DPCA essentially perform PCA by using the image row (or column) as elements, and it would lose the covariance information between the pixel within an image row (or column) [17]. The extension algorithms, such as DiaPCA [18], alleviate this problem by considering the covariance information between the diagonal pixels or combining the information from both row-based and column-based 2DPCA. There inevitably is covariance information loss in the 2D image matrix based PCA methods.

In addition, the image matrix based PCA methods can only suitably applied to the image pixel feature. They may not be suitable to other texture descriptors such as the Gabor feature and LBP, which have been shown to more discriminative for face recognition. In this respect, the standard PCA technique is more flexible than the 2D image-based PCA techniques.

2.2. Fundamental PCA limitation: local confusion

Given the objective function (2), an in-depth understanding on PCA can be derived by reformulating the covariance matrix as follows:

$$\begin{aligned}
 \Sigma &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T \\
 &= \frac{1}{N} \sum_{i=1}^N \left(x_i - \frac{x_1 + x_2 + \dots + x_N}{N} \right) \left(x_i - \frac{x_1 + x_2 + \dots + x_N}{N} \right)^T \\
 &= \frac{1}{N} \frac{1}{N^2} \sum_{i=1}^N \left[\sum_{j=1}^N (x_i - x_j) \sum_{k=1}^N (x_i - x_k)^T \right] \\
 &= \frac{1}{N} \frac{1}{N^2} \sum_{i=1}^N \left[\sum_{j=1}^N \sum_{k=1}^N (x_i - x_j)(x_i - x_k)^T \right] \\
 &= \frac{1}{N} \frac{1}{N^2} \frac{1}{2} \left\{ \sum_{i=1}^N \left[\sum_{j=1}^N \sum_{k=1}^N (x_i - x_j)(x_i - x_k)^T \right] \right. \\
 &\quad \left. + \sum_{i=1}^N \left[\sum_{j=1}^N \sum_{k=1}^N (x_j - x_i)(x_j - x_k)^T \right] \right\} \\
 &= \frac{1}{N} \frac{1}{N^2} \frac{1}{2} \sum_{i=1}^N \left[\sum_{j=1}^N \sum_{k=1}^N (x_i x_i^T - x_i x_k^T - x_j x_i^T + x_j x_k^T) + (x_j x_j^T - x_j x_k^T - x_i x_j^T + x_i x_k^T)^T \right] \\
 &= \frac{1}{2N^2} \sum_{i=1}^N \sum_{j=1}^N (x_i - x_j)(x_i - x_j)^T \tag{3}
 \end{aligned}$$

From the above decomposition, the covariance matrix in essence describes the pairwise difference between any two samples in the data set. Hence, the objective function of PCA is to measure the sum of pairwise distance between any two projected samples as follows:

$$\begin{aligned}
 w^T \Sigma w &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N w^T (x_i - x_j)(x_i - x_j)^T w \\
 &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \|w^T x_i - w^T x_j\|^2 \tag{4}
 \end{aligned}$$

In face recognition, the Eigenfaces space characterizes the image (feature vector) difference between any two face images. We can analyze the Eigenface space in two separated cases. In the first case, there are multiple images per person in the training data set. The image pairs may come from the same person, or from different persons. Obviously, if the image pairs are from the same person, maximizing their distance in the subspace would deteriorate the recognition performance. Therefore, one tend to apply the supervised methods that distinguish the intra-class and inter-class pairwise distance. In the second case, there is a single training sample per person. The pairwise differences characterized by the covariance matrix all come from different persons. The PCA subspace

can roughly maximize the pairwise distance between all classes. From this respect, of global feature space, the PCA subspace is reasonable to conduct recognition in the one sample situation.

We would like to point out that PCA has a fundamental limitation even on the one sample problem. The intuition is that there are many persons who look very similar to each other, as shown in Fig. 1. The similar persons may be corresponding to the close samples in the feature space. On one hand, it is the close samples that reduce the recognition accuracy on the one sample problem. On the other hand, since the PCA objective function focuses on the sample pairs with large distance, the reduced PCA subspace tends to merge the close samples together. Fig. 2 illustrates an example where PCA based dimensionality reduction merges two similar persons together. In this respect, the PCA based feature space is not optimal for one sample problem.

Although several PCA extensions [12–16,19,18] have been developed to improve face recognition performance on the one sample problem, these methods cannot address the “local confusion” limitation of PCA. It is this fundamental PCA limitation that motivates us to propose a novel feature extraction method called uniform pursuit for the one sample problem.

3. Uniform pursuit

This section first introduces two methods to address the “local confusion” problem of PCA, namely the whitened PCA and locality dispersing projection, and then integrates them to a novel uniform pursuit algorithm.

3.1. Whitened PCA

A zero-mean random vector is said to be white if its elements are uncorrelated and have unit variance. Hence, the whitening transformation is usually performed by first decorrelating the data using PCA and then scaling each principal direction to uniform the spread of the data. It is well known that the principal component directions (vectors) are exactly the eigenvectors of the data covariance matrix

$$\Sigma = U \Lambda U^T \tag{5}$$

The random vector x is decorrelated as follows:

$$\hat{x} = U^T (x - \bar{x}) \tag{6}$$

The whitening process weights each

$$\bar{x} = \Lambda^{-1/2} \hat{x} \tag{7}$$

in decreasing order of corresponding eigenvalues. The physical meaning of the eigenvalue is the data variance along corresponding eigenvector.



Fig. 1. Similarity of frontal faces between (a) two former US presidents (downloaded from BBC news, news.bbc.co.uk); (b) two Chinese superstars.

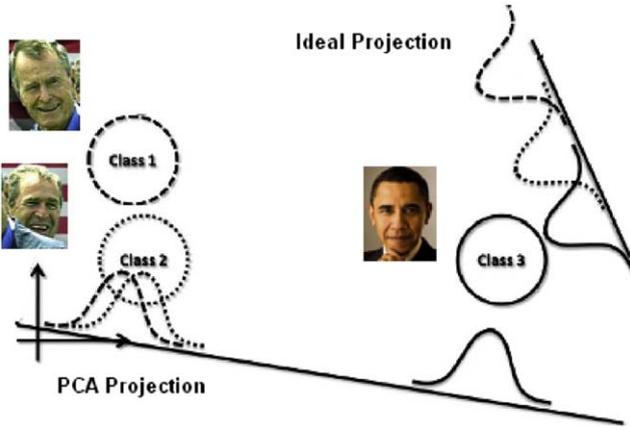


Fig. 2. The geometric interpretation of the fundamental limitation of PCA: the PCA subspace tends to collapse the similar classes with small pairwise distance.

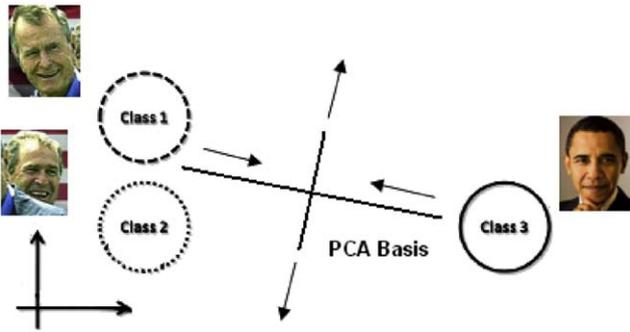


Fig. 3. The geometric intuition of whitening process. The figure shows the whitening process tends to weights heavily on the direction that characterizes the close prototypes, and thus uniform the pairwise distance of all prototypes.

In the situation where there are multiple samples per person, it is well known that the whitened PCA can address the shortcomings of PCA: the leading eigenvectors encode mostly illumination and expression, rather than discriminating information. Several studies has been empirically demonstrated for superior performance on some large-scale experiments [20–22,1], such as those on the FERET database and the FRGC database. Recently, Liu related the whitened similarity measure with the Bayes decision rule by a specific distributional assumption [23].

In the situation where there is only single sample per person, our previous work has demonstrated that the PCA method followed by the whitening process, i.e., whitened PCA, can achieve high recognition accuracy, especially when it is combined with informative low level features, such as the Gabor based feature, and the cosine similarity measure [24]. We called this highly accurate face recognition scheme as “Gabor–Eigen–Whiten–Cosine” (GEWC). Recently, Nguyen et al. extended the GEWC scheme by using a more complex low-level representation [25], and reported more accurate recognition results. Fig. 3 illustrates the rationale of the whitening transformation on the one sample problem. By normalizing the data spread, the whitening process tend to map the close classes, which distribute along the direction of the trivial component, to be far apart.

Although excellent performance has reported on the whitened PCA method with cosine similarity measure, there are still two weaknesses. First, in order to preserve the distance between the close classes, the whitened PCA usually needs to retain the small principal components to achieve good performance [24], resulting in a high-dimensional feature space. Second, since whitening process only uniforms the inter-sample pairwise distance in a very

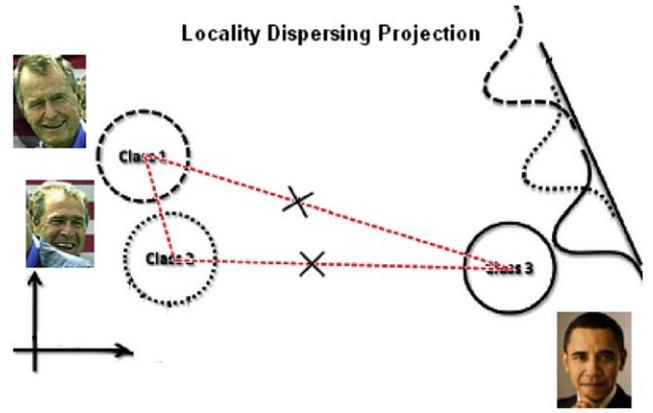


Fig. 4. An interpretation of the locality dispersing projection. By eliminating the connection between the distant prototypes, LDP tends to find the projection that discriminates the close prototypes.

rough way, there may still be some local confusions in the whitened PCA space. To simultaneously address these two problems, we append a novel dimension reduction technique to the whitened PCA space, which will be detailed in the following section.

3.2. Locality dispersing projection

The faces of some people are especially similar to each other, which leads to class confusion and miss recognition. If there is single training image per person, the close samples would form local concentrations in the feature space. Obviously, these local clusters or concentrations are deleterious for the recognition purpose. In the PCA based dimension reduction framework, in order to make the objective function focus on the local structure of the data, a straightforward method is to define a “local” version of covariance matrix as follows:

$$\Sigma_L = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N A_{ij} (x_i - x_j)(x_i - x_j)^T \quad (8)$$

where A_{ij} defines whether sample x_i and sample x_j are of a local concentration structure. The local proximity relationship can be defined by the absolute distance between two samples,

$$A_{ij} = \begin{cases} e^{-\|x_i - x_j\|^2 / t} & \text{if } \|x_i - x_j\|^2 < \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

This neighborhood has clear physical meaning from thermodynamics, however, the kernel size t and distance threshold ε are difficult to determine in practice, especially when the samples are very sparse in the feature space. Given the application on the one sample problem, we applied a more practical definition of neighborhood as follows:

$$A_{ij} = \begin{cases} 1 & \text{if } x_i \text{ is among the } k \text{ nearest neighbors of } x_j, \text{ or} \\ & x_j \text{ is among the } k \text{ nearest neighbors of } x_i \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

The neighborhood is determined by the k nearest-neighbor graph of the data set.

Analogy to the PCA technique, the objective function of the LDP is naturally defined as follows:

$$\max_{\|w\|=1} w^T \Sigma_L w \quad (11)$$

Note that this objective function has clear physical meaning as follows:

$$w^T \Sigma_L w = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N A_{ij} w^T (x_i - x_j)(x_i - x_j)^T w$$

$$= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N A_{ij} \|w^T x_i - w^T x_j\|^2 \quad (12)$$

where $w^T x_i$ and $w^T x_j$ are the projected samples on the direction w . Therefore, the objective function (11) aims to maximize the sum of the pairwise distance between the samples involved in the local concentration structures. Obviously, the eigenvectors of Σ_L corresponding to the largest $d < M$ eigenvalues span the optimal subspace of the LDP, where the local concentration of the original space can be dispersed (see Fig. 4).

According to the recently proposed graph embedding framework [26], the local covariance matrix can be reformulated in a graph form as follows:

$$\begin{aligned} \Sigma_L &= \frac{1}{2N^2} \left(\sum_{i=1}^N \sum_{j=1}^N A_{ij} x_i x_i^T + \sum_{i=1}^N \sum_{j=1}^N A_{ij} x_j x_j^T - 2 \sum_{i=1}^N \sum_{j=1}^N A_{ij} x_i x_j^T \right) \\ &= \frac{1}{2N^2} \left(2 \sum_{i=1}^N D_{ii} x_i x_i^T - 2 \sum_{i=1}^N \sum_{j=1}^N A_{ij} x_i x_j^T \right) \\ &= \frac{1}{N^2} (XD X^T - X A X^T) = \frac{1}{N^2} X L X^T \end{aligned} \quad (13)$$

where $X = [x_1, x_2, \dots, x_N]$ is the data matrix, the matrices A , D and L can be regarded as the weight, degree, and Laplacian matrices of a weighted graph, respectively. Note that popular graph-based feature extraction methods, such as LPP [11,27] and UDP [28,29], aim to *preserve* the local concentration of the samples, with the underlying assumption that there are multiple samples per class and the close samples tend to be of the same class. Compared to previous methods, the proposed LDP provides a novel application of the graph-based data structure to *desperate* the local concentration and address the one sample problem.

3.3. The uniform pursuit algorithm

The methodology for face feature encoding is shown in Fig. 5. The main goal is to find out an optimal basis along which faces can be projected leading to a compact and efficient face encoding in terms of recognition ability. Towards this end, PCA first projects the face feature vectors into a lower dimensional space. The next step is the whitening transformation and it counteracts the fact that the sum of pairwise distance maximization principle underlying PCA ignores the confusion between the similar faces. Lower dimensional space for enhanced recognition performance is then driven by the locality dispersing criterion. The locality dispersing criterion in search of the optimal face basis considers the maximization of the pairwise distance between the faces that are easily confused.

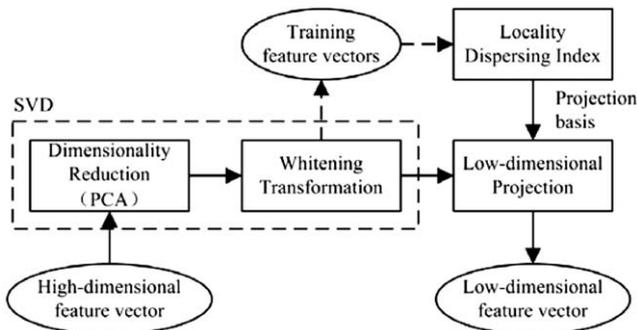


Fig. 5. The methodology of the UP algorithm for face feature encoding.

Specifically, the uniform pursuit algorithm integrates the whitened PCA and the new locality dispersing projection for feature extraction purpose. The detail algorithmic procedures are as follows:

1. *Data centering and SVD based whitened PCA*: We remove the weighted mean of x from each x

$$\hat{x} \leftarrow x - \bar{x}, \quad \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (14)$$

The centered data matrix $X = [\hat{x}_1, \dots, \hat{x}_N]$ is decomposed by the singular value decomposition as follows:

$$X = U D V^T \quad (15)$$

where U and V are the matrices of the left and right singular vectors and D is the diagonal matrix of singular values. Let \tilde{D} be the diagonal matrix of the p largest singular values, and \tilde{U} be the matrix of corresponding left singular vectors, the transformation matrix of whitened PCA is defined as $W_{wpc a} = \tilde{U} \tilde{D}^{-1}$. By the SVD based whitening projection, the centered feature vector x is transformed to

$$\tilde{x} = W_{wpc a}^T x \quad (16)$$

After this step, the data matrix X becomes the whitened data matrix $\tilde{X} = [\tilde{x}_1, \dots, \tilde{x}_N] \in \mathbb{R}^{p \times N}$

2. *Constructing the adjacency graph*: Let G denote a graph with N nodes. The i th node corresponds to the face \tilde{x}_i . We put an edge between nodes i and j if \tilde{x}_i and \tilde{x}_j are “close,” i.e., \tilde{x}_i is among k nearest neighbors of \tilde{x}_j or \tilde{x}_j is among k nearest neighbors of \tilde{x}_i . If nodes i and j are connected, put $A_{ij} = 1$; otherwise, put $A_{ij} = 0$. The weight matrix A of graph G models the local structure of the whitened face space.
3. *Locality dispersing projection*: Compute the eigenvectors and eigenvalues for the eigenproblem:

$$\tilde{X} L \tilde{X}^T u = \lambda u \quad (17)$$

where L is the Laplacian matrix of the adjacency graph. The Laplacian matrix of the graph can be computed as $L = D - A$, where D is a diagonal matrix with $D_{ii} = \sum_j A_{ij}$. Let u_1, u_2, \dots, u_p be the solutions of (17), ordered according to their eigenvalues, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. The optimal projection matrix of the LDP is constructed as $W_{ldp} = [u_1, \dots, u_d]$, where $d < p$.¹

Combining the WPCA and LDP stages, the uniform pursuit method encodes a face feature vector x as follows:

$$x \rightarrow y = W^T (x - \bar{x}) \quad (18)$$

$$W = W_{wpc a} W_{ldp} \quad (19)$$

and

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (20)$$

Let $y_k, k = 1, 2, \dots, L$, be the prototype for person ω_k after the UP projection. For a novel feature vector y that represents a probe image, the classifier applies the nearest neighbor rule for recognition using some similarity (distance) measure δ

$$\delta(y, y_k) = \min_j \delta(y, y_j) \rightarrow y \in \omega_k \quad (21)$$

The similarity (distance) measures used in our experiments to evaluate the efficiency of different representation methods include L_1 distance measure, δ_{L_1} , L_2 distance measure, δ_{L_2} , cosine

¹ Dimension reduction is necessary in the LDP stage. Since u_1, u_2, \dots, u_p is a set of orthogonal basis, if $d = p$, the projection W_{ldp} would perform a rotation of the whitened PCA space, which does not change the recognition performance at all.

similarity measure δ_{\cos} , which are defined as follows:

$$\delta_{L_1}(\mathcal{U}, \mathcal{V}) = \sum_i |\mathcal{U}_i - \mathcal{V}_i| \quad (22)$$

$$\delta_{L_2}(\mathcal{U}, \mathcal{V}) = (\mathcal{U} - \mathcal{V})^t (\mathcal{U} - \mathcal{V}) \quad (23)$$

$$\delta_{\cos}(\mathcal{U}, \mathcal{V}) = -\frac{\mathcal{U}^t \mathcal{V}}{\|\mathcal{U}\| \|\mathcal{V}\|} \quad (24)$$

The novelty of the UP algorithms comes from its sequential usage of “global” and “local” covariance matrix to reduce the confusion of similar faces. The whitened PCA stage globally spheres the feature space by weighting the principal components, and the LDP stage seeks the locality dispersing projections that separate the locally confused faces. One may wonder, as LDP stage relies on a “local” version of covariance matrix, could it make two different faces look similar? For example, P1 is a face feature vector in a similar face cluster C1, and P2 is a face feature vector in another similar face cluster C2. The distance between C1 and C2 is large. After applying the LDP, P1 is distinguished from cluster C1 now, and P2 is distinguished from cluster C2 now. But is it possible the distance between P1 and P2 is small²?

Indeed, there is a trade-off between the requirement of reducing local confusion and the risk of mapping distant faces to be too close. Fortunately, this problem can be circumvented by suitably setting the local neighborhood size of LDP, i.e., the parameter k . Note that the LDP is applied in the whitened PCA space, where the data projected on any direction would have unit variance. This means that, when the projected variance (sum of pairwise distance) among close faces increase, the variance among the distant faces would decrease by an equal quantity. By controlling the neighborhood size of LDP, i.e., the parameter k , not to be too large, the projected results would increase the variance of close faces by only a small quantity. Hence, the distant faces would not be mapped to be close together. Our experimental results show that an appropriate setting of k could balance the requirement of reducing local confusion and the risk of mapping distant faces to be too close, and thus achieve improved recognition accuracy in the whitened PCA space. For instance, we set $k = 20$ in a training set of 1196 faces in the experiment.

4. On performance evaluation

4.1. Previous performance evaluation

As pointed out by Tan et al., although many new algorithms have been developed for the one sample problem, there is still not any standard testing procedure for the one sample problem [4]. Most of the algorithms have been tested on different data sets with different training/testing partitions. Most of them are tested on the probe images that contain only simple facial variations, such as different expression, or man-made occlusions. For the real world applications, however, the largest challenge comes from the unpredictable image variations, such as those caused by the uncontrolled lighting conditions and aging effects.

To address this problem, Wang et al. [5] collected a large-scale data set by combing several face databases, which contains the complex lighting and aging effects. An important contribution of Wang et al.’s work is the demonstration that the experimental results on the simple situation is not sufficient to predict the algorithms’ applicability on the more complex situations. For example, the $E(PC)^2A$ [13], SVD permutation [14], and SOM [30]

Table 1

The empirical comparison results between the PCA baseline and improved methods, namely $E(PC)^2A$, SVD permutation, and SOM, are different with the simple image variation and complex image variations.

Image variations	PCA	$E(PC)^2A$ [13]	SVD [14]	SOM [30]
Expression	83.50	85.50	85.00	89.50
Expression+Lighting+Aging [5]	66.00	59.59	64.13	58.75

methods have been validated to improve the PCA based face recognition performance, based on the FERET subset with simple expression variation. However, Wang’s large-scale experiment with complex Expression+Lighting+Aging variation shows that these methods perform worse than basic PCA (see Table 1). Obviously, Comparing to the previous evaluation procedures [13,14,30], Wang’s experimental settings are arguably more closely related to the practical usages, and therefore one can infer that these methods may not be applicable to the real-world environment. Therefore, to guarantee a new method really work in practice, one should evaluate the effectiveness of the new method on a sufficiently large and complex image set, rather than a small and simple one.

Although Wang’s evaluation procedure is arguably more scientific than the previous ones, it still has some limitations. First, there are not settled training, gallery and probe image lists, so that it is hard to fairly compare different methods. The researchers must re-implement the old methods which they want to compare the new method to. In these cases, the implementation details and the parameter selections may bias the comparative result of different algorithms. Second, there are not partitions of the probe image, so that one cannot investigate the algorithms’ strength and weakness on different kinds of image variations. Third, the gallery size, i.e., 80 persons, is relatively small, so that the algorithm performs well on such experiment may not necessarily be applicable to large-scale situations.

4.2. A proposal on standard evaluation procedure

To address the limitations of previous evaluation methods, we propose a new standard testing procedure for evaluating the one-sample face recognition, which considers both the traditional specific learning [4] and the generic learning [5] methodology. Specifically, we suggest to use the standard data partitions in the FRGC and FERET databases, which are distributed by NIST and already has been made worldwide available in the face recognition community.

As pointed out by Tan et al. [4], the standard FERET protocol requires the algorithm’s training is completely prior to the start of the test, but lacks a settle “generic learning” database. Hence, different participant collected different training data set, which makes different algorithms hard to fairly compare. Fortunately, the recently released FRGC version 2 database [21,22] provides a high-quality training database, which is the largest-size training data set so far in the literature. We suggest that one can train an algorithm in such a large training database, and test it on the FERET database.

The proposal standard evaluation procedure utilizes following settled data partitions:

- *Generic training set* (12,766 images), contains frontal images of 222 people, the images are taken in both controlled and uncontrolled conditions.
- *Specific training set* (1196 images), contains frontal images of 1196 people.

² We would like to thank the anonymous reviewer who pointed out this potential weakness of the UP algorithm.



Fig. 6. Example Images in the FERET database. (a) The original images offered by the FERET database; (b) the cropped images used in our experiments.

- *Gallery set* (1196 images), contains frontal images of 1196 people.
- *fb Probe* probe set (1195 images). The subjects were asked for an alternative facial expression than in the gallery photograph.
- *fc Probe* probe set (194 images). The photos were taken under different lighting conditions.
- *duplicate I Probe* set (722 images). The photos were taken later in time.
- *duplicate II Probe* set (234 images). This is a subset of the dup I set containing those images that were taken at least a year after the corresponding gallery image.

The standard image list files, as well as the eye coordinates³ for image normalization, of the above data sets are distributed accompanying with the FRGC and FERET databases by NIST, so that one can perform the standard evaluation procedure once obtaining the databases.

5. Experiments

This section evaluates the effectiveness of the UP algorithm on one sample face recognition problem using the large-scale FERET database. Specifically, we take 1196 images from the FERET gallery (the “fa” set) and use them both as a training set and as a gallery in the recognition phase. Then they used other FERET sets (fb, fc, duplicate I, and duplicate II) to test the algorithm’s accuracy. In addition, we also consider some practical issues such as the recognition performance using generic learning when the gallery images cannot be used for training.

The image is first normalized by an affine transformation that sets the centered inter-eye line horizontal and 70 pixel apart, and then cropped to the size of 128×128 with the centers of the eyes located at (29, 34) and (99, 34) to extract the pure face region. No further preprocessing procedure is carried out in our experiments. Fig. 6(b) shows some cropped images which are used in our experiments. One can see from the figure, there are large image variations between the gallery images and the probe images with lighting and aging effects. A characteristic of our experimental study is the usage of various face representations. Besides the

commonly used appearance (pixel) representation, our study also includes the Gabor wavelet representation [1], and the LBP face descriptor [31].

The appearance-based method represents the face image by the pixel value of the gray scale image. The pixels in the image are simply vectorized into a 4096-dimensional vector by reading pixel values within the 64×64 image in a raster-scan manner, where the 64×64 image is generated by resizing the 128×128 image using the bilinear interpolation method.

The Gabor wavelet representation is the convolution (using FFT) of the image with a family of Gabor kernels with five scales and eight orientations. The Gabor filter responses are down-sampled by a 8×8 uniform lattices. We construct a 256 dimensional vector out of each down-sampled responses, normalize it to zero mean and unit variance, and concatenate them into a single 10,240 dimensional feature vector. For the detailed parameters, refer to [1].

The LBP feature vector, in a dimensionality of 15,104, is extracted by the histograms of the 59 uniform binary patterns in each 8×8 pixel cells in the 128×128 image, where the patterns are generated by thresholding 8 pixels in a circle of radius 2. For the detailed LBP parameters, refer to [32].

5.1. PCA vs. whitened PCA

For the PCA and WPCA based feature, we apply L_1 distance measure, δ_{L_1} , L_2 distance measure, δ_{L_2} , cosine similarity measure, δ_{\cos} , in the recognition stage. Note that the widely used Mahalanobis distance and whitened cosine similarity are equivalent to the L_2 distance measure and cosine similarity measure in the whitened PCA space, respectively. Note that similar performance study was performed in [20], but with multiple samples per person.

Since a gallery of 1196 persons is used for training, there are at most 1195 principal component vectors are available for feature extraction. For a fair comparison, we settle the dimensionality of PCA and WPCA to 500, and evaluate their rank-1 recognition rate for the four types of probe images. Fig. 7 summarizes the recognition rate of different types of probe images at the PCA dimensionality of 500. Across all the three low level features, the WPCA encoding with cosine similarity measure, i.e., the whitened cosine similarity measure, performs better than other methods by a large margin.

³ We have found that there were slight errors on the eye coordinates of some FERET images, and remarked the eye coordinates of all the FERET images accurately. To obtain the updated eye-coordinate files, one could contact the author by whdeng@bupt.edu.cn.

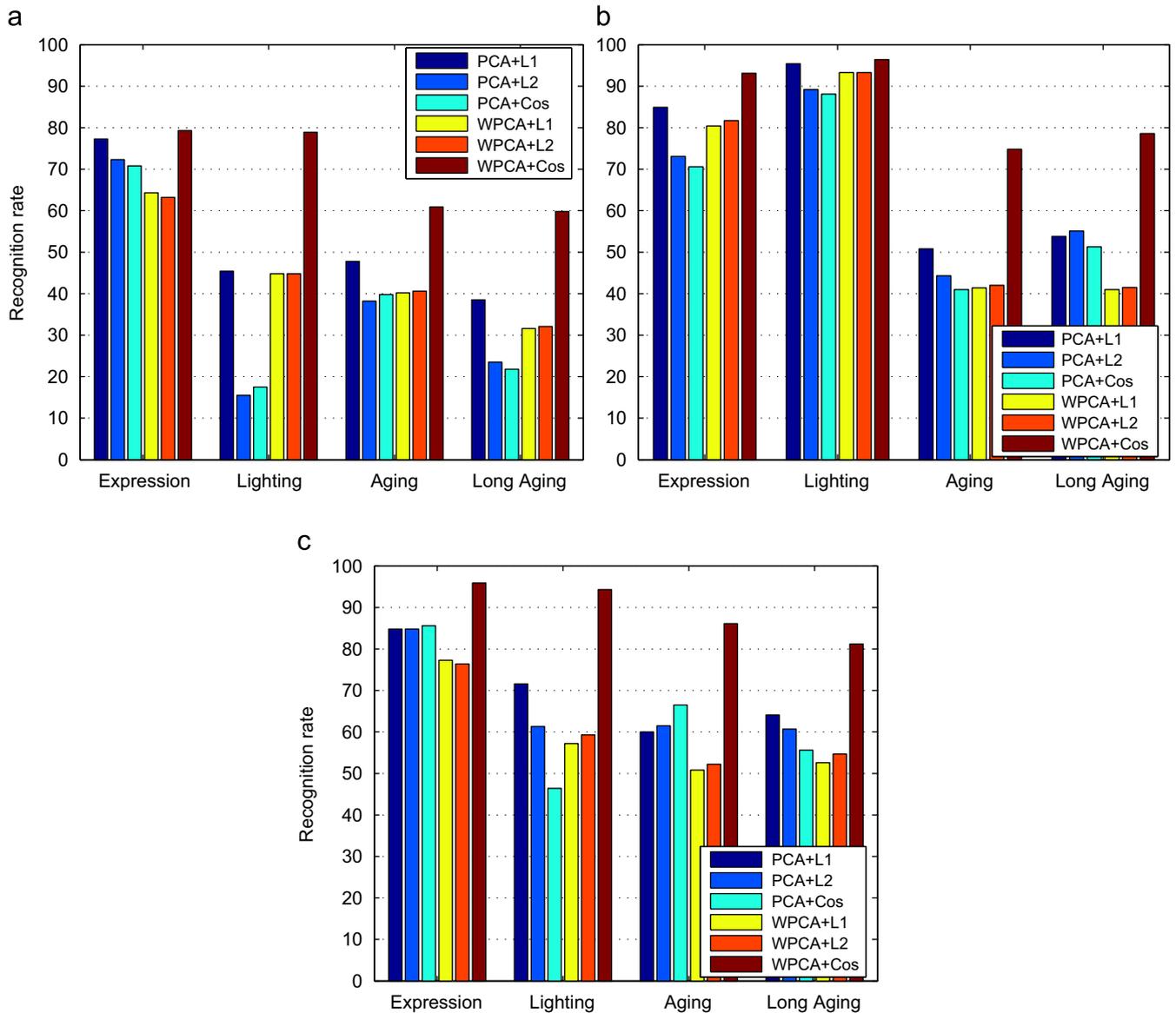


Fig. 7. Comparative FERET face recognition rate of PCA vs. WPCA on the probe image sets with different types of image variations: expression (fb), lighting (fc), aging (duplicate I) and long aging (duplicate II). For a fair comparison, both PCA and WPCA are based on 500-dimensional space and the accuracies based on different distance metric are evaluated. (a) Pixel. (b) Gabor. (c) LBP.

Across all similarities, the Gabor wavelet and LBP based methods provides a more discriminative face representation than the pixel based method. While most one-sample method use the pixel value as the low level feature, our experimental results clearly suggest the informative low level features, such as the Gabor wavelet and LBP based texture descriptors, lead to much better accuracy on the one sample problem. Furthermore, our results reveal that the Gabor wavelet is better at tackling the illumination changes (fc probes), while the LBP is more proficient at counteracting the image variability caused by different acquisition times and locations (dup I and dup II probes).

5.2. Uniform pursuit performance

As indicated in the previous experiment, we proceed our experiment with the Gabor wavelet and LBP features, and reduced the feature dimension using the proposed UP algorithm. There are two parameters in the UP algorithm, i.e., p and k . To keep the

consistency with the previous experiment, p is set to 500. In the $p=500$ dimensional whitened PCA space, the UP algorithm determines the “closeness” of the sample by the cosine similarity measure, and we have tried the number of nearest neighbors as $k = \{10, 20, 30, 40, 50\}$, and report the result with $k=20$ that provides a slightly better result than others.

As in most previous studies, we use the Eigenfaces (PCA with L_2 distance measure) method as a standard baseline to test the robustness of the proposed methods. As summarized in [4], previous improved PCA methods have shown the 0.5% to 10% accuracy improvement over the Eigenfaces method. We therefore use the Eigenfaces as our baseline and Fig. 8 show the comparative performance using the Eigenfaces method and the proposed uniform pursuit method. One can see from the figure that, on different types of probe images, the UP algorithm dramatically improves the Eigenfaces method by large margins ranging from 9.8% to 36.2% (The numerical results given in Table 2.). Note that in our experiment both the Eigenfaces method and the UP method are based on the 500 dimensional subspace

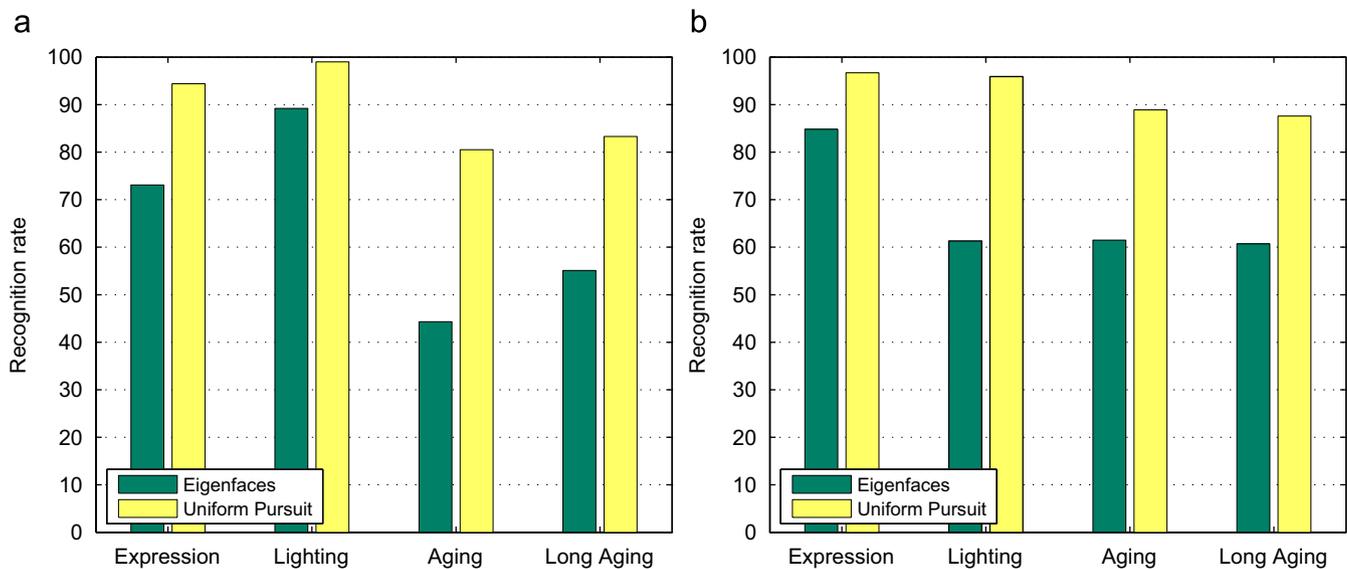


Fig. 8. Comparative FERET face recognition performance using the Eigenfaces method and uniform pursuit method. (a) Gabor. (b) LBP.

derived from PCA, so that the performance enhancement is purely because the UP algorithm addresses the fundamental PCA limitation on the “locally confusion”.

Certainly, as the UP algorithm has two stages, the dramatic accuracy improvement comes from both the WPCA and the LDP stage, and it is interesting to study how large improvement is obtained by each of the two stages. Fig. 7 has clearly shown the performance boost from PCA to WPCA. Fig. 9 illustrates the improvement achieved by the LDP stage only. As shown in Fig. 9, for the fb and fc probes, the UP algorithm achieves slightly better accuracy than the WPCA method. Given the fact that the WPCA accuracy on the fb and fc probes is already very high ($> 93\%$), the improvement by the LDP stage is considerable. The contribution of the LDP stage is more remarkable on the dup I and dup II probe images. This may be because the aging and long aging probes contain more severe variations that makes the probes deviate far from the corresponding gallery images. In this complex situation, the dispersion of the local concentration (in the training data) can significantly reduce the risk of class confusion. This result indicates the LDP stage enhance the *robustness* of the UP algorithm against the aging effect on the face image. As shown in Table 2, the UP algorithm use lower feature dimension (listed in subscript bracket) than the WPCA method, so the LDP stage can also reduce the computation and memory costs of the recognition algorithm. Note that the optimal dimensions of UP are selected for different probe images, which may be hard to operate in practice. For a more assessable result, we simply settle the UP dimension at 400 and measure the accuracy on each probe sets. As shown in Table 2, the UP method outperforms WPCA method using lower (400) dimensional features on all probe sets.

For the comparison purpose, we also implement the independent component analysis (ICA) based algorithms, which have been also used to find better projection basis than PCA-based methods. There are a number of ICA implementations, but pervious studies have shown the face performance difference between them is trivial [33]. We therefore choose the FastICA [34] because of its highest speed. Specifically, we implemented both architectures of ICA [35], known as ICA-I and ICA-II, and we find that the ICA-I method has similar performance with the Eigenfaces (PCA) method, and the ICA-II method has similar

performance with WPCA, a result consistent with Yang et al. [33]. We test FastICA using contrast functions $G_1(u) = u^3$, $G_2(u) = \tanh(u)$, and $G_3(u) = u \exp(-u^2/2)$, and report the result of $G_1(u)$ that is slightly better than others. For all probe types, ICA-II outperforms ICA-I by a large margin, we therefore include the performance of ICA-II for comparison. Note that algorithmic procedure of ICA-II is similar to that of UP, since both algorithms first whiten the data and pursuit low dimensional projection. For a fair comparison, both UP and ICA-II are implemented in 500 dimensional whitened PCA subspace, and Table 2 lists the recognition rate of ICA-II using 400 features and optimal number of features (obtained by exhaustive search with a interval of 10), respectively, and one can summary that the UP algorithm outperforms ICA on all probe sets. The results also suggest that the ICA-II cannot improve the performance of whitened PCA in most cases, even using the optimal projection dimension. It seems that the ICA projection in the whitened PCA space is trivial for the recognition. Yang et al. [33] has made a similar observation on ICA by using multiple images per person for ICA training, and they also concluded that ICA projection has trivial effects on face recognition. In contrast, the UP projection improves the whitened PCA on all probe image sets.

As far as computation cost is concerned, as shown in Fig. 10, the training time of the UP algorithm, which is similar to those of Eigenfaces and Whitened PCA, is about one order of magnitude less than that of ICA-II (FastICA). In summary, UP algorithm is better than ICA-II in term of both the recognition accuracy and the computational efficiency.

To further demonstrate the effectiveness of the UP-based feature transformation, we would like to compare the UP method with other state-of-the-art face recognition methods. Unfortunately, in the one sample situation, few methods have been tested on the large-scale FERET database. Table 3 summarizes the comparison between the UP method and two recent methods which apply very high dimensional texture feature to one sample problem. The results show that the UP method outperform these two methods on the fc, dup I and dup II probe sets, using much lower (about three orders of magnitude lower) feature dimension. This clearly suggests that UP based feature encoding can fulfill very efficient and accurate face recognition based on one sample per person.

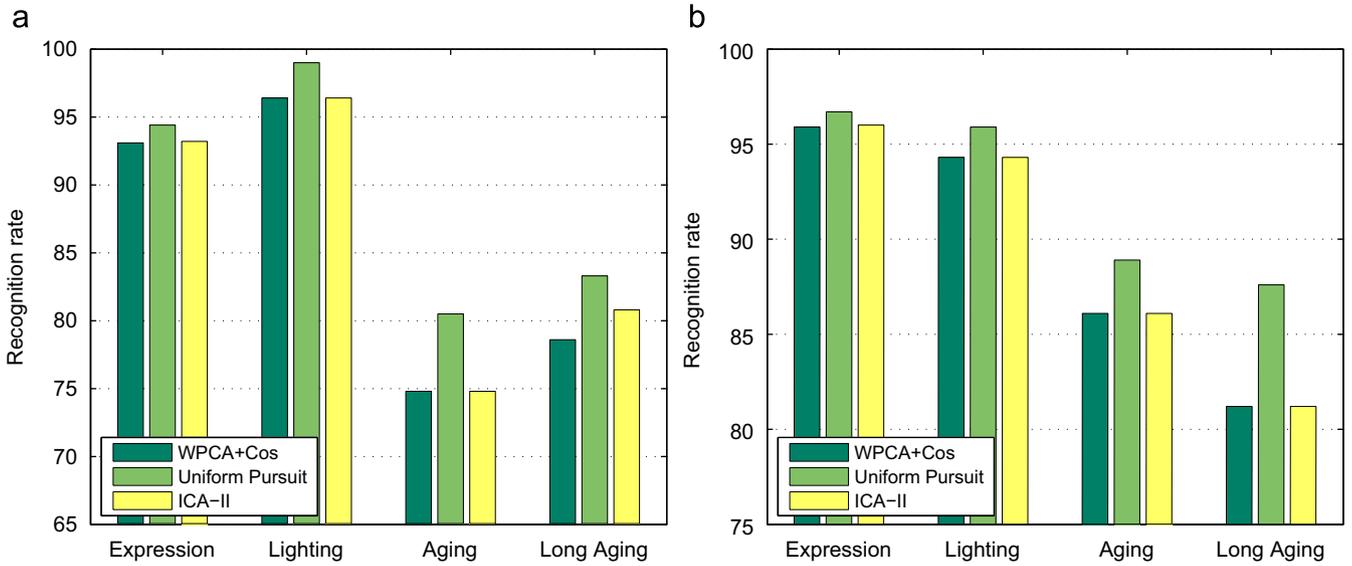


Fig. 9. Comparative FERET face recognition performance using the whitened PCA, uniform pursuit, and ICA (Architecture II) methods. (a) Gabor. (b) LBP.

Table 2

The recognition rates of different similarity or distance measures using Pixel Gray value, Gabor wavelet, and LBP based face representations, respectively.

Feature	Method	Dim	Sim	fb	fc	dupl	dup II
Gabor	Eigenfaces	500	L2	73.1	89.2	44.3	55.1
	WPCA	500	Cos	93.1	96.4	74.8	78.6
	UP _(opt)	Varying	Cos	94.4 ₍₄₁₀₎	99.0 ₍₂₉₀₎	80.5 ₍₄₂₀₎	83.3 ₍₄₁₀₎
	UP ₍₄₀₀₎	400	Cos	94.2	98.5	79.8	82.1
	ICA-II _(opt)	Varying	Cos	93.2 ₍₄₂₀₎	96.4 ₍₅₀₀₎	74.8 ₍₅₀₀₎	80.8 ₍₄₁₀₎
	ICA-II ₍₄₀₀₎	400	Cos	92.3	94.3	73.1	79.5
LBP	Eigenfaces	500	L2	84.8	61.3	61.5	60.7
	WPCA	500	Cos	95.9	94.3	86.1	81.2
	UP _(opt)	Varying	Cos	96.7 ₍₄₂₀₎	95.9 ₍₃₈₀₎	88.9 ₍₄₄₀₎	87.6 ₍₄₅₀₎
	UP ₍₄₀₀₎	400	Cos	96.5	94.8	88.5	86.3
	ICA-II _(opt)	Varying	Cos	96.0 ₍₄₄₀₎	94.3 ₍₅₀₀₎	86.1 ₍₅₀₀₎	81.2 ₍₅₀₀₎
	ICA-II ₍₄₀₀₎	400	Cos	94.9	89.7	81.9	74.4

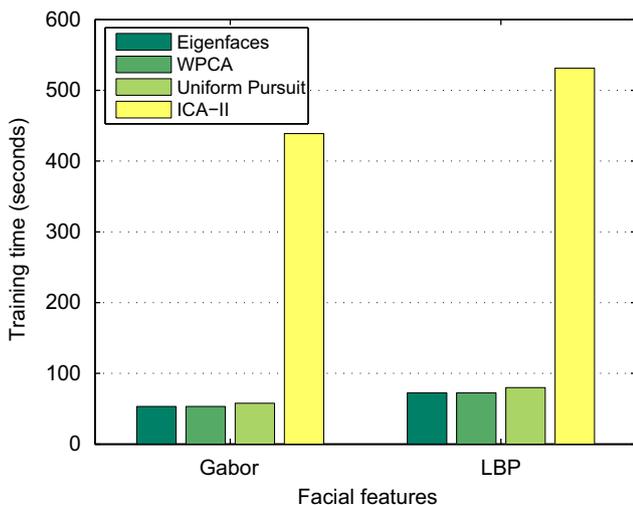


Fig. 10. Comparative training time using the Eigenfaces, Whitened PCA, uniform pursuit, and ICA (Architecture II) methods, on a training set of 1196 images.

Table 3

The performance comparison between the uniform pursuit method and other state-of-the-art face recognition methods.

Method	Dimension	fb	fc	dupl	dup II
FERET'97 Best [36]	N/A	96	82	59	52
LBP [31]	2891	97	79	66	64
HGPP [6]	737,280	97.6	98.5	77.7	76.1
Gabor+UP	< 500	94.4	99.0	80.5	83.3
LBP+UP	< 500	96.7	95.9	88.9	87.6

5.3. Experiment on a small FERET subset

Several one sample based algorithms have been tested on a FERET subset of 200 person.⁴ All these algorithms have reported improved accuracy compared to the Eigenfaces method. To compare the proposed UP algorithm to these previous methods,

⁴ The index file which lists the filenames of the images is download on <http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/annex/singleface-datafile.htm>.

Table 4

The performance comparison between the uniform pursuit method and the improved PCA based face recognition methods on the FERET subset with 200 persons.

Method	Feature	Dimension	Accuracy
PCA	Pixel	73	83.0 [12]
$(PC)^2A$ [12]	Pixel	60	83.5
$E(PC)^2A$ [13]	Pixel	60	85.5
SVD Perturbation [14]	Pixel	40	85.0
2DPCA [16]	Pixel	13×60	84.5 [19]
$(2D)^2PCA$ [19]	Pixel	13×60	85.0
DiaPCA [18]	Pixel	5×60	90.5
UP	Pixel	50	90.0
UP	Gabor	50	96.5
UP	LBP	50	97.0

we conduct an supplementary experiment on this FERET subset. For this small data set, the UP algorithm is implemented with $p=70$ and $k=5$. Table 4 summarizes the performance comparison between the uniform pursuit method and the improved PCA based face recognition methods. Note that this is only a rough comparison, because the experiments in the cited papers use different image cropping procedures. The comparative result in the table shows the UP algorithm outperforms most one-sample based methods, except the DiaPCA method, with similar feature dimension.

A vital advantage of the UP algorithm over the 2D image matrix based PCA methods is that the former can be applied to many other face descriptors besides the image pixel. Similar to the large-scale experiment, applying the texture features to facial image can largely improve the face recognition performance. Specifically, based on the same UP feature extraction, the Gabor and LBP based face representations raise the recognition accuracy from 90% to 96.5% and 97%, respectively. Since the 2D image matrix based PCA methods cannot be applied to the Gabor and LBP features, we can conclude that the proposed UP algorithm is much more accurate than other methods on this popular small-scale experiment.

5.4. Generic vs. specific learning: comparison and combination

Recently, Wang et al. suggested to use “generic learning” methodology to circumvent the one sample problem [5]. The basic idea is that the feature transformation that learned from a large generic database can be used to extract the discriminative features of the unseen faces. Many state-of-the-art face recognition solutions can be readily integrated in the generic learning framework. Their large-scale experiments demonstrated that the LDA-based generic learning methods, which was trained on the large generic database, can significantly outperformed the PCA method, as well as other one-sample methods, which was trained on the specific database [5].

Since we have shown that UP method can also largely outperform the PCA methods, as well as other one sample methods, it is interesting to compare the UP-based specific learning method and the LDA-based generic learning method. For the generic learning, we use the FRGC version 2 training database which contains 12,766 images of 222 persons. For the general learning with multiple images per person, we apply our previously proposed robust discriminant model (RDM) [37], a regularized LDA method which has been shown to be better than many other LDA algorithms. We first reduce the dimension to 1000 using PCA, and then apply RDM with the regularized factor $E=0.98$. Fig. 11 shows the comparative performance of the generic learning (RDM) and specific learning (UP) are very similar

in most cases. When the Gabor wavelet representation are used, the UP-based method performs much better than the RDM-based method on the dup I and dup II probe sets.

Given the fact that both specific learning (based on single image of specific person) and generic learning (based on multiple images of generic person) methods can achieve perform accurate recognition, it is a natural idea to combine both methods. For the simplicity, we use the sum-rule based similarity level fusion with z-score normalization [38]. Fig. 11 shows the combination of the generic and specific learning achieves slightly better accuracy on all probe image sets. In particular, based on the LBP face descriptor, the combination of the generic and specific learning achieves a recognition rate over 90% for duplicate I and duplicate II probes. While it is already known the combination of distinct low-level features, such as Gabor wavelet and LBP [32], can improve face recognition performance, this experiment newly find that the combination of the generic learning and specific learning can also improve performance.

5.5. Most challenging experiment: one-sample generic learning

Our previous evaluation of the UP algorithm uses the images of the same classes (persons) for both training and testing (recognition). This means that the UP method is fitted to those persons. Now, it can easily be envisioned that we are some day to develop a system (using your algorithm) to recognize a set of persons that are not known in advance and of which you do not have the images.⁵ Certainly, one can collect a training set with multiple images per person, and apply the LDA-like supervised feature extraction methods, such as RDM, to learn a feature space as in the pervious experiment. However, a more interesting question is: Can our proposed UP algorithm can be used for generic learning with single training image per person? Does the UP algorithm needs to be re-trained to adapt such “unseen” persons?

To evaluate whether the UP method can be applied to recognize such “unseen” persons, we design a more difficult experiment on the duplicate II probe set, which is regarded as the most difficult probe set [36]. The 234 images of duplicate II set contains 75 persons. By excluded these 75 persons in the training set, the UP algorithm have to learn the subspace without knowing the people to be recognized in advance. The effectiveness of the UP method on one-sample generic learning is evaluated in term of two aspects. One aspect is the comparative performance against other methods, such as Eigenfaces, WPCA, and ICA. The other aspect is how many accuracy drops with and without these 75 persons involved in the training set.

Fig. 12 show the comparative results, where “specific learning” means the training set contains all the 1196 images in the gallery, and “generic learning” means that the algorithms using the 1121 ($1196-75=1121$) images for training. The experimental results clearly show that the performance of UP is only slightly affected (by 1–2%), when the 75 person of the duplicate II probe set are excluded from the training set. Surprisingly, the performance of Eigenfaces (PCA) methods increase by 5–10% after excluding the 75 persons in the training set. This may because the Eigenfaces extracted from generic facial images represent the intrinsic facial difference well and simultaneously ignore (suppress) some within-class facial variations. Most importantly, the proposed UP algorithm still outperforms other methods in the generic

⁵ We would like to thank the anonymous reviewer who pointed out this interesting situation. Actually, we have already encountered this problem when demonstrating the mobile intelligent robot (introduced in next section) to the guests. The new guest who registered into the gallery database on the spot, and robot recognized the guest using the feature space learned from other persons, without re-training the whole system.

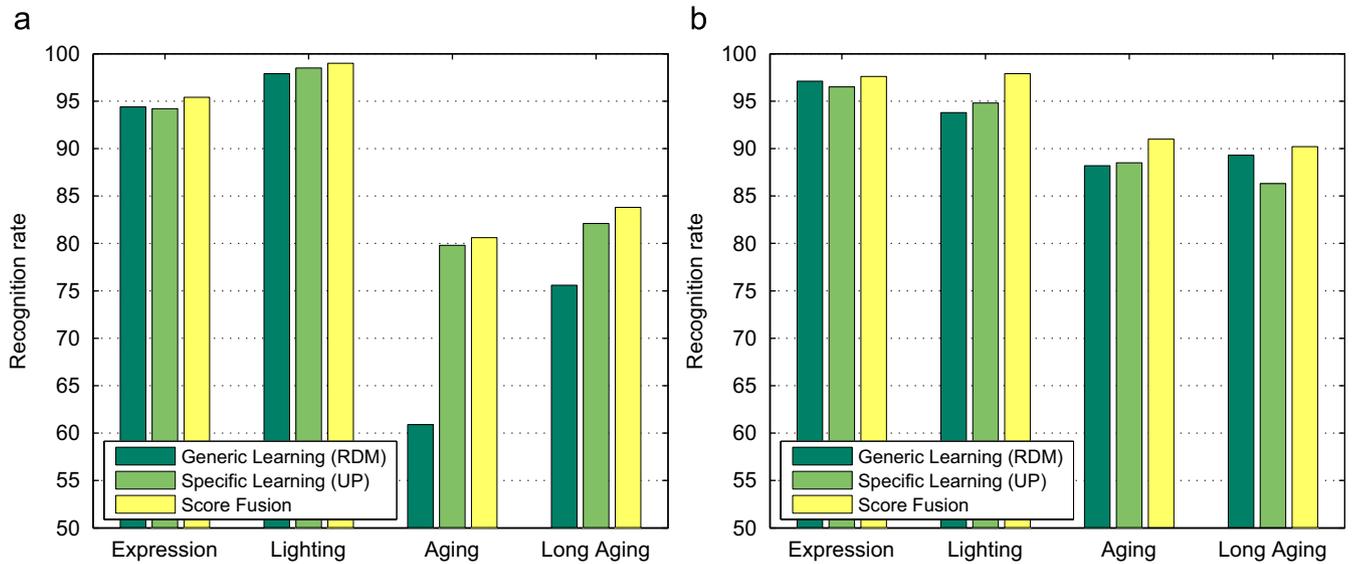


Fig. 11. Comparative FERET face recognition performance using the generic learning and specific learning. The generic learning is performed using the RDM method on the FRGC version 2 training database with 12,766 images, and the specific learning is performed using the UP algorithm on the 1196 FERET gallery images. (a) Gabor. (b) LBP.

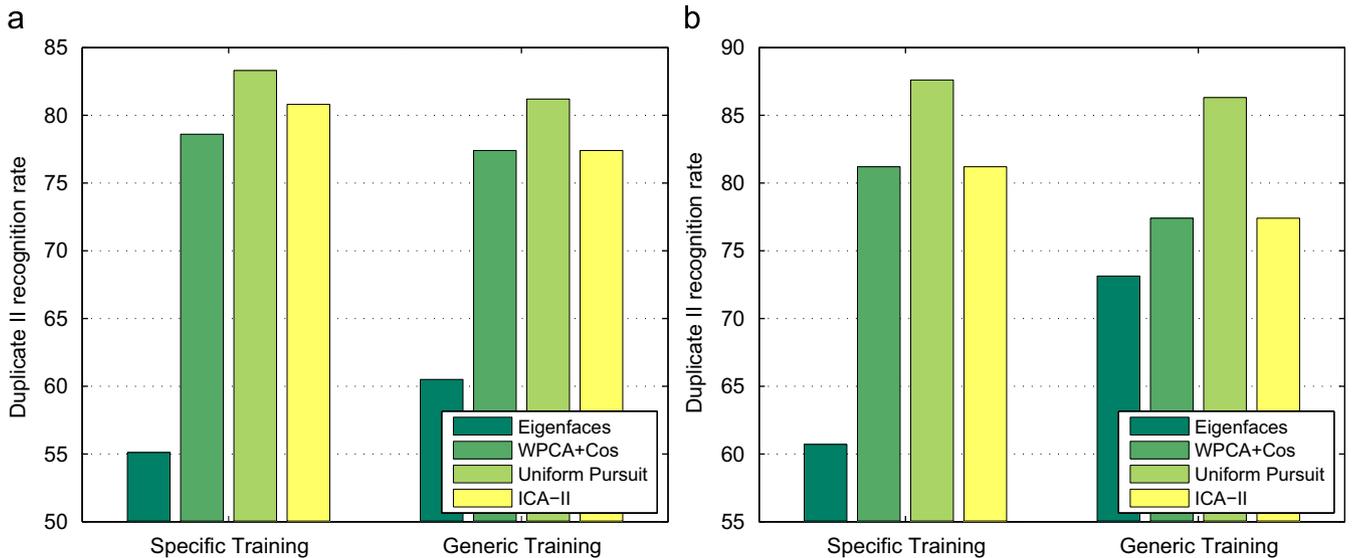


Fig. 12. Comparative FERET face recognition performance (duplicate II probe set) using the Eigenfaces (PCA), whitened PCA, uniform pursuit, and ICA (Architecture II) methods. Note that the term “generic learning” indicates that all the persons involved in the duplicate II probe set are excluded from the training image set in the subspace learning stage. (a) Gabor. (b) LBP.

learning situation. The stable performance of the UP algorithm in the generic learning manner can be explained in two ways. Firstly, the “unseen” face can be well reconstructed by a large training set of generic faces [39]. Secondly, the low dimensional features derived by the UP algorithm represent the discriminative characteristics between similar faces, and such discriminative characteristics can be commonly shared by different faces no matter they are in the training set or not.

5.6. Error analysis and practical issues

By analyzing the experimental results, we find that most recognition errors in our experiments are caused by two major factors. First, the large expression variations, such as the large smile and wink, between the probe image and corresponding gallery image. In fact, this cause of error has been addressed by

many security applications. Smiling in passport photos, especially the smile which exposes the teeth, has indeed been banned in a number of countries. Passport applicants can be asked to pose for a new photo if the first one is deemed too distorted by the act of smiling. Smiling and wink in passport photos can distort the subject’s eyes and change the relationship between facial feature points. At the same time, it is beneficial to include some side information in the practical face recognition system. For example, rejecting to recognize the facial image with large smile or wink may improve the robustness and accuracy of the system.

Second, there are some persons looking particularly similar. As shown in Fig. 13, although the images of the second row look very similar to those of the first row, they do not represent the same person. This phenomenon supports the objective of the proposed UP method which aims to encode the face in manner that distinguish the similar faces. Indeed, the UP algorithm also failed to distinguish some very similar faces as shown in Fig. 13. Note



Fig. 13. Error analysis of our experiments. First row: the probe images that are wrongly recognized. Second row: the recognition result, i.e., the most similar gallery image, determined by the UP algorithm. Third row: the ground true gallery image of corresponding person.



Fig. 14. The mobile robot recognizes the human face in the real-time video captured in the complex environment. The recognition is based on the single example stored in the gallery database as shown in (b). (a) Mobile robot. (b) Robot–human interface.

that, in some cases, the facial difference is so trivial so that even human observers are hard to judge the two faces are not from the same person. The discrimination between such similar faces may rely on the development of the more refined face descriptors that can capture the subtle facial differences between images. Certainly, other biometric features, such as fingerprint and voice, could be included.

Beside on the benchmark database, the feasibility of the UP algorithm also tested in the real-world circumstance. We implement the UP based face recognition system on a mobile robot in order to test its recognition accuracy in the varying circumstances, as shown in Fig. 14. Since the video based image quality is relatively low, the robot vision use the Gabor wavelet based features, rather than LBP, as the low-level representation. Up to now, the mobile robot has “memorized” more than 2000 persons, including the members of our laboratory (102 persons), the guests who visited our lab (56 persons), and the 1040 persons from the CAS-PEAL database [40], and the 1196 persons from the FERET database. With a gallery of this scale with one image per person, the recognition accuracy of the robot is about 94%.⁶ Our

ongoing work is to apply online learning technique to allow the robot to “memorizes” a new face in an online way and encode the face in manner that separates it from similar ones.

6. Conclusions and discussions

Current face recognition techniques rely heavily on the large size and representativeness of the training sets, and most methods suffer degraded performance or fail to work if there is only one training sample per person available. This so-called “one sample problem” is challenging in face recognition. In this paper, we propose a novel feature extraction method, called uniform pursuit, to address the one sample problem. Specifically, the UP method pursues, in the whitened PCA space, the low dimensional projections that reduce the local confusion between the similar persons. The resulting low dimensional projection features are robust against the complex image variations such as those caused by lighting and aging. The robustness, accuracy and efficiency of the new UP method has been successfully evaluated through experiments on challenging FERET frontal face images of 1196 subjects using only one training sample per person. In particular, the UP method achieves 96.7%, 99.0%, 88.9%, and 87.6% recognition rate on the fb, fc, dup I, and dup II probes, respectively, using less than 500 dimensional linearly transformed features.

⁶ The recognition accuracy is evaluated based on more than 1000 times recognition of our laboratory members in varying locations, such as entrance, corridor, and outdoor circumstance.

We also demonstrate that both generic learning and specific learning methodology are effective to address the one sample problem, and a combination of them is a promising way to increase recognition accuracy. Although it is not the main point of this paper, we would like to mention that the methods that are based on the image pixel may not be a good solution for one sample problem. Practical solutions of the one sample problem rely heavily on the informative low level image features, rather than raw pixel value. Finally, although we focus strictly on the one sample problem of face recognition in this paper, the basic idea of the UP might benefit a wide range of pattern recognition tasks besides face recognition and one sample problem. For instance, the idea of UP can be applied to maximize the pairwise distance between the centroids of the confusing classes, such as similar objects or similar characters, with additional considerations on the class conditional covariance.

Acknowledgments

The authors would like to thank the anonymous reviewers for their critical and constructive comments and suggestions, which are very helpful to improve both the technical and the literary quality of this paper. This work was done when W. Deng was a postgraduate exchange student in School of Information Technologies, University of Sydney. W. Deng would like to thank Prof. David Feng and Tom Cai for providing the excellent research environment in the BMIT Laboratory. This work was partially sponsored by CSC (China Scholarship Council) under Grant no. 2007104170, NSFC (Natural Science Foundation of China) under Grant no. 60905017, National High-tech Development Plan of China under Grant no. 2007AA01Z417 and 111 Project under Grant no. B08004.

References

- [1] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, *IEEE Trans. Image Process.* 11 (4) (2002) 467–476.
- [2] W. Deng, J. Guo, J. Hu, H. Zhang, Comment on “100% accuracy in automatic face recognition”, *Science* 321 (5891) (2008) 912.
- [3] A.J. O’Toole, P.J. Phillips, F. Jiang, J. Ayyad, N. Penard, H. Abdi, Face recognition algorithms surpass humans matching faces over changes in illumination, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (9) (2007) 1642–1646.
- [4] X. Tan, S. Chen, Z.-H. Zhou, F. Zhang, Face recognition from a single image per person: a survey, *Pattern Recognition* 39 (9) (2006) 1746–1762.
- [5] J. Wang, K. Plataniotis, J. Lu, A. Venetsanopoulos, On solving the face recognition problem with one training sample per subject, *Pattern Recognition* 39 (9) (2006) 1725–1745.
- [6] B. Zhang, S. Shan, X. Chen, W. Gao, Histogram of Gabor phase patterns (hgpp): a novel object representation approach for face recognition, *IEEE Trans. Image Process.* 16 (1) (2006) 57–68.
- [7] M. Turk, A. Pentland, Eigenfaces for recognition, *J. Cognitive Neurosci.* 3 (1) (1991) 71–86.
- [8] B. Moghaddam, W. Wahid, A. Pentland, Beyond eigenfaces: probabilistic matching for face recognition, in: *Automatic Face and Gesture Recognition, Proceedings of the Third IEEE International Conference, April 1998*, pp. 30–35.
- [9] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (7) (1997) 711–720.
- [10] C. Liu, H. Wechsler, Evolutionary pursuit and its application to face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (6) (2000) 570–582.
- [11] X. He, S. Yan, Y. Hu, P. Niyogi, H.-J. Zhang, Face recognition using Laplacianfaces, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (3) (2005) 328–340.
- [12] Z.-H.Z. Jianxin Wu, Face recognition with one training image per person, *Pattern Recognition Lett.* 23 (14) (2002) 1711–1719.
- [13] D. Zhang, S. Chen, Z.-H. Zhou, Enhanced (pc)2a for face recognition with one training image per person, *Pattern Recognition Lett.* 25 (10) (2004) 1173–1181.
- [14] D. Zhang, S. Chen, Z.-H. Zhou, A new face recognition method based on svd perturbation for single example image per person, *Appl. Math. Comput.* 163 (2) (2005) 895–907.
- [15] J. Wang, K. Plataniotis, A. Venetsanopoulos, Selecting discriminant eigenfaces for face recognition, *Pattern Recognition Lett.* 26 (10) (2005) 1470–1482.
- [16] J. Yang, D. Zhang, Two-dimensional pca: a new approach to appearance-based face representation and recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (1) (2004) 131–137.
- [17] L. Wang, X. Wang, J. Feng, On image matrix based feature extraction algorithms, *IEEE Trans. SMC-B* 36 (1) (2006) 194–197.
- [18] D. Zhang, Z.-H. Zhou, S. Chen, Diagonal principal component analysis for face recognition, *Pattern Recognition* 39 (1) (2006) 140–142.
- [19] D. Zhang, Z.-H. Zhou, (2d)2pca: 2-directional 2-dimensional pca for efficient face representation and recognition, *Neurocomputing* 69 (1) (2004) 131–137.
- [20] H. Moon, P.J. Phillips, Computational and performance aspects of pca-based face-recognition algorithms, *Perception* 30 (2001) 303–321.
- [21] P.J. Phillips, P.J. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: *IEEE International Conference on Computer Vision and Pattern Recognition, 2005*.
- [22] P.J. Phillips, P.J. Flynn, T. Scruggs, K. Bowyer, W. Worek, Preliminary face recognition grand challenge results, in: *Proceedings of the Seventh International Conference on Automatic Face and Gesture Recognition, 2006*, pp. 15–24.
- [23] C. Liu, The Bayes decision rule induced similarity measures, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (6) (2007) 1086–1090.
- [24] W. Deng, J. Hu, J. Guo, Gabor–Eigen–Whiten–Cosine: A Robust Scheme for Face Recognition, in: *Lecture Notes in Computer Science*, vol. 3723, Springer, Berlin, 2005.
- [25] L.B. Hieu, V. Nguyen, L. Shen, Local Gabor Binary Pattern Whiten PCA: A Novel Approach for Face Recognition from Single Image Per Person, in: *Lecture Notes in Computer Science*, vol. 5558, Springer, Berlin, 2009, pp. 269–278.
- [26] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Graph embedding: a general framework for dimensionality reduction, in: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005, pp. 830–837.
- [27] X. He, P. Niyogi, Locality preserving projections, in: *Proceedings of the Conference on Advances in Neural Information Processing System, 2003*.
- [28] J. Yang, D. Zhang, J.-Y. Yang, B. Niu, Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (4) (2007) 650–664.
- [29] W. Deng, J. Hu, J. Guo, H. Zhang, C. Zhang, Comments on “globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics”, *IEEE Trans. Pattern Anal. Mach. Intell.* 30 (8) (2008) 1503–1504.
- [30] D. Zhang, Z.-H. Zhou, S. Chen, Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft KNN ensemble, *IEEE Trans. Neural Networks* 16 (4) (2005) 875–886.
- [31] T. Ahonen, A. Hadid, M. Pietikinen, Face description with local binary patterns: application to face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (12) (2006) 2037–2041.
- [32] X. Tan, B. Triggs, Fusing Gabor and lbp Feature Sets for Kernel-based Face Recognition, in: *Lecture Notes in Computer Science*, vol. 4778, 2007, pp. 235–249.
- [33] J. Yang, D. Zhang, J. Yu Yang, Is ICA significantly better than pca for face recognition? in: *International Conference on Computer Vision, 2005*.
- [34] A. Hyvarinen, Fast and robust fixed-point algorithms for independent component analysis, *IEEE Trans. Neural Networks* 10 (3) (1999) 626–634.
- [35] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, Face recognition by independent component analysis, *IEEE Trans. Neural Networks* 13 (6) (2002) 1450–1460.
- [36] P.J. Phillips, H. Moon, P. Rizvi, P. Rauss, The Feret evaluation method for face recognition algorithms, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (2000) 0162–8828.
- [37] J. Hu, W. Deng, J. Guo, Robust Discriminant Analysis of Latent Semantic Features for Text Categorization, in: *Lecture Notes in Artificial Intelligence*, vol. 4223, 2006, pp. 400–409.
- [38] J. Kittler, M. Hatef, R.P. Duin, J. Matas, On combining classifiers, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (3) (1998) 226–239.
- [39] M. Meytlis, L. Sirovich, On the dimensionality of face space, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (7) (2007) 1262–1267.
- [40] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, D. Zhao, The cas-peal large-scale Chinese face database and baseline evaluations, *IEEE Trans. System Man Cybern. (Part A)* 38 (1) (2008) 149–161.

About the Author—WEIHONG DENG received the B.E. degree in information engineering and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2004 and 2009, respectively. From October 2007 to December 2008, he was a postgraduate exchange student in the School of Information Technologies, University of Sydney, Australia, under the support of the China Scholarship Council. He is currently a lecturer in School of Information and Communication Engineering, BUPT. His research interests include statistical pattern recognition and computer vision, with a particular emphasis in face

recognition. He has published over 10 technical papers in international journals and conferences, including a technical comment on face recognition in science magazine. He also serves as a reviewer for several international journals, such as IEEE TPAMI, Pattern Recognition and IEEE SMC-B.

About the Author—JIANI HU received the B.E. degree in telecommunication engineering from China University of Geosciences in 2003, and the Ph.D. degree in signal and information processing from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2008. She is currently a lecturer in School of Information and Communication Engineering, BUPT. Her research interests include information retrieval, statistical pattern recognition and computer vision.

About the Author—JUN GUO JUN GUO received B.E. and M.E. degrees from BUPT, China, in 1982 and 1985, respectively, Ph.D. degree from the Tohoku-Gakuin University, Japan, in 1993. At present he is a professor and the dean of School of Information and Communication Engineering, BUPT. His research interests include pattern recognition theory and application, information retrieval, content based information security, and network management. He has published over 200 papers, some of them are on world-wide famous journals or conferences including SCIENCE, IEEE Transactions on PAMI, IEICE, ICPR, ICCV, SIGIR, etc. His book "Network management" was awarded by the government of Beijing city as a finest textbook for higher education in 2004.

His team got a number of prizes in national and international academic competitions including: the first place in a national test of handwritten Chinese character recognition 1995, the first place in a national test of face detection 2004, the first place in a national test of text classification 2004, the first place of paper design competition held by IEEE Industry Application Society 2005, the second place in the competition of CSIDC held by IEEE Computer Society 2006.

About the Author—WEIDONG CAI received the B.S. degree in Computer Science from HuaQiao University, Quanzhou, China, in 1989, and the Ph.D. degree from the University of Sydney, Sydney, Australia, in 2001, both in computer science. Prior to his doctoral study, he worked in industry for five years. After graduation, he was a Postdoctoral Research Associate at the Centre for Multimedia Signal Processing (CMSP), Hong Kong Polytechnic University. In 2001, he was a Lecturer and is currently a Senior Lecturer in the School of Information Technologies, University of Sydney. His research interests include computer graphics, image processing and analysis, data compression and retrieval, and multimedia database and computer modeling with biomedical applications.

About the Author—DAGAN FENG received the M.E. degree in electrical engineering and computing science from Shanghai JiaoTong University, Shanghai, China, in 1982, and the M.Sc. degree in biocybernetics and the Ph.D. degree in computer science from the University of California, Los Angeles, in 1985 and 1988, respectively. After briefly working as an Assistant Professor at the University of California, Riverside, he joined the University of Sydney, Sydney, Australia, at the end of 1988, where he was a Lecturer, Senior Lecturer, Reader, Professor, Head of Department of Computer Science, Head of School of Information Technologies, and is currently Associate Dean of the Faculty of Science. He is also the Honorary Research Consultant, Royal Prince Alfred Hospital, Sydney; the Chair-Professor of Information Technology, Hong Kong Polytechnic University, HongKong; the Advisory Professor, Shanghai JiaoTong University; and a Guest Professor with Northwestern Polytechnic University, Xian, China, with Northeastern University, Shenyang, China, and with Tsinghua University, Beijing, China. He is the Founder and Director of the Biomedical and Multimedia Information Technology Research Group. He has published over 400 scholarly research papers, pioneered several new research directions, and made a number of landmark contributions in his field with significant scientific impact and social benefit. His research area is biomedical and multimedia information technology.

Dr. Feng is a Fellow of the Australia Computer Society, the Australian Academy of Technological Sciences and Engineering, Hong Kong Institution of Engineers, and the Institution of Electrical Engineers, UK, and the Institute of Electrical and Electronics Engineers. He is also the special Area Editor of the IEEE Transactions on Information Technology in Biomedicine and is the current Chairman of IFAC-TC-BIOMED. He is the recipient of the Crump Prize for Excellence in Medical Engineering (USA).