# Stepwise Correlation Metric Based Discriminant Analysis and Multi-Probe Images Fusion for Face Recognition

Zhen Lei       Shengcai Liao       Stan Z. Li*

Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun Donglu, Beijing 100190, China.
{zlei,scliao,szli}@cbsr.ia.ac.cn

## Abstract

*Face recognition is a great challenge in practice. Subspace learning method is one of the dominant methods and has achieved great success in face recognition area. In subspace learning, many researches have found that correlation similarity (e.g. cosine distance) usually achieves better classification results than L2 distance with nearest neighbor (NN) classifier in Euclidean space. However, in traditional methods, most of them are devoted to optimize the objective function based on L2 distance, which is not coincident with the classification rule. It is reasonable to obtain better results by optimizing the objective function with correlation metric directly. In this paper, following traditional linear discriminant analysis (LDA), we redefine the between and with-in class scatter with correlation metric and propose an efficient Stepwise Correlation metric based Discriminant Analysis (SCDA) method to derive the sub-optimal discriminant subspace to be classified with correlation similarity. Moreover, we propose a novel weighted fusion mechanism to learn the optimal combination of multi-probe images to be classified. Extensive experiments on PIE and extended Yale-B databases validate the effectiveness of SCDA and the learning based weighted image fusion method.*

## 1. Introduction

Face recognition is a great challenge in practice, especially in surveillance scenarios due to low resolution images and variant illuminations. During various methods, subspace learning is one of the dominant methods and has achieved great success in face recognition. The objective of subspace learning is to find the low embedding data structure from the original high dimension space by preserving the neighborhood relationship or exploiting the discriminant properties.

Principle component analysis (PCA) [16] and linear discriminant analysis (LDA) [1] are two representative ones in subspace learning. PCA uses the Karhunen-Loeve transform to produce the most expressive subspace for object representation and recognition by minimizing the residua of the reconstruction. However, for classification task, it does not consider any class information, so it may drop some important clues for classification. LDA is then proposed to seek optimum discriminant subspace of features best separating different object classes by maximizing the ratio of the between-classes scatter to the within-class scatter. There are also a lot of other work for subspace learning, such as ICA [3], ISOMAP [15], LLE [13], LPP [8], NPE [7] etc. Recently, Wang and Tang [17] unify PCA, LDA and Bayesian methods into the same framework and present the method to find the optimal configuration for LDA. Yan et al. [18] summary the existing various subspace learning methods and re-interpret them from the view of graph embedding.

During these subspace learning methods, most of them adopt L2-norm metric (Eq. 1) in Euclidean space to measure the dissimilarity of samples in training phase.

$$d_{L_2}(\mathbf{x_1}, \mathbf{x_2}) = (\mathbf{x_1} - \mathbf{x_2})^{\mathbf{T}}(\mathbf{x_1} - \mathbf{x_2}) \tag{1}$$

$$d_{cos}(\mathbf{x_1}, \mathbf{x_2}) = \frac{\mathbf{x_1^T x_2}}{\sqrt{\mathbf{x_1^T x_1 x_2^T x_2}}} \tag{2}$$

However, in test phase, many researches have reported that the correlation (cosine) based similarity (Eq. 2) could achieve better results than L2-norm metric with nearest neighbor (NN) classifier (see e.g. [10]). In this case, the similarity metric used in test phase is not accordant with that in training phase. It is reasonable that learning the discriminant subspace using correlation similarity metric in training phase may be helpful for improving the face recognition performance. This is the main motivation of this work.

---

*Corresponding author.

Recently, there are some work to optimize correlation metric based criterion. Ma et al. [12] propose a discriminant analysis method with correlation similarity measure. However, their method does not take into account the reduction problem and actually only re-scales the samples in original space. Fu et al. [5] propose a dimensionality reduction method based on correlation metric. It utilizes the traditional LDA result as an initialization and adopts gradient descend optimization method to derive the most discriminant projective directions. However, it needs to restart the whole optimization process if the reduced dimension is adjusted. Nonetheless, the computational cost for the optimization process is very high and limits its application.

In this paper, we present an effective and efficient stepwise way to solve the discriminant feature dimensionality reduction problem with correlation (cosine) similarity measure. That is, at each step, the derived subspace is only one dimension lower than the former one. Comparing with existing methods, our method is much more efficient during training phase and can retain arbitrary dimension of feature conveniently like traditional ones do. In addition, in recognition phase, we propose a novel learning based weighted multi-probe images fusion method to enhance the face recognition performance.

The remainder of the paper is organized as follows. Section 2 reviews the traditional discriminant method and formulates the correlation similarity based problem. Section 3 details the stepwise correlation metric based discriminant analysis solution in training and test phase. Section 4 describes the weight optimization process for fusing multi-probe images. Experiments compared with prevailing methods are presented in Section 5 and in Section 6, we conclude the paper.

## 2. Formulation of LDA based on L2-norm and Correlation Metric

Linear discriminant analysis (LDA), as a representative method in supervised learning, has been widely used in computer vision and image classification. The essential idea of LDA is to disperse the samples from different classes and meanwhile gather the samples from the same class in subspace. In traditional way, given the sample set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$, the between class scatter matrix $S_b$ and with-in class scatter matrix $S_w$ based on L2-norm are defined as

$$
\begin{aligned}
S_w &= \frac{1}{n} \sum_{i=1}^{L} \sum_{\mathbf{x}_j \in C_i} (\mathbf{x}_j - \mathbf{m}_i)(\mathbf{x}_j - \mathbf{m}_i)^T \\
S_b &= \frac{1}{n} \sum_{i=1}^{L} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T
\end{aligned}
\tag{3}
$$

where $\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x}_j \in C_i} \mathbf{x}_j$ is the mean vector of data in class $C_i$, and $\mathbf{m} = \frac{1}{n} \sum_{i=1}^{L} \sum_{X_j \in C_i} \mathbf{x}_j$ is the global mean vector. LDA aims to learn the projective directions $\mathbf{W}$ which maximize the ratio of between class scatter matrix to with-in class scatter one as

$$
\begin{aligned}
J &= \frac{tr(\sum_{i=1}^{L} n_i \mathbf{W}^T (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \mathbf{W})}{tr(\sum_{i=1}^{L} \sum_{\mathbf{x}_j \in C_i} \mathbf{W}^T (\mathbf{x}_j - \mathbf{m}_i)(\mathbf{x}_j - \mathbf{m}_i)^T \mathbf{W})} \\
&= \frac{tr(\mathbf{W}^T S_b \mathbf{W})}{tr(\mathbf{W}^T S_w \mathbf{W})}
\end{aligned}
\tag{4}
$$

The optimal projection matrix $\mathbf{W}_{opt}$ can be obtained by solving the following eigen-value problem

$$
S_w^{-1} S_b \mathbf{W} = \mathbf{W}\mathbf{\Lambda}
\tag{5}
$$

where $\mathbf{\Lambda}$ is the diagonal matrix whose diagonal elements are eigenvalues of $S_w^{-1} S_b$.

Similarly, we can define the between and with-in class scatter based on correlation metric by summing up the correlation values of sample pairs from different classes and the same class respectively as

$$
\begin{aligned}
S_b &= \sum_{k=1}^{L} \sum_{m=k+1}^{L} \sum_{\substack{\mathbf{x}_i \in C_k \\ \mathbf{x}_j \in C_m}} \frac{\mathbf{x}_i^T \mathbf{x}_j}{\sqrt{\mathbf{x}_i^T \mathbf{x}_i \mathbf{x}_j^T \mathbf{x}_j}} \\
S_w &= \sum_{k=1}^{L} \sum_{\substack{\mathbf{x}_i, \mathbf{x}_j \in C_k \\ \mathbf{x}_i \neq \mathbf{x}_j}} \frac{\mathbf{x}_i^T \mathbf{x}_j}{\sqrt{\mathbf{x}_i^T \mathbf{x}_i \mathbf{x}_j^T \mathbf{x}_j}}
\end{aligned}
\tag{6}
$$

The purpose of correlation metric based LDA is to find the projective directions $\mathbf{W}$ that maximize the correlation of samples from the same class and meanwhile minimize the correlation of samples from different classes which equals to minimize the ratio of between class correlation scatter to with-in one.

$$
\mathbf{W} = \arg\min_{\mathbf{W}} J(\mathbf{W})
$$

$$
J(\mathbf{W}) = \frac{\sum_{k=1}^{L} \sum_{m=k+1}^{L} \sum_{\substack{\mathbf{x}_i \in C_k \\ \mathbf{x}_j \in C_m}} \frac{\mathbf{x}_i^T \mathbf{W}\mathbf{W}^T \mathbf{x}_j}{\sqrt{\mathbf{x}_i^T \mathbf{W}\mathbf{W}^T \mathbf{x}_i \mathbf{x}_j^T \mathbf{W}\mathbf{W}^T \mathbf{x}_j}}}{\sum_{k=1}^{L} \sum_{\substack{\mathbf{x}_i, \mathbf{x}_j \in C_k \\ \mathbf{x}_i \neq \mathbf{x}_j}} \frac{\mathbf{x}_i^T \mathbf{W}\mathbf{W}^T \mathbf{x}_j}{\sqrt{\mathbf{x}_i^T \mathbf{W}\mathbf{W}^T \mathbf{x}_i \mathbf{x}_j^T \mathbf{W}\mathbf{W}^T \mathbf{x}_j}}}
\tag{7}
$$

Unfortunately, the above optimization problem is usually non-convex and lacks close solution. Fu et al. [5] tries to adopt gradient descend method to solve this problem. However, the optimization process is really inefficient and is easy to converge to local extremum.

# 3. Stepwise Correlation Metric based Discriminant Analysis

As we know, if the L2-norm length of sample vector is unit, the cosine distance is equivalent with the L2 distance regardless of a constant value. In this way, we can transform the correlation related problem to the formulation that maximizes the ratio of between class scatter to the with-in class scatter computed based on L2 distance and meanwhile preserves the unit L2-norm property of samples. Consequently, the problem in Eq. 7 can be reformulated as

$$\mathbf{W} = \arg\max_{\mathbf{W}} J(\mathbf{W})$$

$$J(\mathbf{W}) = \frac{\sum_{k=1}^{L}\sum_{m=k+1}^{L}\sum_{\substack{\mathbf{x}_i \in C_k \\ \mathbf{x}_j \in C_m}} (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{W}\mathbf{W}^T (\mathbf{x}_i - \mathbf{x}_j)}{\sum_{k=1}^{L}\sum_{\substack{\mathbf{x}_i,\mathbf{x}_j \in C_k \\ \mathbf{x}_i \neq \mathbf{x}_j}} (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{W}\mathbf{W}^T (\mathbf{x}_i - \mathbf{x}_j)}$$

$$= \frac{\sum_{k=1}^{L}\sum_{m=k+1}^{L}\sum_{\substack{\mathbf{x}_i \in C_k \\ \mathbf{x}_j \in C_m}} tr(\mathbf{W}^T (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{W})}{\sum_{k=1}^{L}\sum_{\substack{\mathbf{x}_i,\mathbf{x}_j \in C_k \\ \mathbf{x}_i \neq \mathbf{x}_j}} tr(\mathbf{W}^T (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{W})}$$

$s.t.$

$$\mathbf{x}_i^T \mathbf{W}\mathbf{W}^T \mathbf{x}_i = 1 \qquad i = 1, 2, \cdots, n$$

(8)

The problem stated above is still difficult to solve directly. In following, we adopt a sub-optimal way to relax the constraints a bit. It is known that PCA aims to find the subspace that minimizes the reconstruction error. Therefore, in this work, we incorporate the PCA goal into the objective function to preserve the constraints in Eq. 8 as much as possible.

$$\mathbf{W} = \arg\max_{\mathbf{W}} J(\mathbf{W})$$

$$J(\mathbf{W}) = \frac{tr(\mathbf{W}^T (\tilde{S}_b + \eta \tilde{S}_t)\mathbf{W})}{tr(\mathbf{W}^T \tilde{S}_w \mathbf{W})}$$

$$\tilde{S}_b = \sum_{k=1}^{L}\sum_{m=k+1}^{L}\sum_{\substack{\mathbf{x}_i \in C_k \\ \mathbf{x}_j \in C_m}} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T$$

$$\tilde{S}_w = \sum_{k=1}^{L}\sum_{\substack{\mathbf{x}_i,\mathbf{x}_j \in C_k \\ \mathbf{x}_i \neq \mathbf{x}_j}} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \qquad (9)$$

$$\tilde{S}_t = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T$$

$s.t.$

$$\mathbf{x}_i^T \mathbf{x}_i = 1 \qquad i = 1, 2, \cdots, n$$

where the second term in numerator of $J(\mathbf{W})$ is the PCA optimization objective item which tries to preserve the unit L2-norm property of sample vector in the reduced subspace and

---

**Training Phase:**
**Input**: Let $d$-dimensional sample set be $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ from $L$ classes $\{C_1, C_2, \ldots, C_L\}$. We expect to find $t$-dimensional discriminant subspace.
**Initialization**: $\mathbf{W} = \mathbf{I}_{d \times d}$
for $s = 1 : d - t$

1. normalize the sample set $\mathbf{X}$ to be of unit L2-norm.

2. compute $\tilde{S}_b, \tilde{S}_w, \tilde{S}_t$ according to Eq. 9 and derive the $d - s$ projective directions $\mathbf{W}_{d-s+1}^{d-s}$ by solving the generalized eigen-value problem (Eq. 10).

3. $\mathbf{W} = \mathbf{W} \cdot \mathbf{W}_{d-s+1}^{d-s}, \qquad \mathbf{X} = \mathbf{W}_{d-s+1}^{d-sT}\mathbf{X}$.

end
**Output**:The projective matrix from $d$ to $t$ dimension $\mathbf{W}$.

**Testing Phase:**
1. project the sample into subspace by the learned projections $\mathbf{W}$: $\mathbf{x}' = \mathbf{W}^T\mathbf{x}$.

2. using correlation metric in projected subspace to measure the dissimilarity of samples.

Figure 1. Stepwise correlation metric based discriminant analysis algorithm.

$\eta$ is the coefficient that controls the trade-off between the unit L2-norm property preservation and discrimination of subspace. In order to preserve the unit L2-norm of sample vectors as much as possible, we adopt a stepwise way to reduce the dimension. At each step, the sample vector in current $d$ space is first normalized to be of unit L2-norm length. Our purpose is to find the $d-1$ subspace that maximizes the ratio of between class scatter to the with-in one based on L2-norm metric and meanwhile preserve the unit L2-norm property as much as possible. After that, we re-normalize the sample vector in the projected subspace and find the next one-lower discriminant subspace. In this way, one can obtain arbitrary dimension of discriminant subspace iteratively and finally obtain a sub-optimal solution to Eq. 8. The solution in each step can be obtained by solving the following generalized eigen-value problem corresponding to the $d-1$ largest eigenvalues efficiently.

$$(S_b + \eta S_t)\mathbf{W} = S_w\mathbf{W}\mathbf{\Lambda} \qquad (10)$$

The whole process of proposed stepwise correlation metric based discriminant analysis (SCDA) algorithm is illustrated in Fig. 1. It is worth noting the proposed SCDA is different from the Foley-Sammon method [4] that in each iteration, the feature vectors are needed to be re-normalized and the derived solutions are not necessarily orthogonal.

In test phase, intuitively, we need to follow the training

procedure to reduce the original sample vector in stepwise way and re-normalize it at each step which is obviously inefficient and increases the burden of computation. Fortunately, from theorem 1, we can see that the lower subspace projection of SCDA can be computed by one-step directly as traditional methods do.

**Theorem 1** *If* $\boldsymbol{x}_t = \frac{1}{N_t}\boldsymbol{W}_{t+1}^{tT}\frac{1}{N_{t+1}}\boldsymbol{W}_{t+2}^{t+1T}\cdots\frac{1}{N_{d-1}}\boldsymbol{W}_d^{d-1T}\frac{1}{||\boldsymbol{x}_d||}\boldsymbol{x}_d$, *where* $N_k = ||\boldsymbol{W}_{k+1}^{kT}\frac{1}{N_{k+1}}\boldsymbol{W}_{k+2}^{k+1T}\cdots\frac{1}{N_{d-1}}\boldsymbol{W}_d^{d-1T}\frac{1}{||\boldsymbol{x}_d||}\boldsymbol{x}_d||$, *then* $\boldsymbol{x}_t = \frac{1}{||\boldsymbol{W}_d^{tT}\boldsymbol{x}_d||}\boldsymbol{W}_d^{tT}\boldsymbol{x}_d$, *where* $\boldsymbol{W}_d^t = \boldsymbol{W}_d^{d-1}\boldsymbol{W}_{d-1}^{d-2}\cdots\boldsymbol{W}_{t+1}^t$.

**Proof:**

$$
\begin{aligned}
\mathbf{x}_t &= \frac{\mathbf{W}_{t+1}^{tT}\mathbf{x}_{t+1}}{N_t} = \frac{\mathbf{W}_{t+1}^{tT}\mathbf{W}_{t+2}^{t+1T}\mathbf{x}_{t+2}}{N_t N_{t+1}} = \cdots \\
&= \frac{\mathbf{W}_{t+1}^{tT}\mathbf{W}_{t+2}^{t+1T}\cdots\mathbf{W}_d^{d-1T}\mathbf{x}_d}{||\mathbf{x}_d||N_{d-1}\cdots N_t} \\
&= \frac{\mathbf{W}_d^{tT}\mathbf{x}_d}{||\mathbf{x}_d||N_{d-1}\cdots N_t}
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
N_t &= ||\mathbf{W}_{t+1}^{tT}\frac{1}{N_{t+1}}\mathbf{W}_{t+2}^{t+1T}\cdots\frac{1}{N_{d-1}}\mathbf{W}_d^{d-1T}\frac{1}{||\mathbf{x}_d||}\mathbf{x}_d|| \\
&= \frac{1}{N_{t+1}N_{t+2}\cdots N_{d-1}||\mathbf{x}_d||}||\mathbf{W}_{t+1}^{tT}\mathbf{W}_{t+2}^{t+1T}\cdots\mathbf{W}_d^{d-1T}\mathbf{x}_d|| \\
&= \frac{1}{N_{t+1}N_{t+2}\cdots N_{d-1}||\mathbf{x}_d||}||\mathbf{W}_d^{tT}\mathbf{x}_d|| \\
&\Longrightarrow N_t N_{t+1}\cdots N_{d-1}||\mathbf{x}_d|| = ||\mathbf{W}_d^{tT}\mathbf{x}_d||
\end{aligned}
\tag{12}
$$

Substituting Eq. 12 into Eq. 11, we obtain:

$$
\mathbf{x}_t = \frac{\mathbf{W}_d^{tT}\mathbf{x}_d}{||\mathbf{W}_d^{tT}\mathbf{x}_d||}
$$

## 4. Learning Optimal Weights for Multi-Probe Images Fusion

In real application such as surveillance scenario or video based face recognition, it is easy to obtain multiple probe images for one person. It is well known multiple images fusion could significantly improve the face recognition performance. Various fusion rules such as sum, max, min, product etc. [9] at score level have been researched. In [2], authors indicate that the simple utilization of mean image of multiple images could achieve better results than single one. It is obvious that simple mean image may not be the best way to combine multi-probe images. Thus, in this paper, we propose to learn the optimal weights for combining probe images to obtain the best representation.

Given a gallery image $\mathbf{g}$ and a series of probe images $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n]$, our purpose is to learn the weights for different probe images so that the difference between gallery image and the combination of probe images is the smallest. The problem can be formulated by minimizing the following objective function.

$$
J = ||\mathbf{g} - \sum_{i=1}^{n} w_i\mathbf{p}_i||^2 = ||\mathbf{g} - \mathbf{Pw}||^2, \quad w_i \geq 0 \tag{13}
$$

where $\mathbf{w} = [w_1, w_2, \cdots, w_n]^T$. Since the image pixel value is non-negative, we impose the non-negative constraints on weights to avoid the meaningless combination result. This problem can then be solved by the similar way as NMF [11] does. The gradient of $J$ with respect to $\mathbf{w}$ is computed as

$$
\frac{\partial J}{\partial w_i} = -2[\mathbf{p}_i^T\mathbf{g} - \mathbf{p}_i^T\mathbf{Pw}] \tag{14}
$$

According to gradient descend algorithm, the solution $w_i$ will be updated by $w_i \leftarrow w_i + \lambda_i[\mathbf{p}_i^T\mathbf{g} - \mathbf{p}_i^T\mathbf{Pw}]$ at each step. Like NMF, if $\lambda_i$ is selected as $\lambda_i = \frac{w_i}{\mathbf{p}_i^T\mathbf{Pw}}$, the updated process becomes $w_i \leftarrow w_i\frac{\mathbf{p}_i^T\mathbf{g}}{\mathbf{p}_i^T\mathbf{Pw}}$ which obviously guarantees the non-negative of $w_i$ and has been proved to make the objective function descending [11]. Therefore, after some iterations, the optimal weights can be learned to synthesize a new probe image which is utilized to be compared with the gallery image later.

## 5. Experiments

### 5.1. Database Preparation

Two databases, PIE [14] and extended Yale-B [6] are used to evaluate the performance of different methods. For PIE database, there are $41,368$ images of $68$ people under different poses, illumination conditions and expressions. Five near frontal poses (C05, C07, C09, C27, C29) and all the images under different illuminations and expressions are selected. So there are totally 170 images for each individual. For extended Yale-B database, there are 2414 images from 38 persons each of which contains nearly $64$ ones. All the images are rotated, scaled and cropped to $32 \times 32$ according to the provided eye positions. No further preprocessing is adopted. Fig. 2 illustrates some cropped face examples of PIE and extended Yale-B databases.

The images are partitioned into gallery and probe sets randomly with different numbers. For ease of representation, in following experiments, $G_m$ means $m$ images per person are selected randomly as gallery set for training and the left images consist the probe set for testing.

### 5.2. Experiment I: SCDA Performance Evaluation

We evaluate the performance of SCDA with PCA and LDA using different metric measurements.

Figure 2. Cropped face images from PIE (a) and extended Yale-B (b) databases.

Tables 1 and 2 illustrate the performance comparisons of different methods on PIE and extended Yale-B databases and Fig. 3 shows the SCDA recognition rate curves with respect to different dimension of reduced feature. Each result is reported as the average of 10 different partitions. For Norm+LDA method, the samples are first normalized to be of unit L2-norm before LDA. Due to high computational cost, we don't compare CEA [5] method in this work. In tables 1 and 2, we report the best results of each method by varying the reduced dimension of feature. From the results, we can observe that for face recognition on PIE and extended Yale-B databases, the performance of correlation based similarity is usually better than that of the L2 distance in Euclidean space. And the normalization preprocessing step is helpful to improve the recognition performance. In most cases, the proposed SCDA method, incorporating the correlation metric in training phase which is consistent with the similarity measure in test phase, achieves better result than other methods which conforms to our motivation.

Table 1. Recognition rates on the PIE database (The result for each method is reported as the best one by varying the retained feature dimension).

|  | PCA+Cos | PCA + L2 | LDA + Cos |
|---|---|---|---|
| G5 | $0.2780 \pm 0.0065$ | $0.2187 \pm 0.0049$ | $0.5031 \pm 0.0629$ |
| G10 | $0.4065 \pm 0.0097$ | $0.3234 \pm 0.0074$ | $0.8358 \pm 0.0087$ |
| G20 | $0.5654 \pm 0.006$ | $0.4716 \pm 0.0055$ | $0.9295 \pm 0.0035$ |
| G30 | $0.6669 \pm 0.0054$ | $0.5730 \pm 0.0051$ | $0.9536 \pm 0.0034$ |
|  | LDA + L2 | Norm+LDA+Cos | SCDA+Cos |
| G5 | $0.4723 \pm 0.0395$ | $0.5448 \pm 0.0483$ | $\mathbf{0.5683} \pm 0.0555$ |
| G10 | $0.7709 \pm 0.0107$ | $0.8537 \pm 0.0077$ | $\mathbf{0.8648} \pm 0.009$ |
| G20 | $0.8942 \pm 0.0047$ | $0.9336 \pm 0.0039$ | $\mathbf{0.9386} \pm 0.0033$ |
| G30 | $0.9341 \pm 0.0031$ | $0.9556 \pm 0.0034$ | $\mathbf{0.9598} \pm 0.0028$ |

Table 2. Recognition rates on the extended Yale-B database (The result for each method is reported as the best one by varying the retained feature dimension).

|  | PCA+Cos | PCA + L2 | LDA + Cos |
|---|---|---|---|
| G5 | $0.3000 \pm 0.012$ | $0.2580 \pm 0.012$ | $0.5917 \pm 0.0542$ |
| G10 | $0.4216 \pm 0.0094$ | $0.3591 \pm 0.0084$ | $0.8382 \pm 0.0102$ |
| G20 | $0.5336 \pm 0.0095$ | $0.4558 \pm 0.0064$ | $0.9189 \pm 0.01$ |
| G30 | $0.5898 \pm 0.0062$ | $0.5157 \pm 0.0125$ | $0.9458 \pm 0.0086$ |
|  | LDA + L2 | Norm+LDA+Cos | SCDA+Cos |
| G5 | $0.5719 \pm 0.0404$ | $\mathbf{0.6291} \pm 0.0397$ | $0.6193 \pm 0.0438$ |
| G10 | $0.7999 \pm 0.0134$ | $0.8441 \pm 0.0112$ | $\mathbf{0.8522} \pm 0.0151$ |
| G20 | $0.8875 \pm 0.0123$ | $0.9281 \pm 0.0135$ | $\mathbf{0.9317} \pm 0.013$ |
| G30 | $0.9181 \pm 0.0067$ | $0.9553 \pm 0.0071$ | $\mathbf{0.9600} \pm 0.0074$ |

### 5.3. Experiment II: SCDA with Multi-Probe Images Fusion

In this experiment, we try to validate the effectiveness of proposed multi-probe images fusion compared with other existing fusion methods such as sum-rule, max-rule and the mean image strategy in [2]. The probe images for every person are divided into several groups each of which contains five images. In recognition phase, these five images are considered as a whole to be compared with the images in gallery set.
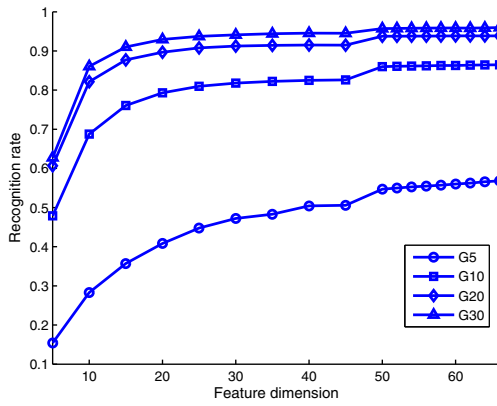
Tables 3 and 4 illustrates the recognition rates of different fusion methods with SCDA on PIE and extended Yale-B databases respectively. The dimension of SCDA subspace remains $C - 1$, where $C$ is the class number. For sum-rule and max-rule, the five probe images are first compared with the gallery image respectively to obtain five similarity scores and then the sum of these scores or the maximum value is used for the classification. For mean image and the proposed weighted mean image method, the five probe images are combined with equal or the learned weights to form a new image and then the similarity score

Table 3. Recognition rates of SCDA with different fusion methods on the PIE database.
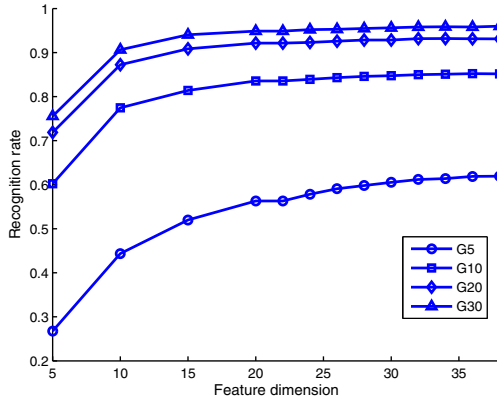
|    | SCDA+Single Image | SCDA+Sum-rule | SCDA+Max-rule | SCDA + Mean | SCDA+Weighted Mean |
|----|-------------------|---------------|---------------|-------------|--------------------|
| G2 | $0.4085 \pm 0.0154$ | $0.4543 \pm 0.0152$ | $\mathbf{0.5165} \pm 0.0148$ | $0.4555 \pm 0.0153$ | $0.4964 \pm 0.0144$ |
| G4 | $0.6405 \pm 0.0268$ | $0.7342 \pm 0.0334$ | $0.7595 \pm 0.0384$ | $0.7301 \pm 0.0335$ | $\mathbf{0.7664} \pm 0.0360$ |
| G6 | $0.7314 \pm 0.0104$ | $0.8310 \pm 0.0114$ | $0.8422 \pm 0.0131$ | $0.8234 \pm 0.0120$ | $\mathbf{0.8517} \pm 0.0113$ |
| G8 | $0.7560 \pm 0.0129$ | $0.8559 \pm 0.0140$ | $0.8624 \pm 0.0127$ | $0.8502 \pm 0.0136$ | $\mathbf{0.8721} \pm 0.0134$ |

Table 4. Recognition rates of SCDA with different fusion methods on the extended Yale-B database.

|    | SCDA+Single Image | SCDA+Sum-rule | SCDA+Max-rule | SCDA + Mean | SCDA+Weighted Mean |
|----|-------------------|---------------|---------------|-------------|--------------------|
| G2 | $0.4371 \pm 0.0487$ | $0.4709 \pm 0.0437$ | $\mathbf{0.5874} \pm 0.0447$ | $0.4832 \pm 0.0473$ | $0.5810 \pm 0.0429$ |
| G4 | $0.6927 \pm 0.0257$ | $0.7866 \pm 0.0300$ | $0.8452 \pm 0.0323$ | $0.7880 \pm 0.0322$ | $\mathbf{0.8539} \pm 0.0355$ |
| G6 | $0.7916 \pm 0.0222$ | $0.9128 \pm 0.0201$ | $0.9284 \pm 0.0161$ | $0.9111 \pm 0.0185$ | $\mathbf{0.9405} \pm 0.0162$ |
| G8 | $0.8355 \pm 0.0149$ | $0.9396 \pm 0.0110$ | $0.9452 \pm 0.0120$ | $0.9343 \pm 0.0106$ | $\mathbf{0.9548} \pm 0.0089$ |



(a)



(b)

Figure 3. Face recognition rate of SCDA with respect to different feature dimension on PIE (a) and extended Yale-B (b) databases.

is computed between the new probe image and the gallery one. For the method based on single image, one probe image is randomly selected from the five ones and com-

pared with the gallery one. Each result is reported as the average of 10 different partitions. From the results, we can observe that image fusion can significantly enhance the face recognition rate and in most cases, the proposed learning based weighted combination mechanism achieves the highest recognition rate which indicates the effectiveness of weighted multi-probe images fusion method.

## 6. Conclusions

There are two main contributions in this paper. First, a stepwise correlation metric based discriminant analysis (SCDA) is proposed in which the similarity measure (i.e. correlation metric) in training and test phases is consistent. Experimental results show SCDA usually outperforms the traditional LDA method, especially in the case of small sample size. The proposed stepwise reduction mechanism not only fits to the LDA, but also can be combined with other subspace learning methods such as LPP, NPE etc. Compared with the existing correlation metric based methods, SCDA avoids the time-consuming gradient computation process and thus is much more efficient and convenient to retain arbitrary dimension feature in training phase. Second, we propose a novel multi-probe images fusion method. Unlike the traditional score level fusion ones, the learning based weighted multi-probe images fusion is conducted on the image data level and exhibits the superiority over the existing fusion methods.

# References

[1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997. 1

[2] S. Chen, E. Berglund, A. Bigdeli, C. Sanderson, and B. C. Lovell. Experimental analysis of face recognition on still and cctv images. In *Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 317–324, 2008. 4, 5

[3] P. Comon. Independent component analysis - a new concept? *Signal Processing*, 36:287–314, 1994. 1

[4] D. H. Foley and J. W. Sammon. An optimal set of discriminant vectors. *IEEE Transactions on Computers*, C-24(3):281–289, March 1975. 3

[5] Y. Fu, S. Yan, and T. S. Huang. Correlation metric for generalized feature extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2229–2235, 2008. 2, 5

[6] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001. 4

[7] X. He, D. Cai, S. Yan, and H. Zhang. Neighborhood preserving embedding. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1208–1213, 2005. 1

[8] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005. 1

[9] J. Kittler, M. Hatel, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, January 1998. 4

[10] J. Kittler, Y. Li, and J. Matas. On matching scores for lda-based face verification. In *Proceedings of the British Machine Vision Conference*, pages 42–51, 2000. 1

[11] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999. 4

[12] Y. Ma, S. Lao, E. Takikawa, and M. Kawade. Discriminant analysis in correlation similarity measure space. In *Proceedings of International Conference on Machine Learning*, pages 577–584, 2007. 2

[13] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(22):2323–2326, 2000. 1

[14] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003. 4

[15] J. Tenenbaum, V. Silva, and J. Langford. A global gemoetric framework for nonlinear dimensinality reduction. *Science*, 290(22):2319–2323, 2000. 1

[16] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 586–591, 1991. 1

[17] X. Wang and X. Tang. A unified framework for subspace face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1222–1228, 2004. 1

[18] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007. 1