2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops

Local Sparse Discriminant Analysis For Robust Face Recognition

Cuicui Kang, Shengcai Liao, Shiming Xiang, Chunhong Pan
Institute of Automation, Chinese Academy of Sciences
{cckang, scliao, smxiang, chpan}@nlpr.ia.ac.cn

Abstract

The Linear Discriminant Analysis (LDA) algorithm plays an important role in pattern recognition. A common practice is that LDA and many of its variants generally learn dense bases, which are not robust to local image distortions and partial occlusions. Recently, the LASSO penalty has been incorporated into LDA to learn sparse bases. However, since the learned sparse coefficients are globally distributed all over the basis image, the solution is still not robust to partial occlusions. In this paper, we propose a Local Sparse Discriminant Analysis (LoSDA) method, which aims at learning discriminant bases that consist of local object parts. In this way, it is more robust than dense or global basis based LDA algorithms for visual classification. The proposed model is formulated as a constrained least square regression problem with a group sparse regularization. Furthermore, we derive a weighted LoSDA (WLoSDA) approach to learn localized basis images, which also enables multi subspace learning and fusion. Finally, we develop an algorithm based on the Accelerated Proximal Gradient (APG) technique to solve the resulting weighted group sparse optimization problem. Experimental results on the FRGC v2.0 and the AR face databases show that the proposed LoSDA and WLoSDA algorithms both outperform the other state-of-the-art discriminant subspace learning algorithms under illumination variations and occlusions.

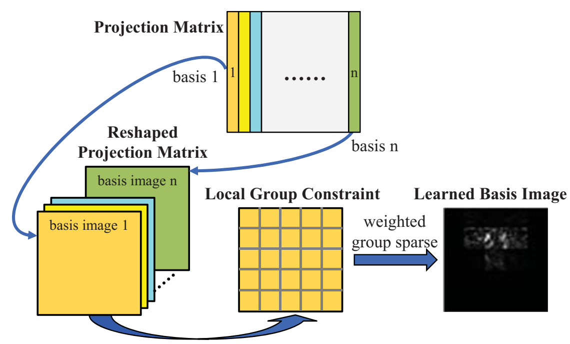

Figure 1. Illustration of the local group sparse regularization for the proposed LoSDA algorithm, which aims at learning local basis.

1. Introduction

Computer vision applications often deal with high dimensional data, which contains redundant information and requires a heavy computational cost to process. Therefore, feature extraction and dimensionality reduction techniques are widely applied in practice. Among them, the Principle Component Analysis (PCA) [8] and the Linear Discriminant Analysis (LDA) [8, 9, 10] are the most popular approaches. PCA aims at finding a subspace that contains the most information from the data, while LDA seeks a subspace that best separates multi classes. LDA has been widely applied for image classification problems including image retrieval [26] and face recognition [4], and demonstrates impressive performances.

The purpose of LDA is to learn a projection matrix, by which the original data can be transformed to a low dimensional subspace, where the mapped data achieves maximum between-class distance and minimum within-class distance simultaneously. However, the classical LDA and many of its variants [6, 7, 26, 32, 33] generally learn dense bases for the projection matrix, which are not robust to local image distortions and partial occlusions. Considering this, Cai *et al*. proposed a sparse LDA algorithm [5] for robust subspace learning, which utilizes the LASSO [28] penalty on the bases to promote a sparse solution. However, since the learned sparse coefficients are globally distributed across the basis image, the derived solution is still not robust to local image distortions and partial occlusions.

To learn a subspace that is robust to partial occlusions and local illumination variations, a good suggestion is to make the basis image only contain local object parts [14, 15, 16]. This is inspired from the field of Nonnegative Matrix Factorization (NMF) [15, 16], where the nonnegative constraint helps to derive local object parts. In this way, the learning algorithm is able to extract local features instead of holistic features for image representation, which is proved to be less affected by partial occlusions [14, 16].

In this paper, we propose a Local Sparse Discriminan-

978-0-7695-4990-3/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPRW.2013.125

t Analysis (LoSDA) approach for robust visual classification. The proposed method aims at learning discriminant bases that consist of local object parts, which enables local feature extraction for image representation (see Fig. 1). First, we divide each basis image of the projection matrix into several local blocks, and treat each block as a group of variables. Then, the proposed model is formulated as a constrained least square regression problem with a group sparse regularization. In this way, most of the groups will be shrunk to zeros, leaving few groups have none zero entries. Furthermore, we derive a weighted LoSDA (WLoSDA) approach to better control the coefficient distribution across groups, enforcing a solution of localized basis images. With different weighting strategies, WLoSDA also enables multi subspace learning and fusion. Finally, we develop an algorithm based on the Accelerated Proximal Gradient (APG) technique [21, 29] to solve the resulting weighted group sparse optimization problem.

Our experimental results on the FRGC v2.0 and the AR face databases show that the proposed LoSDA and WLoSDA algorithms both outperform the other state-of-the-art discriminant subspace learning algorithms. Especially, on face images with both synthesized and real partial occlusions, the proposed WLoSDA algorithm outperforms existing algorithms for more than 5% in recognition accuracy, which demonstrates the effectiveness of the proposed algorithm for robust visual classification.

The contribution of this work is summarized as follows.

- A novel discriminant subspace learning approach called LoSDA is proposed, which learns local object parts as bases, and hence enables local feature extraction for robust image representation.
- A weighted LoSDA (WLoSDA) method is proposed, which learns more localized basis than LoSDA, and with different weighting strategies, it enables multi-subspace learning and fusion.
- The proposed model is formulated as a least square regression problem regularized by a weighted group sparse constraint, and an algorithm based on the APG technique is developed to solve the resulting weighted group sparse optimization problem.

The rest of this paper is organized as follows. First, we briefly review the related work in Section 2. Then, we formulate the proposed LoSDA model and present its optimization in Section 3, as well as the WLoSDA. The experiments are shown in Section 4 with a useful discussion. Finally, we summarize the paper in Section 5.

2. Related Work

LDA is a supervised learning algorithm, which learns a discriminant subspace that minimizes the within-class distance and maximizes the between-class distance simultaneously. The dimensions of the derived subspace of the classical LDA can be very low; it is no more than $c-1$ dimensions, where c is the number of classes of the original data [8, 10]. However, one of the drawback of LDA is that it often trapped into the curse of high dimension, especially when there are limited number of training samples. It is known as the small sample size problem (SSS) [9]. Based on this, many extensions of the LDA algorithm have been proposed to overcome it, such as PCA+LDA [4, 26], null LDA [7], the directed LDA [33], and the Regularized LDA [32]. All of the algorithms were developed to deal with the SSS problem. The proposed method follows the regularization idea, and the proposed group sparse regularization makes the algorithm also robust to the SSS problem.

There are also other extensions of LDA, such as the Local LDA [24, 25], spectral regression based LDA (SRDA) [6], and the sparse LDA [5]. It should be noted that the local LDA and our proposed LoSDA are totaly different models. The local LDA in [25] deals with different local clusters of samples in the same class, which is done in the sample space. The study in [24] is also done in the sample space as well. *In contrast, the proposed LoSDA algorithm aims at learning local basis image, which adds spatial constraints on basis images.* The SRDA algorithm proposed by Cai *et al*. is based on the spectral clustering algorithm [6], which can be divided into two steps. On the first step, the SRDA learns the target subspace representation for each sample by the spectral technique based on a designed undirected graph. Second, the least square regression is applied to the original data and the subspace representation learned in the first step to calculate the transformation matrix. For the sparse LDA (SSRDA) in [5], it is the sparse version of the SRDA algorithm. The difference is that SSRDA applies the LASSO penalty in the second step of SRDA to compute the transformation matrix. While, our algorithm does not need to learn the transformation matrix and the subspace representation separately. The LoSDA is done directly in a regression form. Therefore, in our model, it just needs to solve the group sparse regularized least square regression problem.

With the development of LDA techniques, researchers have realized that there are some connections between the LDA and the least square regression problem [8, 10, 30, 34]. Duda has proved the equality between the LDA algorithm and the least square regression with a binary-class classification problem in [8]. Zhang and Riedel also proposed a discriminant learning analysis algorithm based on the ridge regression [34]. Recently, Ye [30] shows that by defining a special correspondence matrix based on the data's labels, the least square regression problem is equal to LDA under a mild condition that can be easily satisfied. The proposed algorithm is closely related with this development, and our model is directly built upon the least square regression formulation of LDA proposed in [30].

Face recognition under partial occlusion is an important problem, for which many algorithms have been developed [11, 13, 14, 16, 20, 22]. Among these, there are the SVM based algorithms [11, 13], the subspace based algorithms [14, 16, 22], and so on. Especially, the Local NMF (LNMF) algorithm suggests that learning local basis for face representation is effective against occlusion [16]. Our algorithm follows the subspace learning direction, which can learn a discriminant subspace with local object parts, and enable multi subspace learning and fusion.

3. Algorithm

In this section, we firstly give a brief review of the classical LDA algorithm. Then, based on the connection of the LDA and the least square regression problem [8, 10, 30], we propose the Local Sparse Discriminant Analysis (LoSDA) formulation along with the Weighted LoSDA. The optimization of the proposed objective function is finally built upon the accelerated proximal gradient method [21, 29].

3.1. Classical Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a basic algorithm in the pattern recognition field. The binary-class LDA is to learn an effective direction in which the two class data can be separated the utmost [8]. For the multi-class problem with c classes, the LDA projects the original data into $c-1$ dimensional subspace to separate them. In the derived subspace, the mapped data achieves the minimum within-class distance, and the maximum between-class distance simultaneously.

Suppose the original data has n samples from c classes, namely $\{\mathbf{a}_i, y_i\}_{i=1}^n$, where $\mathbf{a}_i \in \mathbb{R}^{d\times 1}$ is the feature vector, y_i is the class label, and each class j has n_j samples. Then the data matrix $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_c]$, where $\mathbf{A}_j$ consists of samples belonging to the j-th class. The objective function of the LDA is [8],

$$\max_{\mathbf{X}} J(\mathbf{X}) = \frac{\text{tr}(\mathbf{X}^T\mathbf{S}_b\mathbf{X})}{\text{tr}(\mathbf{X}^T\mathbf{S}_w\mathbf{X})}, \tag{1}$$

where $\mathbf{X}$ is the projection matrix that needs to be learned, $\mathbf{S}_b$ and $\mathbf{S}_w$ are called the between-class scatter matrix and the within-class scatter matrix, respectively. By denoting the centroid $\mathbf{m}_j = \frac{1}{n_j}\sum_{y_i=j}\mathbf{a}_i$ for the j-th class and the global centroid $\mathbf{m} = \frac{1}{n}\sum_{j=1}^c n_j\mathbf{m}_j$, the between-class scatter matrix $\mathbf{S}_b$, and the within-class scatter matrix $\mathbf{S}_w$ can be written as [8]

$$\mathbf{S}_w = \sum\nolimits_{j=1}^c \sum\nolimits_{y_i=j}(\mathbf{a}_i - \mathbf{m}_j)(\mathbf{a}_i - \mathbf{m}_j)^T,$$

$$\mathbf{S}_b = \sum\nolimits_{j=1}^c n_j(\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^T,$$

respectively.

The Linear Discriminant Analysis has a strong connection with the least square regression problem, which has been shown in [8] with a binary-class classification problem. Recent studies on LDA have proved that, the multi-class LDA problem is equivalent to the multivariate linear regression with a predefined correspondence matrix under a mild condition [30]. The equivalent formulation is,

$$\min_{\mathbf{X}}\{\frac{1}{2}\|\mathbf{A}^T\mathbf{X} - \mathbf{H}\|_F^2\}, \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenious-norm, and the element of the correspondence matrix $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_c]$ is constructed as

$$h_{i,j} = \begin{cases} \sqrt{\frac{n}{n_j}} - \sqrt{\frac{n_j}{n}}, & if \;\; y_i = j, \\ -\sqrt{\frac{n_j}{n}}, & otherwise. \end{cases} \tag{3}$$

The equality holds under a mild condition that rank$(\mathbf{S}_t) =$ rank$(\mathbf{S}_b)$ + rank$(\mathbf{S}_w)$, where $\mathbf{S}_t$ is the total scatter matrix defined as

$$\mathbf{S}_t = \sum\nolimits_{i=1}^n(\mathbf{a}_i - \mathbf{m})(\mathbf{a}_i - \mathbf{m})^T = \mathbf{S}_w + \mathbf{S}_b.$$

This mild condition can be easily satisfied in many practical applications [31].

3.2. Local Sparse Discriminant Analysis

The motivation of the proposed LoSDA algorithm is to learn discriminant subspaces consisting of local basis images, which is robust to local image distortion and partial occlusions. The basic idea is, we can divide each basis image of the projection matrix into several local blocks, and treat each block as a group of variables. Then, the proposed model is formulated as a constrained least square regression problem with a group sparse regularization. In this way, most of the groups will be shrunk to zeros, leaving few groups have none zero entries.

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_c]$ be the projection matrix as in Eq. (1), where each $\mathbf{x}_j$ is a basis or component to learn. We define g groups for each basis $\mathbf{x}_j$ (see Fig. 2 for an illustration of group definition), and denote the k-th group of variables in the j-th basis as $\mathbf{x}_{j,G_k}$. Then, using the notations defined in the above subsection, the proposed LoSDA model can be formulated as

$$\min_{\mathbf{X}}\{\frac{1}{2}\|\mathbf{A}^T\mathbf{X} - \mathbf{H}\|_F^2 + \lambda\sum_{j=1}^c\sum_{k=1}^g w_k\|\mathbf{x}_{j,G_k}\|_2\}, \tag{4}$$

where $\mathbf{H}$ is constructed as in Eq.(3), λ is the regularization parameter, and w_k is the weight of the k-th group for each basis. Here we use equal weights for the LoSDA algorithm.

It can be observed that the second term of Eq.(4) is a local group sparse constraint. By rewriting Eq.(4) as

$$\min_{\mathbf{X}} \frac{1}{2}\sum_{j=1}^{c} \|\mathbf{A}^T\mathbf{x}_j - \mathbf{h}_j\|_2^2 + \lambda \sum_{j=1}^{c}\sum_{k=1}^{g} w_k \|\mathbf{x}_{j,G_k}\|_2, \quad (5)$$

it can be easily discovered that the original problem can be decomposed into c sub-problems, which have the same simplified formulation as follows

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}^T\mathbf{x} - \mathbf{h}\|_2^2 + \lambda \sum\nolimits_{k=1}^{g} w_k \|\mathbf{x}_{G_k}\|_2, \quad (6)$$

where $\mathbf{x}$ and $\mathbf{h}$ are column vectors from the matrix $\mathbf{X}$ and $\mathbf{H}$, respectively, and $\mathbf{x}_{G_k}$ is the k-th group of variables in $\mathbf{x}$.

In fact, Eq.(6) is a kind of sparse regression problem, which is a convex but non-smooth optimization problem. In the following, we adopt the accelerated proximal gradient (APG) algorithm [3, 12, 21, 29] to solve Eq.(6). The proximal gradient method is a first order gradient descent method, who has received popular attentions in recent years because of its fast convergence rate $O(1/t^2)$ when dealing with the non-smooth convex problem, where t is the iteration number [3, 12, 21, 29]. By decomposing the objective function Eq.(6) into a smooth term and a non-smooth term, we have

$$l(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}^T\mathbf{x} - \mathbf{h}\|_2^2, \quad (7)$$

$$\min_{\mathbf{x}} f(\mathbf{x}) = l(\mathbf{x}) + \lambda \sum\nolimits_{k=1}^{g} w_k \|\mathbf{x}_{G_k}\|_2. \quad (8)$$

APG finds the proximate solution $\mathbf{x}$ based on another sequence of search points $\{\mathbf{v}_i\}$, where $\mathbf{v}_i$ is a linear combination of $\mathbf{x}_i$ and $\mathbf{x}_{i-1}$ (see Eq.16). First, we construct a proximal operator of the Eq.(8) to find the solution $\mathbf{x}$,

$$\begin{aligned}\min_{\mathbf{x}} g(\mathbf{x}) =& l(\mathbf{v}_t) + \langle l'(\mathbf{v}_t), \mathbf{x} - \mathbf{v}_t\rangle + \frac{1}{2\eta_t}\|\mathbf{x} - \mathbf{v}_t\|_2^2 \\ &+ \lambda \sum\nolimits_{k=1}^{g} w_k \|\mathbf{x}_{G_k}\|_2,\end{aligned} \quad (9)$$

where

$$l'(\mathbf{v}_t) = \mathbf{A}(\mathbf{A}^T\mathbf{v}_t - \mathbf{h}). \quad (10)$$

and η_t is the update step. By replacing the constant term $l(\mathbf{v}_t)$ by another constant term $\eta_t\|l'(\mathbf{v}_t)\|_2^2/2$, with respect to $\mathbf{x}$, the above proximal operator is equal to

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x} - \mathbf{u}_t\|_2^2 + \lambda\eta_t \sum\nolimits_{k=1}^{g} w_k \|\mathbf{x}_{G_k}\|_2 \quad (11)$$

where

$$\mathbf{u}_t = \mathbf{v}_t - \eta_t l'(\mathbf{v}_t), \quad (12)$$

As for Eq.(11), we can further expand it as the sum of g local groups,

$$\min_{\mathbf{x}} \frac{1}{2}\sum_{k=1}^{g} \|\mathbf{x}_{G_k} - \mathbf{u}_{t,G_k}\|_2^2 + \lambda\eta_t \sum_{k=1}^{g} w_k \|\mathbf{x}_{G_k}\|_2. \quad (13)$$

Then, optimizing Eq.(11) is actually equivalent to optimizing k sub-problems, where the k-th sub-problem for the local group G_k is defined as

$$\min_{\mathbf{x}_{G_k}} \frac{1}{2}\|\mathbf{x}_{G_k} - \mathbf{u}_{t,G_k}\|_2^2 + \lambda\eta_t w_k \|\mathbf{x}_{G_k}\|_2. \quad (14)$$

Finally, by applying the soft-thresholding technique [18], the solution of Eq.(14) is obtained by

$$\mathbf{x}_{G_k} = \begin{cases} (1 - \frac{\lambda\eta_t w_k}{\|\mathbf{u}_{t,G_k}\|_2})\mathbf{u}_{t,G_k}, & if ||\mathbf{u}_{t,G_k}||_2 > \lambda\eta_t w_k \\ \mathbf{0}, & otherwise. \end{cases} \quad (15)$$

The APG algorithm accelerates the proximal gradient descent by updating $\mathbf{v}_{t+1}$ as,

$$\mathbf{v}_{t+1} = \mathbf{x}_{t+1} + \frac{\alpha_{t+1}(1-\alpha_t)}{\alpha_t}(\mathbf{x}_{t+1} - \mathbf{x}_t), \quad (16)$$

where $\alpha_t = \frac{2}{t+2}$ [29].

As a result, the accelerated proximal gradient algorithm updates $\mathbf{x}_t$ and $\mathbf{v}_t$ alternatively to find the solution. The summary of the derived algorithm to solve our objective function Eq.(6) is shown in Algorithm 1.

Algorithm 1: APG solution for Eq.(6).

Input: $\mathbf{A} \in \mathbb{R}^{d\times n}$, $\mathbf{h} \in \mathbb{R}^{n\times 1}$, group index G, and group weight $\mathbf{w}$.
Output: $\mathbf{x}$

Initialization $\mathbf{v}_t = \mathbf{0}, t = 0$;
repeat
  Compute $\mathbf{u}_t$ as in Eq.(12);
  for *each group* k **do**
    if $\|\mathbf{u}_{t,G_k}\|_2 \le \lambda\eta_t w_k$ **then**
      $\mathbf{x}_{t+1,G_k} = \mathbf{0}$;
    else
      $\mathbf{x}_{t+1,G_k} = (1 - \frac{\lambda\eta_t w_k}{\|\mathbf{u}_{t,G_k}\|_2})\mathbf{u}_{t,G_k}$;
    end
  end
  Compute $\mathbf{v}_{t+1}$ as in Eq.(16).
until *convergence*;

3.3. Weighted Local Sparse Discriminant Analysis

In this section, we derive a weighted LoSDA (WLoSDA) algorithm which considers different weights in Eq. (4). In our formulation, it can be observed that, we can learn different subspaces according to different weighting strategies. Our weighting strategy considers to pursue localized basis images. To this end, we can design a weighting strategy that selects a certain group as the center group, and assign weights to other groups according to the spatial distance. An illustration of this weighting strategy is shown in Fig. 2.

Generally, we put more weights on groups that are farther away from the center group. As a result, the optimization of the algorithm is guided to penalize less on the center group and its neighbors, and more on other far away groups. Consequently, the algorithm is expected to get a solution with localized basis images.

Furthermore, with different selection of the center group, we can learn a different subspace. Therefore, the proposed WLoSDA algorithm naturally enables multi subspace learning, which is different from existing subspace learning algorithms. In order to make the learned multi subspaces complementary to each other, we select the center groups that emphasizing on different object parts (see Fig. 2 for an example). We also develop a multi subspace fusion method for the distance metric. Suppose there are m weighting strategies, *e.g.* $\{\mathbf{w}^i\}_{i=1}^m$, then, we can learn m projection matrices accordingly. Let $\mathbf{p}^k$ and $\mathbf{g}^k$, $k = 1, 2, \cdots, m$, be the projected vectors of two original samples to the m subspaces, respectively, then, we can compute the distance of the two samples by $dist(\mathbf{p}, \mathbf{g}) = \min_{k=1}^m dist(\mathbf{p}^k, \mathbf{g}^k)$, which adopts the min fusion rule.

4. Experiments

To evaluate the performance of the proposed LoSDA and WLoSDA algorithms, we apply them to the face recognition problem, with experiments conducted on the FRGC v2.0 [23] and the AR [19] face databases. We also compared several state-of-the-art discriminant analysis algorithms, including the classical LDA [8], regularized LDA (RLDA) [32], Spectral Regression Discriminant Analysis (SRDA) [6][1], the Sparse regularized version of the SRDA (SSRDA) [5], and the famous local object part based learning algorithms, the NMF [15] and LNMF [16]. For both the FRGC and the AR databases, we split the data into a training set and a testing set, where there was no overlapping of classes (AR) or a partial overlapping of classes but no overlapping of images (FRGC), between the two sets. For the testing set, we selected *one sample per class* as enrolled in the gallery set, and used the rest for the probe set. The testing task is to recognize the identity of each face image in the probe set. We projected each face image in the testing set in the learned subspace, and calculated the Cosine metric between two projected features as the similarity measure. Finally, we applied the K Nearest Neighbor (KNN) classifier with $K = 1$ for face identification, and used the rank-1 recognition accuracy for performance report.

Some of the compared algorithms in the experiments, such as the RLDA, SSRDA and the proposed LoSDA, have a regularization parameter. The optimal value of the regularization parameter for each algorithm was cross validated on the training set in the range of $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.2, \ldots, 1\}$. For this task, we randomly split the training set into two subsets, one for training and the other one for testing, and repeated this for 5 times for cross validation of the parameters. The average rank-1 recognition accuracy on the testing subset was calculated with each parameter, and the optimal parameter value was determined accordingly for each algorithm. As for the NMF and the LNMF algorithms, the only parameter is the dimension of the learned subspace. Considering they are unsupervised algorithms, we set the learned dimension as $5 \times c$, which is 5 times as the learned dimension of the proposed LoSDA algorithm.

[1]The source code is available on the author's webpage [1].

In all experiments, we used the gray scale pixel intensity values as features. All face images were cropped and resized into 35×35 pixels, and then each face image was pulled as a long feature vector. For the proposed LoSDA and WLoSDA algorithms, we simply divided each basis image (in columns of $\mathbf{X}$ in Eq. 4) into 25 (5×5) equivalent blocks, so that each block contained 7×7 pixels as a group. Fig. 2 shows how we defined the groups on a basis image.

Figure 2. The 5×5 groups on a basis image (left), and two illustration examples where the defined groups are attached on face images. The example images are from the FRGC (middle) and the AR (right) databases.

For the WLoSDA algorithm, we applied five weighting strategies, and the weights for each group were defined as follows. For each weighting strategy, we chose a block as the centroid, and then set the weights of other blocks as the square of the Euclidian distance (blockwise) to the centroid. In the experiments, we chose five blocks as centroids as shown in Fig. 2 (number 7, 9, 13, 17, and 19). In this way, for each weighting strategy, the farther a group is to the centroid, the more penalty is placed to the group. Therefore, the optimization of Eq. (4) will promote a solution of basis images localized at the selected centroid.

4.1. Experiments On FRGC

The Face Recognition Grand Challenge (FRGC) v2.0 database [23] is a large and challenging face database which contains about 50,000 frontal face images. There are 2D face images taken in both controlled and uncontrolled environments, as well as 3D face images, which were not used in our experiments. The controlled images were taken indoor with two lighting conditions and two expressions (smiling and neutral). The uncontrolled images were taken outdoor, which also contains two expressions, smiling and neutral. The FRGC v2.0 database contains a training set and a testing set for the 2D face images. The training set contains 12,776 face images of 222 individuals, in which

6,360 are controlled images and 6,416 are uncontrolled images. The testing set consists of 466 individuals, including 16,028 controlled images captured indoor and 8,014 uncontrolled images captured outdoor. The testing set contains some individuals present in the training set, but there is no overlapping of images between the two sets. The top row of Fig. 4 shows some cropped face images from the database.

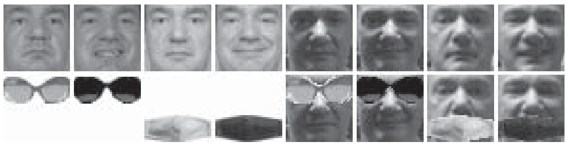

Figure 3. Top row: example indoor face images (first 4) and outdoor face images (last 4) cropped from the FRGC database. Bottom row: some sunglasses and masks used in the experiments, and some example face images with synthesized occlusions.

In the experiments, the face images were cropped and resized to 35×35 pixels according to the provided eye coordinates, followed by an illumination preprocessing proposed in [27]. We selected the first image of each individual in the indoor set as the gallery image, and the remaining images constituted the probe set, including 15,562 indoor images and 8,014 outdoor images. Then we randomly separated the indoor images of each person as two parts, about half images as the indoor probe images, and the other half for probe images with synthesized occlusions. Therefore, the probe images were separated into three subsets, indoor images, outdoor images, and the occlusion images, with a total number of 8,014, 8,014, and 7,548 images, respectively. For the occlusion subset, some sunglasses (a total of 4 types) or masks (a total 4 types) were randomly selected and placed on the cropped images. Some face images with synthesized occlusions are shown in Fig.3. Considering the computational cost, we just randomly selected 10 indoor face images and 10 outdoor face images per class from the training set for the cross validation of parameter values. Then, 5 indoor images and 5 outdoor images per subject were randomly selected to learn the final projection matrix. Table 1 displays the recognition accuracy of the compared algorithms on the indoor subset, outdoor subset, occlusion subset, and the whole testing set, respectively.

From Table 1, it can be observed that, except WLoSDA, the proposed LoSDA algorithm outperforms all other existing algorithms, with an improvement of 3% in average over the best performer, SSRDA, among the existing algorithms. This shows that, thanks to the local group sparse regularization, the proposed algorithm learns a more robust subspace than the existing state-of-the-art algorithms, against both illumination variations and partial occlusions. Besides, WLoSDA achieves a further improvement over LoSDA, especially on the occlusion subset (about 6% improvement), showing that more localized basis is more robust for face recognition under partial occlusion. It is obviously that the NMF and LNMF fail to learn an highly discriminant subspace. The most important reason is that both of them are unsupervised algorithms, while the others are supervised. It should also be noted that the NMF and LNMF solve nonconvex problems, thus they are not guaranteed to find the right solution, and may often be caught in a local minimum.

Table 1. Face recognition performance (%) on the FRGC database.

Method	Indoor	Outdoor	Occlu.	All
LDA	60.64	19.78	4.37	28.74
SRDA	56.66	14.19	3.97	25.36
RLDA	77.79	39.18	17.17	45.26
SSRDA	80.40	39.62	16.92	46.21
NMF	37.62	4.97	4.90	16.05
LNMF	36.06	1.30	4.66	14.19
LoSDA	83.45	**44.91**	18.52	49.56
WLoSDA	**84.35**	44.71	**24.17**	**51.61**

4.2. Experiments On AR

The AR database is a color facial database which contains more than 3,000 face images corresponding to 135 individuals (76 men and 59 women). These images are captured in two sessions, separated by two weeks (14 days). In each session, frontal face images under four different scenarios have been taken for each individual. Particularly, there are four face images for different facial expressions under normal light, three images under different illumination conditions with normal expression, three images under sunglasses occlusion and another three images under scarf occlusion. Images under the same conditions have been taken in both sessions so that there are 26 images per subject. There are also some people who just participated in one session. Fig.4 shows some examples from the AR database.

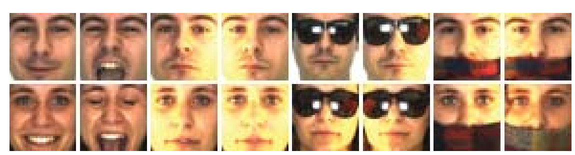

Figure 4. Example images cropped from the AR database, under expressions, illumination variations, occlusions with sunglasses and scarf.

In the experiments, we randomly separated the database into a training set, which contains 60 individuals, and a testing set consisting of the remaining subjects (75 individuals). The recognition accuracy of the compared algorithms on the testing set are shown in Table 2. We summarize results for the overall recognition accuracy, as well as on the facial expression subset (Expr.), illumination subset (Illum.), sunglasses subset (Glass), and the scarf subset. From the overall recognition accuracy (the last column in Table 2), we can

Table 2. Face recognition performance (%) on the AR database.

Method	Expr.	Illum.	Glass	Scarf	All
LDA	37.58	49.38	7.90	21.98	29.50
SRDA	35.42	46.42	5.68	19.26	27.00
RLDA	77.54	89.88	37.78	55.31	65.55
SSRDA	**79.27**	91.36	36.79	57.53	66.69
NMF	62.20	15.80	8.64	1.98	23.54
LNMF	62.42	13.83	7.41	2.22	22.88
LoSDA	78.83	**92.10**	37.04	62.72	68.06
WLoSDA	76.89	91.60	**41.23**	**70.62**	**70.32**

see that the proposed WLoSDA algorithm achieved the best performance, followed by LoSDA. It can be observed that, except WLoSDA, LoSDA performs the best in overall and in all subsets except the facial expression subset. Especially, LoSDA outperforms existing algorithms by more than 5% recognition rate on the scarf occluded images. This shows that the proposed method is more effective against occlusion, which benefits from the local group sparse regularization. Furthermore, the WLoSDA algorithm outperforms LoSDA by about 8% on the scarf subset and about 4% on the sunglasses subset. The notable improvement is because WLoSDA learns various localized components and boosts the performance by fusing information from different local parts on face images.

4.3. Discussion

The proposed LoSDA and the WLoSDA algorithms have shown their excellent performance and robustness on discriminant subspace learning for the face recognition problem. Another important property of an algorithm that we care about is the parameter sensitivity. With varying parameters, we expect the algorithm to be stable, making parameter selection an easy task. Considering this, we made an analysis on the regularization parameter λ in the proposed algorithms. The experiments were done on the training set of the AR database that we previously used, under the cross validation manner. Parameters values in $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.2, \ldots, 1\}$ were evaluated for λ. Fig. 5 shows the face recognition accuracy with respect to the value of λ, averaged for five trials on the training set. It can be observed that the performances of both the LoSDA and WLoSDA algorithms are improved when the λ increases up to 0.3. Then, the algorithm of WLoSDA becomes stable, while the performance of LoSDA drops. A good range of the parameter value λ for WLoSDA is [0.3, 1.0], while that for LoSDA is [0.1, 0.4].

In this paper, we applied the proposed algorithms to the face recognition problem, but they can also be applied to other visual classification applications that needs part-based representation rather than holistic representation to overcome the occlusion and local illumination variation challenges. In practical use, it should be noted that, though WLoSDA performs the best in our experiments, which is especially robust against occlusion, the multi subspace learning and fusion method generally involves more computation. Therefore, a balance between accuracy and efficiency should be considered for the adoption between the LoSDA and the WLoSDA algorithms.

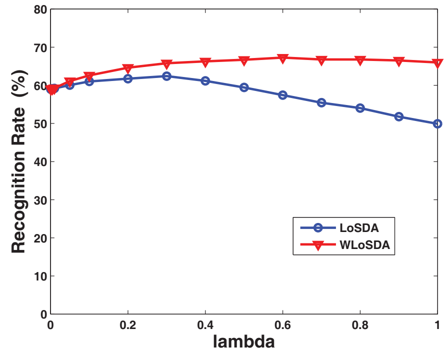

Figure 5. Recognition accuracy with respect to the value of parameter λ on the AR database.

Besides, we just utilized the raw pixel values as our features in the experiments, but the proposed algorithms are not limited to pixels. It can be extended to more complicated local descriptors to achieve better performance, such as the LBP [2] and Gabor features [17]. Particularly, the local histogram based features (e.g. LBP) with local regions can be conveniently treated as groups for our algorithm. Furthermore, the proposed algorithm can make use of overlapping groups to exploit different local structures for object image. We will consider this in our future study.

5. Conclusion

In the paper, we have proposed a novel discriminant analysis algorithm called LoSDA, which aims at learning discriminant bases that consist of local object parts. Furthermore, we derive a weighted LoSDA (WLoSDA) approach to learn localized basis images, which also enables multi subspace learning and fusion. The advantage of the proposed approach is that, it is more robust against partial occlusions and local illumination variations than traditional dense or global basis based LDA algorithms for visual classification. Experimental results on the FRGC v2.0 and the AR face databases show that the proposed LoSDA and WLoSDA algorithms both outperform the other state-of-the-art discriminant subspace learning algorithms under illumination variations and occlusions.

6. Acknowledgement

This work was supported in part by the National Basic Research Program of China under Grant 2012CB316304, and the National Natural Science Foundation of China under Grants 61272331 and 61175025.

References

[1] http://www.zjucadcg.cn/dengcai/.

[2] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009.

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

[5] D. Cai, X. He, and J. Han. Spectral regression: A unified approach for sparse subspace learning. In *International Conference on Data Mining*, 2007.

[6] D. Cai, X. He, and J. Han. Spectral regression for efficient regularized subspace learning. In *International Conference on Computer Vision*, 2007.

[7] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 33(10):1713–1726, 2000.

[8] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley-Sons, 2001.

[9] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.

[10] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: Data mining, inference, and prediction*. Springer, 2001.

[11] K. Hotta. Robust face recognition under partial occlusion based on support vector machine with local gaussian summation kernel. *Image and Vision Computing*, 26(11):1490 – 1498, 2008.

[12] S. Ji and J. Ye. An accelerated gradient method for trace norm minimization. In *International Conference on Machine Learning*, pages 457–464, 2009.

[13] H. Jia and A. Martinez. Support vector machines in face recognition with occlusions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 136–141, 2009.

[14] J. Kim, J. Choi, J. Yi, and M. Turk. Effective representation using ICA for face recognition robust to local distortion and partial occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1977–1981, 2005.

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

[16] S. Z. Li, X. Hou, H. Zhang, and Q. Cheng. Learning spatially localized, parts-based representation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 207–212, 2001.

[17] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 11(4):467–476, 2002.

[18] J. Liu, S. Ji, and J. Ye. Multi-task feature learning via efficient $\ell_{2,1}$-norm minimization. In *Uncertainty in Artificial Intelligence*, pages 339–348, 2009.

[19] A. Martinez and R. Benavente. The AR face database. Technical Report 24, Computer Vision Center, June 1998.

[20] A. M. Martinez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):748–763, 2002.

[21] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, 2003.

[22] H. J. Oh, K. M. Lee, and S. U. Lee. Occlusion invariant face recognition using selective local non-negative matrix factorization basis images. *Image and Vision Computing*, 26(11):1515–1523, 2008.

[23] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.

[24] X. Shu, Y. Gao, and H. Lu. Efficient linear discriminant analysis with locality preserving for face recognition. *Pattern Recognition*, 45(5):1892–1898, 2012.

[25] M. Sugiyama. Local fisher discriminant analysis for supervised dimensionality reduction. In *International Conference on Machine Learning*, volume 148, pages 905–912, 2006.

[26] D. L. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):831–836, 1996.

[27] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2007.

[28] R. Tibshirani. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.

[29] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. *submitted to SIAM Journal on Optimization*, 2008.

[30] J. Ye. Least squares linear discriminant analysis. *The Twenty-Fourth International Conference on Machine Learning*, pages 1087–1093, 2007.

[31] J. Ye and T. Xiong. Computational and theoretical analysis of null space and orthogonal linear discriminant analysis. *Journal of Machine Learning Research*, 7:1183–1204, 2006.

[32] J. Ye, T. Xiong, Q. Li, R. Janardan, J. Bi, V. Cherkassky, and C. Kambhamettu. Efficient model selection for regularized linear discriminant analysis. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, pages 532–539, 2006.

[33] H. Yu and J. Yang. A direct LDA algorithm for high dimensional data with application to face recognition. *Pattern Recognition*, 34(10):2067–2070, 2001.

[34] P. Zhang and N. Riedel. Discriminant analysis: A unified approach. In *The Fifth IEEE International Conference on Data Mining*, pages 514–521, 2005.