ICIC Express Letters
Part B: Applications
Volume **9**, Number **6**, June **2018**
ICIC International ©2018 ISSN 2185-2766
pp. **591–598**

A NOVEL LOG-BASED WEIGHTED 2DLDA ALGORITHM FOR FACE RECOGNITION

Yongjie Chu and Lindu Zhao*

Institute of Systems Engineering
Southeast University
No. 2, Sipailou, Nanjing 210096, P. R. China
yongjiechu@yeah.net; *Corresponding author: ldzhao@seu.edu.cn

Received December 2017; accepted March 2018

Abstract. *It is observed that 2DLDA may confuse the contribution of the projection vectors when used for feature extraction. To overcome the drawback of 2DLDA, we present a novel image feature extraction algorithm called log-based weighted two dimensional linear discriminant analysis (LW-2DLDA). It is an improvement of direct 2DLDA by introducing a logarithmic function to weigh the projection vectors of both within-class and between-class subspaces. The proposed approach extracts discriminative features by subsequently mapping the images into the weighted intra-class subspace and the weighted extra-class subspace. As LW-2DLDA loads different vectors with proper weights, it is capable of avoiding overemphasizing or over-underrating any projection vectors, so that more discriminative features can be obtained in the projection subspace. Moreover, LW-2DLDA is image matrix-based rather than vector-based, and it deals well with small sample size problem. The new algorithm is tested on FERET and LFW face databases, and the experimental results demonstrate its effectiveness.*
Keywords: Two-dimensional linear discriminant analysis (2DLDA), Feature extraction, Log-based weights, Face recognition

1. **Introduction.** Linear discriminant analysis (LDA), mainly as a technique for feature extraction, has been successfully applied to face recognition field in the past two decades. The optimal linear transformation of LDA achieves maximum class separability by minimizing the within-class distance and maximizing the between-class distance simultaneously. Though LDA is a classical method, numerous variants of it are still successively proposed recently. These variants are designed to enhance the quality of feature representation, such as local LDA [1], or to address other machine learning tasks, such as semi-supervised LDA [2] and multi-label LDA [3]. To boost the discriminative ability of the subspace of LDA, regularized LDA methods [4-7] and weighted LDA approaches [8-11] have been designed. Some regularized LDA methods propose to impose constraints on the optimal transformation of LDA to avoid overfitting [4,5], while others impose local structure information of data on the scatter matrices to achieve robustness [6,7]. The motivation of various weighted LDA is to alleviate the imbalance influence of the between-class distances by redefining the between-class or within-class matrix, both of which need to introduce weighting functions.

All of the above LDA is one dimensional LDA (1DLDA), which means they must convert the 2D image matrices into 1D high dimensional vectors. Since 1DLDA operates on the 1D high dimensional vectors, it usually encounters singularity of the within-class scatter matrix, which is also called small sample size (S3) problem. Although 1DLDA can be modified to solve the problem, such as sparse uncorrelated LDA [12], it is laborious. An efficient solution to the S3 problem is two dimensional LDA (2DLDA), which can perform LDA directly on image matrices, overleaping the process of turning image matrices into vectors. 2DLDA was first proposed in [13]. Since then, dozens of 2DLDA based algorithms have been designed. [14] presents a 2DLDA method for face recognition, which proves to be more efficient than 1DLDA. Considering the local information of data, [15] designs a weighted 2DLDA algorithm to map the images such that pairwise between-class distances can be well balanced. Similarly to regularized 1DLDA, [16] proposes robust L1-norm 2DLDA. To extract sufficient features from images, [17,18] propose bidirectional 2DLDA, both of which learn two transformation matrices for images. Besides, [18] integrates multiple kernels into 2DLDA to handle non-linear mappings.

All of the LDA methods, including 1DLDA and 2DLDA, work well when the number of classes is small, but degenerate obviously when that is large. One possible reason is that in the process of generating the discriminant matrices or projection subspaces, and some important eigenvectors are overemphasized or some less important eigenvectors are over-discounted. To address the problem, we propose a novel log-based weighted 2DLDA in this paper to reweigh the coefficients of projection vectors by using logarithmic function.

Our major contributions are summarized as follows. 1) We introduce a logarithmic function to endow various weights to eigenvectors according to their corresponding eigenvalues. The new weight function helps properly emphasize the important projection vectors of within-class and between-class subspaces, while avoids over-emphasizing or over-underrating effect. 2) We develop a new algorithm to solve the LW-2DLDA based on singular value decomposition and simultaneous diagonalization of the scatter matrices. The new algorithm is much more efficient, since it avoids the expensive computation of inverse matrix. 3) Extensive experiments on two challenging face datasets are conducted to show the effectiveness of our method.

The rest of the paper is organized as follows. Section 2 reviews 2DLDA and presents the proposed approach, and some related discussions are all covered. In Section 3, the proposed approach is applied for face recognition and compared with some other related methods. Finally, a brief conclusion is given in Section 4.

2. The Proposed Approach.

In this section, we first review the framework of 2DLDA, and explicitly demonstrate the limitations of existing 2DLDA. Then we describe the proposed LW-2DLDA algorithm and its advantages.

2.1. Overview of 2DLDA and its limitations.

Suppose there are C known individuals in the training set. M is the total number of training samples and M_i is the number of training samples of the ith individual. The jth training sample of the ith individual is denoted by an $m \times n$ matrix $A_j^{(i)}$. The mean image of the ith individual is denoted by $\overline{A}^{(i)}$ and the mean image of all training samples is denoted by $\overline{A}$.

Based on the given training image samples, the image between-class scatter matrix and image within-class scatter matrix are defined as

$$G_b = \frac{1}{M} \sum_{i=1}^{C} M_i \left(\overline{A}_i - \overline{A}\right)^T \left(\overline{A}_i - \overline{A}\right) \tag{1}$$

$$G_w = \frac{1}{M} \sum_{i=1}^{C} \sum_{j=1}^{M_i} \left(A_j^{(i)} - \overline{A}^{(i)}\right)^T \left(A_j^{(i)} - \overline{A}^{(i)}\right) \tag{2}$$

By the definition, it is easy to verify that both G_b and G_w are $n \times n$ non-negative definite matrices. It should also be mentioned that G_w is invertible when there are at least two training samples of each individual. The optimal projection matrix W of 2DLDA can be obtained by solving the following optimization problem:

$$\begin{cases} W = [w_1, \ldots, w_d] = \arg\max \dfrac{tr\left(W^T G_b W\right)}{tr\left(W^T G_w W\right)} \\ w_k^T w_l = 0,\ k \neq l,\ k, l = 1, \ldots, d \end{cases} \tag{3}$$

To solve the above problem efficiently, [9] designed a simple algorithm named direct LDA, which is readily manipulated on image matrices to form direct 2DLDA. The projection matrix W computed by direct 2DLDA can be expressed as the following:

$$\begin{aligned} W &= YD_b^{-1/2}VD_w^{-1/2} \\ &= \underbrace{[u_1, \ldots, u_d]\begin{bmatrix} \frac{1}{\sqrt{\mu_1}} & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\mu_d}} \end{bmatrix}}_{\text{weighted extra-class subspace}} * \underbrace{[v_1, \ldots, v_d]\begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\lambda_d}} \end{bmatrix}}_{\text{weighted intra-class subspace}} \end{aligned} \quad (4)$$

In Equation (4), $Y = [u_1, \ldots, u_d]$ is a matrix consisting of eigenvectors of G_b corresponding to the first largest d eigenvalues $\mu_1, \ldots, \mu_d$; $V = [v_1, \ldots, v_d]$ is a matrix constructed by the eigenvectors of $\left(YD_b^{-1/2}\right)^T G_w \left(YD_b^{-1/2}\right)$ corresponding to the first largest d eigenvalues $\lambda_1, \ldots, \lambda_d$. u_k is an n-dimensional vector and v_k is a d-dimensional vector for $k = 1, \ldots, d$.

For the right side of Equation (4), the two diagonal matrices $D_b^{-1/2}$ and $D_w^{-1/2}$ can be treated as weighting matrices on corresponding eigenvectors if we associate the first two terms and latter two terms respectively. We call them weighted extra-class subspace and weighted intra-class subspace. The D-2DLDA algorithm seems to work well; however, the weighting process may confuse the contribution of the eigenvectors, leading to either the loss or the redundancy of information in representation. Specifically, for the between-class scatter matrix, the projection vector u_k is more important than the projection vector u_t if $\mu_k > \mu_t$; while for the within-class scatter matrix, the projection vector v_k is more important than the projection vector v_t when $\lambda_k > \lambda_t$, where $k, t = 1, \ldots, d$. In order to extract more discriminative information, it is beneficial to emphasize the important projection vectors and underrate the less important ones. For the extra-class subspace in Equation (4), however, $1/\sqrt{\mu_k} < 1/\sqrt{\mu_t}$ implies that the weight coefficients deemphasize the important projection vectors and highlights the unimportant ones of between-class scatter matrix, which is contrary to our expectation. For the intra-class subspace, the weight coefficients do highlight the important projection vectors and underrate the less important ones since $1/\sqrt{\lambda_k} < 1/\sqrt{\lambda_t}$. However, in some cases, when an eigenvalue λ_k is either very small or large, the corresponding weighting coefficient $1/\sqrt{\lambda_k}$ might be either too large or small, resulting in over-emphasizing projection vectors with tiny eigenvalues or over-depressing projection vectors with huge eigenvalues. As a consequence, features extracted by D-2DLDA may involve redundant information or loss helpful information.

2.2. **LW-2DLDA.** To overcome the drawbacks of the D-2DLDA algorithm, we improve it by endowing various weights to the projection vectors according to the importance, so that the important vectors will be emphasized and less important vectors will be deemphasized properly. A logarithmic function is introduced to obtain proper weight coefficients of the projection vectors according to their eigenvalues. We call our method log-based weighted 2DLDA (LW-2DLDA). The proposed LW-2DLDA algorithm is depicted as follows. The difference and advantages of LW-2DLDA will be discussed shortly later.

It can be noticed that the key difference of LW-2DLDA over D-2DLDA is the revision of Equation (4), i.e., the logarithmic function (or the reciprocal of logarithmic function), instead of square root is introduced to compute weight coefficients of the projection vectors

The LW-2DLDA Algorithm

Input: Training face images of all individuals: $A_j^{(i)}$, the parameter value of α.
Output: Features extracted from input images: $B_j^{(i)}$.

Step 1: Calculate the image between-class scatter matrix G_b and image within-class scatter matrix G_w according to Equation (1) and Equation (2), respectively.
Step 2: Diagonalize G_w by finding matrix V such that $V^T G_w V = D_w$, where $V^T V = I$ and $D_w = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix with the diagonal values sorted in decreasing order.
Step 3: Calculate $H = diag(\log_2 \lambda_1, \log_2 \lambda_2, \ldots, \log_2 \lambda_n)$, $Z = VH^{-1}$ and $\tilde{G}_b = Z^T G_b Z$.
Step 4: Diagonalize $\tilde{G}_b$ by eigen-analysis $U^T \tilde{G}_b U = D_b$, where $U^T U = I$ and $D_b = diag(\mu_1, \mu_2, \ldots, \mu_n)$ is a diagonal matrix with values in the diagonal position sorted in decreasing order.
Step 5: Calculate the discriminant matrix $W = ZQP$, where $Q = (u_1, u_2, \ldots, u_d)$ is constructed by the first d columns of U and $P = diag(\log_2 \mu_1, \log_2 \mu_2, \ldots, \log_2 \mu_d)$.
Step 6: Output the extracted features: $B_j^{(i)} = A_j^{(i)} (VH^{-1}QP)$.

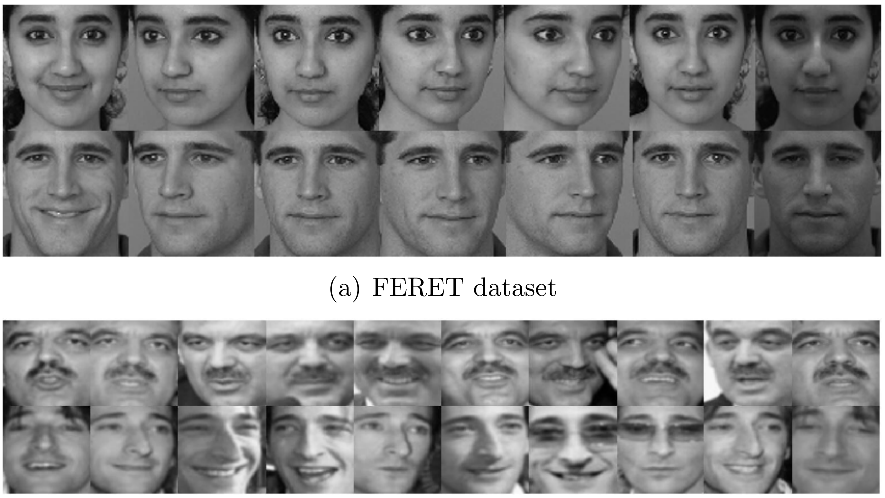
according to their eigenvalues, just as Equation (5) shows.

$$\begin{aligned} W &= VH^{-1}QP \\ &= \underbrace{[v_1, \ldots, v_d] \begin{bmatrix} \frac{1}{\log_\alpha \lambda_1} & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & \frac{1}{\log_\alpha \lambda_n} \end{bmatrix}}_{\text{weighted intra-class subspace}} * \underbrace{[u_1, \ldots, u_d] \begin{bmatrix} \log_\alpha \mu_1 & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & \log_\alpha \mu_d \end{bmatrix}}_{\text{weighted extra-class subspace}} \end{aligned} \tag{5}$$

By calculating the logarithmic value of each eigenvalue, the ratio between the largest and smallest weight coefficients is massively reduced, i.e., the weights are distributed in a narrower range than before, which is helpful for controlling the emphasizing to the projection vectors. According to Equation (5), the important projection vectors of both the intra-class subspace and extra-class subspace are gradually emphasized by the weight coefficients while the unimportant ones are deemphasized. Concretely, for the intra-class subspace, with an increase of the eigenvalues of the within-class matrix, the important eigenvectors corresponding to small eigenvalues are gradually stressed and the unimportant eigenvectors corresponding to large eigenvalues are slowly underrated; for the extra-class subspace, the eigenvectors corresponding to large eigenvalues are gradually enhanced and those corresponding to small eigenvalues are gradually depressed. Briefly, the input images are first mapped onto the weighted intra-class subspace to reduce the intra-class variations efficiently, and then are further mapped onto the weighted extra-class subspace to enhance the distinguishable information. Therefore, the features extracted through LW-2DLDA are more discriminative in representation.

3. **Experimental Results.** The proposed LW-2DLDA algorithm is used as a feature extraction method for face recognition. After projecting all the samples into the projective feature space, the nearest neighbor classifier (NNC) with Euclidean distance is applied to performing classification in the feature space.

To evaluate the performance of the proposed LW-2DLDA for face recognition over other related algorithms, we carry out several experiments on two popular datasets, i.e., the FERET and LFW (Label Faces in the Wild) face datasets. To assess the performance of algorithms, we select a subset of FERET dataset containing 1358 images of 194 individuals. Each individual has 7 face images resized to 80×80 pixels. For LFW dataset, we gather the subjects containing no less than ten samples and then get a dataset with 158 subjects from LFW-a database, and further choose the first 10 images of 158 individuals to construct the face subset for evaluation. All images in this subset are resized to 64×64 pixels. Figure 1 displays some sample images in FERET and LFW subset.



(a) FERET dataset

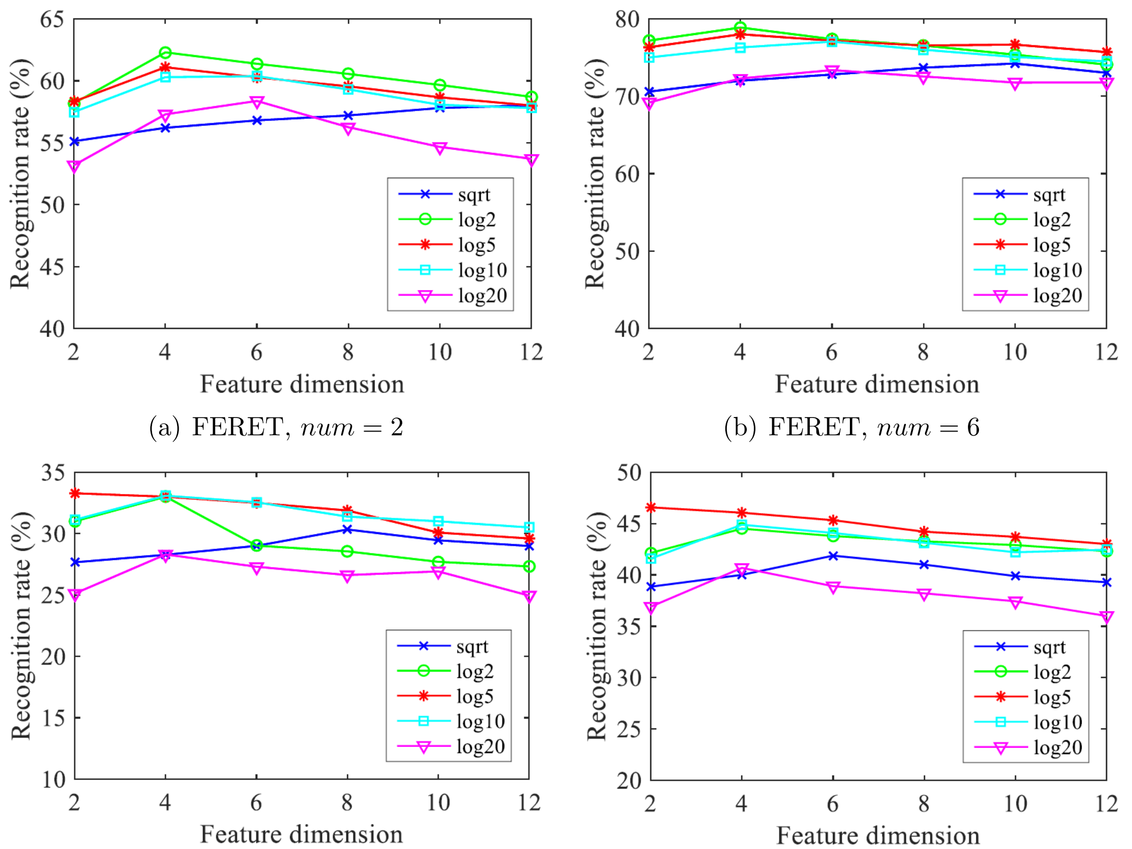
(b) LFW dataset

Figure 1. Sample face images from (a) FERET dataset and (b) LFW dataset

3.1. **The impact of parameter in logarithmic function.** To investigate the impact of values of α, which is a vital parameter in logarithmic function, we set the value of α to be 2, 5, 10, 20, respectively to calculate the weight coefficients of projection vectors according to the eigenvalues. Meanwhile, as an alternative weighting function of logarithmic functions, the square root function is also tested to be compared with. We randomly select num ($num = 2, 6$) images of each subject for training and the rest for testing in both FERET and LFW databases. We run each experiment for 10 times and Figure 2 displays the average recognition rates over various dimensions with respect to different weighting functions.

It shows in Figure 2 that when the value of α is 2, 5 or 10, the best average recognition accuracies achieved by logarithmic function make little difference in each group of experiments, and possess evident advantage over the square root function, especially in the more complicated LFW dataset. However, when the value of α is 20, which is a much larger base of logarithmic function, the highest recognition rates are significantly lower than the performance when small parameter values are used. According to Figure 2, when $\log 20$ is employed to calculate the weights, the performance is even worse than the square root function. This implies that smaller parameter values may more properly weigh the projection vectors and generate satisfying recognition performance compared with larger ones. It can also be observed from the figure that the parameter value is dataset dependent, since $\log 2$ produces the best performance on FERET dataset, while $\log 5$ achieves the highest recognition rate on LFW dataset. In the following experiments, we select $\log 2$ and $\log 5$ as the weighting functions for FERET and LFW dataset to calculate the weight coefficients, respectively.

3.2. **Performance comparison.** In this section, the proposed LW-2DLDA is tested with different training sample numbers on the two face databases, and compared with the other related 2DLDA methods. These compared methods are direct 2DLDA (D-2DLDA)



(a) FERET, $num = 2$ (b) FERET, $num = 6$

(c) LFW, $num = 2$ (d) LFW, $num = 6$

FIGURE 2. The effect of various weighting functions

TABLE 1. Recognition accuracy on FERET face database (mean $\pm$ std-dev)%

Method	$num = 2$	$num = 3$	$num = 4$	$num = 5$	$num = 6$
D-2DLDA	58.78 $\pm$ 6.03	65.50 $\pm$ 7.52	69.54 $\pm$ 8.31	72.20 $\pm$ 9.01	74.01 $\pm$ 11.32
L's 2DLDA	56.65 $\pm$ 9.06	60.25 $\pm$ 8.64	62.96 $\pm$ 8.25	66.25 $\pm$ 9.06	69.66 $\pm$ 10.44
P-2DLDA	60.58 $\pm$ 5.50	66.79 $\pm$ 9.13	71.98 $\pm$ 9.39	74.33 $\pm$ 10.67	77.02 $\pm$ 10.11
Re-2DLDA	57.61 $\pm$ 8.22	60.95 $\pm$ 9.24	64.06 $\pm$ 8.85	68.22 $\pm$ 9.60	71.72 $\pm$ 9.64
WF-2DLDA	61.05 $\pm$ 7.46	67.39 $\pm$ 7.55	71.88 $\pm$ 10.24	75.26 $\pm$ 10.18	78.52 $\pm$ 10.32
LW-2DLDA	**62.29 $\pm$ 6.82**	**68.86 $\pm$ 8.92**	**73.39 $\pm$ 9.25**	**76.56 $\pm$ 10.62**	**78.87 $\pm$ 11.41**

[9], parameterized 2DLDA (P-2DLDA) obtained according to [11], Loog's 2DLDA (L's 2DLDA) obtained by expanding the one dimensional weighted LDA in [8] to the two-dimensional case, regularized 2DLDA (Re-2DLDA) [4] and weighted fuzzy 2DLDA (WF-2DLDA) [10]. All of the above methods are used for feature extraction in our experiments. In each database, we randomly choose *num* (from 2 to 6) samples respectively from each class for training and the rest for testing purpose. For each value of *num*, we run the experiment for 10 times and report the best averaged performance. Table 1 and Table 2 respectively display the comparison results among different methods on FERET and LFW dataset.

As displayed in Table 1, the proposed LW-2DLDA achieves the best recognition performance in each experiment, and possesses the advantage of about 2 percent over the second-best method, i.e., P-2DLDA on FERET dataset, when the number of training samples in each class is small, i.e., $num = 2, 3, 4$. It shows in Table 2, the recognition rates of all methods on LFW database are much lower than those on FERET database, which implies that recognizing faces in LFW database is more challenging than FERET

TABLE 2. Recognition accuracy on LFW face database (mean ± std-dev)%

Method	$num = 2$	$num = 3$	$num = 4$	$num = 5$	$num = 6$
D-2DLDA	29.52 ± 9.50	34.88 ± 10.77	36.90 ± 12.62	38.23 ± 15.12	39.14 ± 16.43
L's 2DLDA	30.46 ± 9.76	32.26 ± 10.14	33.61 ± 12.58	35.26 ± 11.96	36.96 ± 17.23
P-2DLDA	31.66 ± 10.34	35.83 ± 12.44	40.07 ± 14.49	44.04 ± 13.26	45.39 ± 16.55
Re-2DLDA	31.94 ± 9.71	34.61 ± 10.39	36.16 ± 11.32	38.24 ± 14.06	39.99 ± 15.36
WF-2DLDA	32.42 ± 9.60	35.53 ± 11.68	38.12 ± 12.57	41.30 ± 14.45	42.64 ± 16.16
LW-2DLDA	**36.28 ± 9.45**	**39.56 ± 11.72**	**42.38 ± 13.31**	**44.76 ± 14.38**	**46.57 ± 15.20**

dataset. Even so, our LW-2DLDA approach defeats all other compared methods and performs the best when different numbers of samples are used for training. Moreover, the superiority of the proposed method over other methods on this dataset is more evident. According to both tables, we can observe that the advantage of LW-2DLDA over other methods is large when small number of samples used for training, is decreasing when more and more samples are used in the training process. This indicates that our proposed method is more suitable for addressing small sample size problem.

4. **Conclusions.** In this paper, we propose a novel and efficient log-based weighted 2DLDA (LW-2DLDA) algorithm for feature extraction. LW-2DLDA introduces a logarithmic function to recalculate the weight coefficients of projection vectors, which prevents the overemphasizing or over-deemphasizing effect in subspace. Therefore, more discriminative features can be produced by using the proposed approach. Encouraging experimental results demonstrate the effectiveness of our algorithm. In addition, the proposed LW-2DLDA can also be generalized to one dimensional case or embedded in methods which use LDA technique to process images.

In future work, we would attempt to expand our method to two-directional 2DLDA, which would be helpful for extracting discriminative features from two directions of images. Also, to integrate our proposed feature extraction idea with feature selection will be one interesting research topic.

Acknowledgment. This work was partially supported by the National Natural Science Foundation of China (No. 71390333), the National Key Technology R&D Program of China during the 12th Five-Year Plan Period (No. 2013BAD19B05), the Special Fund for Basic Research in Central University, i.e., Fund for Graduate Research Project of Jiangsu Province (KYZZ15_0072). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] Z. Fan, Y. Xu and D. Zhang, Local linear discriminant analysis framework using sample neighbors, *IEEE Trans. Neural Networks*, vol.22, no.7, pp.1119-1132, 2011.
[2] M. Zhao, Z. Zhang, T. W. Chow and B. Li, A general soft label based linear discriminant analysis for semi-supervised dimensionality reduction, *Neural Networks*, vol.55, no.10, pp.83-97, 2014.
[3] H. Wang, C. Ding and H. Huang, Multi-label linear discriminant analysis, *European Conference Computer Vision*, pp.126-139, 2010.
[4] H. Wang, X. Lu, Z. Hu and W. Zheng, Fisher discriminant analysis with L1-norm, *IEEE Trans. Cybernetics*, vol.44, no.6, pp.828-834, 2014.
[5] Y. Pang, S. Wang and Y. Yuan, Learning regularized LDA by clustering, *IEEE Trans. Neural Networks & Learning Systems*, vol.25, no.12, pp.2191-2201, 2014.
[6] S. Huang, D. Yang and J. Zhou, Graph regularized linear discriminant analysis and its generalization, *Pattern Analysis & Applications*, vol.18, no.3, pp.639-650, 2015.
[7] G. Shao and N. Sang, Regularized max-min linear discriminant analysis, *Pattern Recognition*, vol.66, pp.353-363, 2017.

[8] M. Loog, R. Duin and R. Haeb-Umbach, Multiclass linear dimension reduction by weighted pairwise Fisher criteria, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.2, no.7, pp.762-766, 2001.

[9] J. Lu, K. N. Plataniotis and A. N. Venetsanopoulos, Face recognition using LDA-based algorithms, *IEEE Trans. Neural Networks*, vol.14, no.1, pp.195-200, 2003.

[10] M. Xue, W. Liu and X. Liu, A novel weighted fuzzy LDA for face recognition using the genetic algorithm, *Neural Computing and Applications*, vol.22, no.7, pp.1531-1541, 2013.

[11] F. Song, D. Zhang, J. Wang, H. Liu and Q. Tao, A parameterized direct LDA and its application to face recognition, *Neurocomputing*, vol.71, no.1, pp.191-196, 2007.

[12] X. Zhang, D. Chu and R. Tan, Sparse uncorrelated linear discriminant analysis for undersampled problems, *IEEE Trans. Neural Networks & Learning Systems*, vol.27, no.7, pp.1469-1485, 2016.

[13] K. Liu, Y. Cheng and J. Yang, Algebraic feature extraction for image recognition based on an optimal discriminant criterion, *Pattern Recognition*, vol.26, no.6, pp.903-911, 1993.

[14] H. Xiong, M. Swamy and M. O. Ahmad, Two-dimensional FLD for face recognition, *Pattern Recognition*, vol.38, no.7, pp.1121-1124, 2005.

[15] C. Lu, S. An, W. Liu and X. Liu, An innovative weighted 2DLDA approach for face recognition, *Journal of Signal Processing Systems*, vol.65, no.1, pp.81-87, 2001.

[16] C. Li, Y. Shao and N. Deng, Robust L1-norm two-dimensional linear discriminant analysis, *Neural Networks*, vol.65(C), pp.92-104, 2015.

[17] J. Zhao, L. Shi and J. Zhu, Two-stage regularized linear discriminant analysis for 2-D data, *IEEE Trans. Neural Networks & Learning Systems*, vol.26, no.8, pp.1669-1681, 2015.

[18] X. Liu and H. Ye, Dual-kernel based 2D linear discriminant analysis for face recognition, *Journal of Ambient Intelligence & Humanized Computing*, vol.6, no.5, pp.557-562, 2015.