TWO-DIMENSIONAL ADAPTIVE DISCRIMINANT ANALYSIS

Yijuan Lu[1], Jie Yu[1], Nicu Sebe[2], Qi Tian[1]

[1]Department of Computer Science, University of Texas at San Antonio, TX, USA
[2]Faculty of Science, University of Amsterdam, Amsterdam, The Netherlands

ABSTRACT

In this paper, we develop a new feature extraction and dimension reduction technique: 2-Dimensional Adaptive Discriminant Analysis (2DADA) based on 2DLDA and our proposed 2DBDA. It effectively exploits favorable attributes of both 2DBDA and 2DLDA and avoids their unfavorable ones. 2DADA can easily find an optimal discriminative subspace with adaptation to different sample distributions. It not only alleviates the problem of high dimensionality, but also enhances the classification performance in the subspace with KNN classifier. Experimental results on hand-written digit database and face databases show an improvement of 2DADA over other traditional dimension reduction techniques.

***Index Terms*—** 2DLDA, 2DBDA, 2DADA, dimension reduction

1. INTRODUCTION

Recently, 2-Dimensional Linear Discriminant Analysis (2DLDA)[1-5] is becoming popular and widely used in face recognition and classification. Compared to 1-Dimensional LDA (1DLDA), 2DLDA works directly with images in their native state, as two-dimensional matrices**,** rather than 1D vectors. Hence the image does not need to be transformed, which not only saves the computational cost but also preserves all spatial information of the original images. In addition, the singularity problem resulting from the high-dimensionality of vectors is artfully solved.

Up to present, there are several variants of 2DLDA. Li *et al*. [1] and Sanguansat *et al*. [2] presented their 2DLDA with only reducing the number of columns and keeping the number of rows unchanged. Instead, Yang *et al.* [3] presented a two-step algorithm, which first reduces the number of columns and then reduces the number of rows. Ye *et al*. [4] proposed to calculate the row and column transformation matrices in an iterative way. Fortunately, they also recommended one iteration is sufficient because the accuracy curves were stable with respect to the number of iterations *T*. Later, Inoue *et al.* [5] proposed two non-iterative algorithms, namely selective and parallel algorithm. However, those two algorithms are more complex than the iterative one with *T*=1.

Based on Ye *et al.'s* 2DLDA [4], we extend 1DBDA to 2DBDA and combine 2DLDA and 2DBDA to propose a novel 2D Adaptive Disriminant Analysis (2DADA). 2DADA merges 2DLDA and 2DBDA in a unified framework and offers more flexibility and a richer set of alternatives to each individual method in the parametric space. 2DADA can easily find an optimal projection with adaptation to different sample distributions and discover a good classification in the subspace with *K*-NN classifier. Extensive experiments on the hand-written digit and face databases exhibit the superior performance of 2DADA.

2. 2D ADAPTIVE DISCRIMINANT ANALYSIS

2.1. 2DLDA

2DLDA tries to find two transformation matrices $L \in \Re^{r \times l_1}$ and $R \in \Re^{c \times l_2}$ that map each $X \in \Re^{r \times c}$ from originally high dimensional space to a low-dimensional space $Y = L^T XR \in \Re^{l_1 \times l_2}$, in which the most discriminant features are preserved. Intuitively it makes samples from the same class cluster to each other and samples from different classes separate from each other.

Mathematically, it could be modeled as finding two optimal projections *L* and *R* that maximizes the following ratios:

$$< L_{opt}, R_{opt} >= \arg\max_{L,R} \frac{| L^T \sum_{j=1}^{C} N_j \cdot (\boldsymbol{m}_j - \boldsymbol{m}) RR^T (\boldsymbol{m}_j - \boldsymbol{m})^T L |}{| L^T \sum_{j=1}^{C} \sum_{i=1}^{N_j} (\boldsymbol{x}_i^{(j)} - \boldsymbol{m}_j) RR^T (\boldsymbol{x}_i^{(j)} - \boldsymbol{m}_j)^T L |} \quad (1)$$

Here, the separability of class centers is measured by the between-class matrix $S_B = \sum_{j=1}^{C} N_j \cdot (\boldsymbol{m}_j - \boldsymbol{m})(\boldsymbol{m}_j - \boldsymbol{m})^T$ and the within-class scatter matrix $S_W = \sum_{j=1}^{C} \sum_{i=1}^{N_j} (\boldsymbol{x}_i^{(j)} - \boldsymbol{m}_j)(\boldsymbol{x}_i^{(j)} - \boldsymbol{m}_j)^T$ measures the within-class variance in the low-dimensional space. $\{\boldsymbol{x}_i^{(j)}, i = 1,..., N\}, j = 1,..., C$ denote the feature matrix of training samples. *C* is the number of classes. When *C*=2, it is 2D-Fisher Discriminant Analysis (2DFDA) and when C>2, it is called 2D-Multiple Discriminant Analysis (2DMDA), a natural extension of 2DFDA to multiple classes. N_j is the number of the samples of the j^{th} class,

$\boldsymbol{x}_i^{(j)}$ is the i^{th} sample from the j^{th} class, $\boldsymbol{m}_j$ is mean matrix of the j^{th} class, and $\boldsymbol{m}$ is grand mean of all examples.

Due to the difficulty of computing the optimal L and R simultaneously, Ye *et al.* [4] derive an iterative algorithm. Initially, $R_0 = \left(I_{l_2}, 0\right)^T$, we can compute the optimal

$$L_{t+1} = \arg\max_L \frac{|L_{t+1}^T \sum_{j=1}^{C} N_j \cdot (\boldsymbol{m}_j - \boldsymbol{m}) R_t R_t^T (\boldsymbol{m}_j - \boldsymbol{m})^T L_{t+1}|}{|L_{t+1}^T \sum_{j=1}^{C}\sum_{i=1}^{N_j} (\boldsymbol{x}_i^{(j)} - \boldsymbol{m}_j) R_t R_t^T (\boldsymbol{x}_i^{(j)} - \boldsymbol{m}_j)^T L_{t+1}|} \quad (2)$$

Next, with the computed L and calculate the optimal

$$R_{t+1} = \arg\max_R \frac{|R_{t+1}^T \sum_{j=1}^{C} N_j \cdot (\boldsymbol{m}_j - \boldsymbol{m}) L_t L_t^T (\boldsymbol{m}_j - \boldsymbol{m})^T R_{t+1}|}{|R_{t+1}^T \sum_{j=1}^{C}\sum_{i=1}^{N_j} (\boldsymbol{x}_i^{(j)} - \boldsymbol{m}_j) L_t L_t^T (\boldsymbol{x}_i^{(j)} - \boldsymbol{m}_j)^T R_{t+1}|} \quad (3)$$

This procedure is repeated for T iterations. In real application, T is often set to be 1.

2.2. 2DBDA

LDA makes the equivalent (unbiased) effort to cluster negative and positive samples. But intuition suggests that clustering the negative samples may be difficult and unnecessary because they may be from different classes (*Fig 1. (a)*). Hence, Biased Discriminant Analysis (BDA) [6] is proposed to cluster only positive samples and makes the negative samples far away from the positive ones. In this paper, we extend BDA to 2DBDA, which differs from 2DLDA in a modification on the computation of transformed between-class scatter matrix S_B and within-class scatter matrix S_W. They are replaced by $S_{N\to P}$ and S_P, respectively.

$$<L_{opt}, R_{opt}> = \arg\max_{L,R} \frac{|L^T \sum_{i\in Negative} (\boldsymbol{x}_i - \boldsymbol{m}_P) RR^T (\boldsymbol{x}_i - \boldsymbol{m}_P)^T L|}{|L^T \sum_{i\in Positive} (\boldsymbol{x}_i - \boldsymbol{m}_P) RR^T (\boldsymbol{x}_i - \boldsymbol{m}_P)^T L|} \quad (4)$$

where $\boldsymbol{m}_p$ is the mean matrix of the positive examples. $S_{N->P} = \sum_{i\in Negative} (\boldsymbol{x}_i - \boldsymbol{m}_P)(\boldsymbol{x}_i - \boldsymbol{m}_P)^T$ is the scatter matrix between the negative examples and the centroid of the positive examples, and $S_P = \sum_{i\in Positive} (\boldsymbol{x}_i - \boldsymbol{m}_P)(\boldsymbol{x}_i - \boldsymbol{m}_P)^T$ is the scatter matrix within the positive examples. $N \to P$ indicates the asymmetric property of this approach, which means the user's biased opinion towards the positive class, thus the name of biased discriminant analysis [6].

2.3. 2DADA

Given that 2DLDA and 2DBDA have their own assumptions and pay attention to different roles of the positive and the negative examples in finding the optimal discriminating subspace, it is our expectation that they can be unified.

In this paper, we propose a new method 2DADA (2D-Adaptive Discriminant Analysis) based on 2DLDA, 2DBDA and ADA [7], which finds an optimal projection.

$$<L_{opt}, R_{opt}> = \arg\max_{L,R} \frac{|L^T[(1-\lambda)\cdot S_{N\to P}^R + \lambda \cdot S_{P\to N}^R]L|}{|L^T[(1-\eta)\cdot S_P^R + \eta \cdot S_N^R]L|} \quad (5)$$

in which
$$S_{N\to P}^R = \sum_{i\in Negative} (\boldsymbol{x}_i - \boldsymbol{m}_P) RR^T (\boldsymbol{x}_i - \boldsymbol{m}_P)^T \quad (6)$$

$$S_{P\to N}^R = \sum_{i\in Positive} (\boldsymbol{x}_i - \boldsymbol{m}_N) RR^T (\boldsymbol{x}_i - \boldsymbol{m}_N)^T \quad (7)$$

$$S_P^R = \sum_{i\in Positive} (\boldsymbol{x}_i - \boldsymbol{m}_P) RR^T (\boldsymbol{x}_i - \boldsymbol{m}_P)^T \quad (8)$$

$$S_N^R = \sum_{i\in Negative} (\boldsymbol{x}_i - \boldsymbol{m}_N) RR^T (\boldsymbol{x}_i - \boldsymbol{m}_N)^T \quad (9)$$

Table I summarizes five special cases of 2DADA. From Table I, we can find that the 2DADA recovers 2DBDA when λ and η are set to be 0 and 0 in Case 1. Case 5 corresponds to a 2DLDA-like projection with λ and η set to 0.5 and 0.5. Case 4 finds a projection that is on the contrary side of 2DBDA, which is called Counter-2DBDA. Case 2 and Case 3 is a couple of contrary distribution scenarios, which assume that the negative (positive) samples are similar and positive (negative) samples might be from different classes. All these five cases fit certain sample distributions and have correspondence with some scenarios as illustrated in Fig. 1.

Table 1. Special cases of Adaptive Discriminant Analysis

(λ, η)	Optimal Projection	Note
(0,0)	$<L_{opt}, R_{opt}> = \arg\max_{L,R} \frac{\lvert LS_{N\to P}^R L^T\rvert}{\lvert LS_P^R L^T\rvert}$	Case 1 (2DBDA)
(0,1)	$<L_{opt}, R_{opt}> = \arg\max_{L,R} \frac{\lvert LS_{N\to P}^R L^T\rvert}{\lvert LS_N^R L^T\rvert}$	Case 2
(1,0)	$<L_{opt}, R_{opt}> = \arg\max_{L,R} \frac{\lvert LS_{P\to N}^R L^T\rvert}{\lvert LS_{P\to N}^R L^T\rvert}$	Case 3
(1,1)	$<L_{opt}, R_{opt}> = \arg\max_{L,R} \frac{\lvert LS_{P\to N}^R L^T\rvert}{\lvert LS_N^R L^T\rvert}$	Case 4 (Counter-2DBDA)
$(\frac{1}{2}, \frac{1}{2})$	$<L_{opt}, R_{opt}> = \arg\max_{L,R} \frac{\lvert L(S_{P\to N}^R + S_{N\to P}^R)L^T\rvert}{\lvert L(S_P^R + S_N^R)L^T\rvert}$	Case 5 (2DLDA-like)

In order to illustrate these five cases and show the advantages of 2DADA over 2DBDA and 2DLDA, we use synthetic data to simulate different sample distributions as shown in Fig. 1. Positive examples are marked with "+" s and negative examples are marked with "o" s. In each case, we apply 2DBDA, 2DLDA and 2DADA to find the best projection direction by their own criterion functions. The resulting projection lines are drawn in dotted, dash-dotted and solid lines, respectively. In addition, the distributions of positive and negative samples along these projections are also drawn like bell-shaped thicker and thinner curves along projection line, assuming Gaussian distribution for each class.

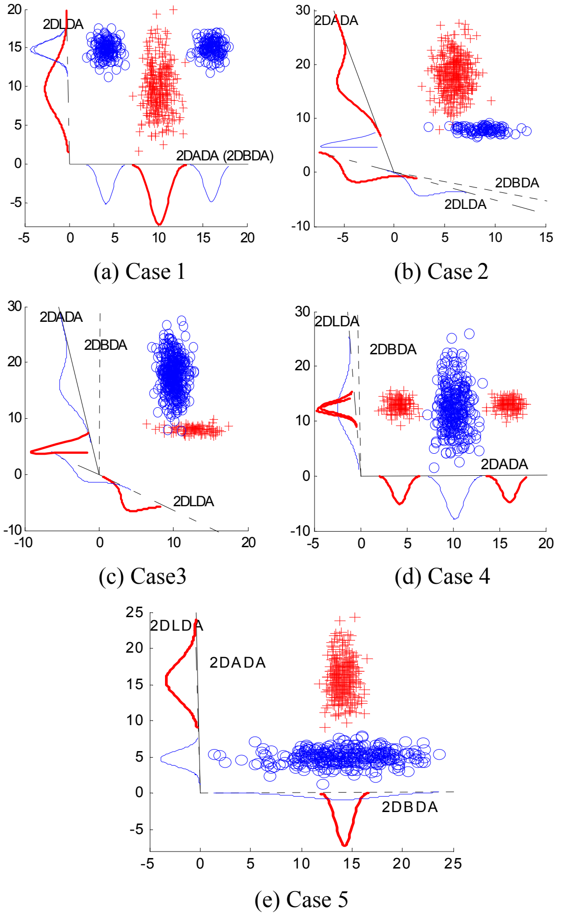


Fig. 1. Comparison of 2DBDA, 2DLDA and 2DADA for dimension reduction from 2-D to 1-D on synthetic data.

From Fig.1, we can see these five cases actually represent several typical data distribution scenarios. Case 1 best fits the distribution where all positive samples are alike while negative ones may be irrelevant (*Fig.1 (a)*). Case 4 is on the opposite side of Case 1, in which negative samples share strong correlations while positive samples may be quite different (*Fig.*1 *(d)*). Case 2 and Case 3 represent the imbalanced data set. In each case, the size of positive (negative) samples is much larger than that of negative (positive) samples (*Fig. 1(b) and Fig.1 (c)*). Case 5 is the scenario where the major descriptive directions of positive samples and negative genes are upright (*Fig. 1 (e)*).

From projection results, we can see 2DLDA treats positive and negative samples equally, *i.e.*, it tries to cluster the positive samples and decrease the scatter of the negative samples, although some positive (negative) samples maybe come from different sub-classes. This makes it a bad choice in Case 1 and Case 4. Similarly, since 2DBDA assumes all positive samples are projected together, it fails in Case 4 and Case 5. In Case 2 and Case 3, 2DBDA and 2DLDA are found not applicable for imbalanced data sets, because they tend to severely bias to cluster the dominating samples.

In all five cases, 2DADA yields good projection with positive samples and negative samples well separated and outperforms both 2DBDA and 2DLDA. It clearly demonstrates that no matter if it is an imbalanced data set or samples are from different sub-class clusters, 2DADA can adaptively fit into different distributions of samples and find a balance between clustering and separating, which are embedded in the criterion function. Here, we only show five special cases of 2DADA. More accurate data model fitting could be achieved by fine parameter tuning.

3. EXPERIMENTS AND ANALYSIS

In this section, we experimentally evaluate the performance of the 2DADA algorithm on hand-written digit recognition and face classification. In all experiments, our 2DADA is tested with (λ,η) evenly samples from 0 to 1 with step size of 0.1. Besides, 10-fold cross validation is used to report the mean accuracy of a *K*-NN query with *K*=10.

3.1. 2DADA for hand-written digit recognition

First, we tested 2DADA, 2DBDA, 2DLDA and their 1D-based methods on a subset of MNIST data set [8], which contains 400 similar hand-written 1's (200) and 7's (200). Some example images are showed in Fig.2.

Fig.2 Examples of hand-written images

In this experiment, the original dimension of images 28×28 is reduced to $d \times d$ (d<28) by all 2D based methods. Correspondingly, the reduced dimension *p* in their 1D based methods is chosen such that both 1D and 2D methods use the same amount of storage for the transformation matrices and the reduced presentations [9]. For examples, on the MNIST dataset, *d*=2,4,6,8,10,12,14 for 2DADA are used, corresponding to *p* =2,6,12,22,34, 49 and 67 for 1DADA.

The average accuracy rate across 10-fold cross validation over the variation of dimension *d* is plotted in Fig. 3, where the x-axis denotes the values of *d* (between 2 to 14).

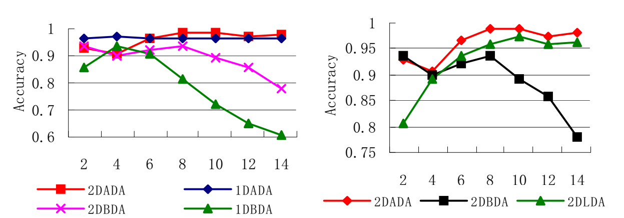


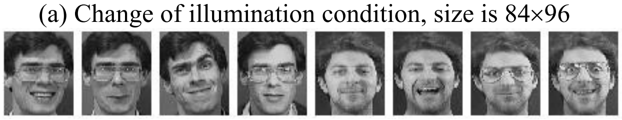
Fig.3. Comparison of accuracy with different dimensions.

From Fig.3, we can clearly find: 1) 2D based approaches (*i.e.* 2DADA, 2DBDA) achieve higher or comparable accuracy with their 1D based approach (1DADA and 1DBDA). In addition, in our experiments, we found 2D methods are almost one order of magnitude faster than 1D methods. It justifies that 2D techniques have lower loss of information and computational effective with the same amount of storage. 2) 2DADA consistently outperforms others irrespective of variation in dimensions. Its stableness verifies that it is a powerful dimension reduction method for classification.

3.2. 2DADA for face classification

To evaluate 2DADA for face classification, we tested it on three well-known face image databases with change of illumination, expression and head pose, respectively. The Harvard Face image database contains images from 10 individuals, each providing total 66 images, which are classified into 10 sets based on increasingly changed illumination condition [10]. The AT&T Face Image database [11] consists of grayscale images of 40 persons. Each person has 10 images with different expressions, open or closed eyes, smiling or non-smiling and wearing glasses or no glasses. The UMIST Face Database [12] consists of 564 images of 20 people, which covers a range of poses from profile to frontal views. Figure 4 gives some example images from the databases.

(a) Change of illumination condition, size is 84×96



(b) Change of expressions, size is 92×112

(c) Change of head pose, size is 92×112

Fig.4. Example Face images from three databases

Table 2. Comparison of 2D based methods on three different face databases.

Accuracy (%)	Harvard Database			AT&T Database	UMIST Database
	Subset1	Subset2	Subset3		
2DADA	**95**	**95.5**	**94.6**	**98.5**	98.5
2DBDA	83.3	91.1	82.3	**98.5**	94.3
2DLDA	91.6	93.3	91.7	97.5	96.1
2DPCA	90	90	90	98	**99.8**

Table 2 shows classification accuracy of 2DADA, 2DLDA, 2DBDA and 2DPCA on these three datasets with reduced dimension *d*=6. For each database, we randomly chose one person's face images as positive and the rest face images of others are considered as negative.

It is clear from Table 2 that the proposed 2DADA still performs better or comparable to other techniques in all tests and more robustness to the changes of illumination, expression and pose than other techniques. It clearly demonstrates that 2DADA could find the most discriminant features that fit different distributions of samples and classification task.

4. CONCLUSIONS

This paper proposes a novel 2-Dimensional Adaptive Discriminant Analysis (2DADA) for high dimensionality problem. 2DADA provides a richer set of alternatives to 2DLDA and 2DBDA. As a result, it takes advantage of both of their merits and finds an optimal projection with adaptation to different sample distributions. The proposed approach is applied to hand-written digit recognition and face classification. Its superior performance demonstrates that 2DADA is a promising and efficient dimension reduction approach.

Acknowledgement: This work was supported in part by San Antonio Life Science Institute (SALSI) and the Army Research Office (ARO) grant under W911NF-05-1-0404.

5. REFERENCES

[1] M. Li and B. Yuan. "2D-LDA: A novel statistical linear discriminant analysis for image matrix," *Pattern Recognition Letters*, 26(5): 527-532, 2005.

[2] P. Sanguansat, W. Asdornwised, *et al.*, "Two-dimensional linear discriminant analysis of principle component vectors for face recognition," *Proc. of ICASSP*, 345-348, *2006*.

[3] J. Yang, D. Zhang, X. Yong, and J. Yang, "Two-dimensional linear discriminant transform for face recognition," *Pattern Recognition*, 38 (7): 1125-1129, 2005.

[4] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," *Advances in Neural Information Processing Systems (NIPS2004)*, 17:1569-1576, 2004.

[5] K. Inoue and K. Urahama, "Non-Iterative Two-Dimensional Linear Discriminant Analysisg," *Proc. of ICPR*, 2006.

[6] X. Zhou and T. S. Huang, "Small sample learning during multimedia retrieval using biasMap," *IEEE CVPR,* 2001.

[7] J. Yu and Q. Tian, "Adaptive discriminant projection for content-based image retrieval," *Proc. of ICPR*, 2006.

[8] Y. LeCun, *et al.*, http://yann.lecun.com/exdb/mnist/.

[9] J. Ye, R. Janardan, and Q. Li, "GPCA: An efficient dimension reduction scheme for image compression and retrieval," *KDD'04*, August 22-25, 2004.

[10] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans, PAMI*, Vol. 19, No.7, July 1997.

[11] H.A. Rowley, S.Baluja, and T.Kanade, "Neural Network-Based Face Detection," *IEEE Trans. PAMI*, Vol.20, 1998.

[12] F.Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," *IEEE workshop on Application of Computer Vision*, Sarasota FL, December 1994.