Similar Handwritten Chinese Character Recognition using Discriminative Locality Alignment Manifold Learning

Dapeng Tao, Lingyu Liang, Lianwen Jin, Yan Gao
College of Electronic and Information
South China University of Technology, China
E-mails: {dapeng.tao, lianglysky, lianwen.jin}@gmail.com

***Abstract* -- The discriminant analysis for Similar Handwritten Chinese Character Recognition (SHCR) is essential for the improvement of handwritten Chinese character recognition performance. In this paper, a new manifold based subspace learning algorithm, Discriminative Locality Alignment (DLA), is introduced into SHCR. Experimental results demonstrate that DLA is consistently superior to LDA (Linear Discriminant Analysis) in terms of discriminate information extraction, dimension reduction and recognition accuracy. In addition, DLA reveals some attractive intrinsic properties for numeric calculation, e.g. it can overcome the matrix singular problem and small sample size problem in SHCR.**

Keywords: Discriminative Locality Alignment, subspace learning, similar handwritten Chinese character recogniton (SHCR), LDA

I. Introduction

In recent years, handwritten Chinese character recognition (HCR) has made great progress in both research and practical application. Unconstrained cursive online HCR, however, is still an open problem remaining to be solved, for it is still challenging to reach high recognition rate considering the high diversity of handwriting styles and large capacity of category set [1][2][3][4]. In constrained HCR, recognition rate can generally reach to over 98.5%; in unconstrained cursive online HCR, the rate falls to 92.39% [2], where there is a decrease of nearly 6%.

Many effective methods have been proposed to promote the recognition rate in unconstrained cursive online HCR. C.L. Liu [4] proposed a hierarch classifier composed of DFE and DLQDF. Jin et. al. [6] presented an incremental LDA (ILDA) model, which can implement writer adaptive recognition by updating the LDA transformation matrix and the classifier prototypes in the discriminative feature space. All of the methods above are concerned with constructing a global optimized model to improve recognition accuracy..

In fact, one of the main reasons for the performance degradation in unconstrained cursive online HCR lies in the subtle resemblance between similar cursive handwritten characters. In CHRC2010 [2], the average rate of the top ten candidates reaches up to 98.95%, whereas the first candidate accuracy falls down to 93%. It reveals that the recognition enhancement for similar handwritten Chinese character recognition (SHCR) is crucial for the improvement of the global recognition performance. Fig. 1 illustrates some similar cursive samples from the CASIA-OLHWD1 database [5].

Therefore, in contrast to the global optimized model, it also makes sense to propose some 'local optimized' methods targeting on the 'confusing' samples.

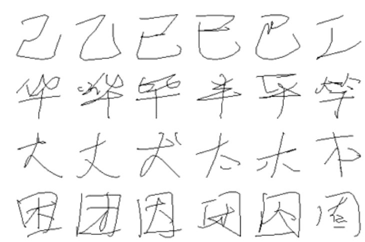

Figure 1. Similar samples in handwritten Chinese character from CASIA-OLHWD1 database

In SHCR, discriminate information extraction is considerably essential [1][8][12]. Liner Discriminate Analysis (LDA) [7] is one of the widely used discriminate feature extraction methods in the literature. Traditional LDA, however, suffers from the following drawbacks. First, it ignores the local structure properties between the samples, which makes it fail to discover the nonlinear structure hidden in the high dimensional data. Second, a large number of training samples are required to make a good model approximation, i.e. LDA is confronted with the small sample size (SSS) problem. Third, the singular problem may arise when computing the projection matrix. Finally, the reduced dimension has an upper bound of $C-1$ where C is the class number.

To overcome the problems of LDA, we introduces a state-of-the-art subspace manifold learning approach into SHCR, which is called Discriminative Locality Alignment (DLA) [10][11]. DLA is a manifold based dimension reduction method presented recently [10]. As a manifold based subspace learning method, DLA have many attractive properties for SHCR compared to LDA. First, DLA model focuses on local discriminate structure for each training sample, which implies the better discriminate performance. Second, relatively small labeled training samples are enough for DLA to obtain satisfactory high recognition accuracy. Third, no singular problem and no upper dimension bound restriction for dimension reduction in DLA. In addition, the experiments in this paper show that smaller projection matrix can be obtained in DLA, whereas the recognition rate maintains high for SHCR. That means DLA is able to keep a high accuracy with smaller computing and storage cost, which is attractive for practical application.

The rest of the paper is organized as follows. Section 2 introduces discriminate feature extraction for SHCR using

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.205

DLA method. Experiments and analysis are presented in Section 3. Section 4 concludes this paper.

II. Proposed Method

In this paper, the DLA-based SHCR is consists of the following three stages:

- **Similar samples collection and feature extraction:** Several sets of similar handwritten Chinese are firstly built using the static candidate generation technique [12]. Then the 8-directional features [9] were extracted with D dimensions.
- **Subspace learning using DLA:** In order to find a proper subspace Y from the training samples set X, the projection matrix U is found after the DLA manifold learning. For the training set X including N samples, i.e. $X=[x_1,...,x_N]\in\mathbb{R}^{D\times N}$, and the subspace $Y\in\mathbb{R}^{d\times N}$ with dimensions $d<D$, the discriminate subspace could be obtained by $Y=U^TX$.
- **Classification:** the minimum Euclidean distance classifier is implemented to classify the data in subspace Y.

A. Similar samples collection and feature extraction

Similar sample set is formed by using the static candidate generation technique presented in [12]. The process is carried out as follows: *First*, central sample vectors for each category and the distance between them are calculated. After comparing all the distances, l categories that are much closer than others are selected. Then, a static similar character table is formed for all categories, each of which has l similar characters. *Second*, an unconstrained cursive online HCR is implemented and the output is the first candidate character. Third, similar samples for each character are collection according to the table formed in the first step.

We used the 8-directional features extraction method proposed by Z.L. Bai and Q.Huo [9] to extract feature for similar samples.

B. Subspace projection using DLA manifold learning

a) LDA subspace learning

LDA aims at maximizing the distances between means of each class and minimizing the distances of within-class scatter simultaneously in the projected sample subspace [11]. The objection function of LDA is given by:

$$\arg\max_U \frac{tr(U^TS_wU)}{tr(U^TS_bU)} \quad s.t. UU^T=I_d \tag{1}$$

where

$$S_w=\sum_{j=1}^{C}\sum_{i=1}^{N_j}(x_i^{(j)}-m_j)(x_i^{(j)}-m_j)^T; \tag{2}$$

$$S_b=\sum_{j=1}^{M}N_j(m_j-m)(m_j-m)^T; \tag{3}$$

S_W is the within-class scatter matrix; S_b is the between-class scatter matrix; m_j is the sample mean for the jth class; m is the sample mean for all samples.

If the original feature space is $X=[x_1,...,x_N]\in\mathbb{R}^{D\times N}$, to find a proper subspace $Y\in\mathbb{R}^{d\times N}$ that preserved the dominative discriminate information to recognize similar character effectively, the projection matrix U should be obtained by maximizing Eq.(1). If S_w is nonsingular, then U can be obtained by solving a conventional eigenvalue problem of $S_w^{-1}S_b$. Suppose the subspace dimension is d; then U is composed of d eigenvector corresponding to the d largest eigenvalue. The projected subspace is obtained by $Y=U^TX$.

b) DLA subspace learning

Different from the global linear optimization principle in LDA, DLA aims to preserve the discriminate information in a local patch instead of the global linear structure of LDA. In each patch, DLA firstly operates "part optimization" to a given sample, so that in a low dimensional subspace, the distance between the given sample and its neighbors in identical class will be as small as possible, whereas its neighbors in different class will be as large as possible. Then, DLA operates "whole alignment" to integrate all the weighted part optimization to form a global subspace structure [10].

1) Part Optimization

The part optimization stage of DLA starts from each given sample and the corresponding patch. Each patch is built by the given sample and its neighbors including the samples from both the same and different classes [10][11].

For a given sample x_i and its corresponding patch, we can find m_1 closest samples $x_{i^1},...,x_{i^{m_1}}$ that from the same class with x_i, and m_2 closest samples $x_{i_1},...,x_{i_{m2}}$ that from different classes. Let the training set be $X_i=[x_i,x_{i^1},...x_{i^{m1}},x_{i_1},...,x_{i_{m2}}]$. The goal of part optimization is to find a new low dimensional subspace $Y_i=[y_i,y_{i^1},...y_{i^{m1}},y_{i_1},...,y_{i_{m2}}]$. In the subspace, the between-class distance is maximized, whereas the within-class distance is minimized.

Fig. 2 illustrates the process of part optimization in the situation when $m_1=3, m_2=2$. It shows that in the projected subspace, x_i is closer to the samples from identical class (x_{i^1} x_{i^2} x_{i^3}), whereas the distance between x_i and the samples from different class (x_{i_1} x_{i_2}) is larger.

The optimization function in part optimization stage is given by:

$$\arg\min\left(\sum_{j=1}^{m_1}\left\|y_i-y_{i^j}\right\|^2-\beta\sum_{p=1}^{m_2}\left\|y_i-y_{i_p}\right\|^2\right), \tag{4}$$

where β is a scaling factor in [0, 1], which can change the contribution to optimization function for within-class and between-class distance.

If we define the coefficients vector and matrix L_i that contains the local geometry property and discriminative information, then Eq. (5) can be reduced to [10]:

$$\arg\min_{Y_i} tr(Y_i L_i Y_i^T) \tag{5}$$

where,

$$\omega_i = [\overbrace{1,...,1}^{m}, \overbrace{-\beta,...,-\beta}^{n}]^T \tag{6}$$

$$L_i = \begin{bmatrix} \sum_{j=1}^{m+n}(\omega_i) & -\omega_i^T \\ -\omega_i & diag(\omega_i) \end{bmatrix} \tag{7}$$

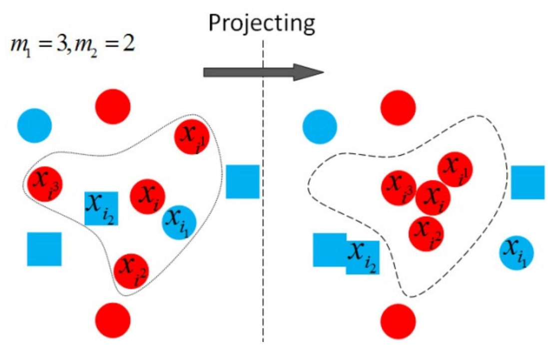

Figure 2. The process of part optimization.

2) *Whole alignment*

After part optimization stage, we obtain N local alignment matrixes $L_i(i = 1...N)$ for each sample. In the whole alignment stage, each L_i are summarized according to sample weighting in a global coordinate to form the global alignment matrix L [13], then the objection function is given as:

$$\arg\min_{Y} tr(YLY^T) \tag{8}$$

Since Y can be obtained by $Y = U^T X,\ s.t. UU^T = I_d$, Eq. (8) can be then rewritten as:

$$\begin{aligned} &\arg\min_{U} tr(U^T XLX^T U) \\ &s.t. U^T U = I_d \end{aligned} \tag{9}$$

where U is the projection matrix.

U can be obtained by solving a conventional eigenvalue problem for XLX^T, i.e. then U is composed of d eigenvector corresponding to the d smallest eigenvalue of XLX^T.

C. *Classifier*

After the 8-directional features [9] are extracted for the similar handwritten characters, the original features are projected into low dimensional discriminant subspace using either LDA or DLA. Then a simple minimum Euclidean distance classifier is used for recognition.

III. EXPERIMENTS

A. *Experimental Data*

In this paper, the benchmark dataset comes from the SCUT-COUCH2009 dataset [14]. SCUT-COUCH2009 is an online unconstrained Chinese handwriting dataset, which contains 11 subsets of different vocabularies, including GB1, GB2, Letters, Digit, Symbol, Word8888 etc, and all the samples are collected from more than 190 subjects. In the following experiments, the GB1 subset is used, which contains 3,755 frequently used simplified Chinese characters in GB-2312-80 standard. In SHCR experiments, 10 sets of similar characters are randomly selected. Table Ⅰ lists the ten similar character sets we used in the following experiments. Fig.3 shows some similar characters and the corresponding handwritten samples.

TABLE I. SIMILAR CHARACTERS SETS

Set#	First Candidate	Similar characters
1	癌	癌瘤瘟瘸痞痴痹瘪痛痊
2	爱	爱受复赏夏曼笺聂赁缓
3	氨	氨氮氦氯氢氛氰姜氘氟
4	盎	盎蛊盅盏盔盗鸯签置盘
5	笆	笆爸琶筐邑笔芭岂色苞
6	茬	茬荐若苯荒茎巷苦花芜
7	差	差羞羌姜善羔美养羡盖
8	柴	柴柒紫染架浆渠梨梁粱
9	敞	敞敝敲敬澈版骸歇淑散
10	大	大丈人太犬木友支术入

Figure 3. Similar charcter samples and the corresponding handwritten samples

B. *Parameters optimizaion for DLA*

Since the parameters setup for DLA is essential for its performance, we carried out the DLA parameter optimization experiments before for SHCR. We aims to find a proper range for the dominant parameters m_1, m_2 in DLA, where m_1 is the number of the samples from identical class in the given patch, and m_2 is the number of the samples from other classes in the same given patch. Parameter β is set to an empirical value 0.15 and the reduced dimension is set to 9.

Suppose n is the training sample number in each class, N is the total training sample number, and C is class number. We have $N = C \times n$ Then, m_1 and m_2 could be chosen in the range of $[1, n-1]$ and $[0, N-n]$ respectively.

Fig. 4 shows the recognition rate against different m_1 and m_2 values using the similar samples table mentioned above (See Table1 and Fig. 3), which character set with used "氨" as the first candidate. When $n = 30$, different combinations of m_1, m_2 pairs result in different recognition rates All the possible combination of m, n and the corresponding accuracy are visualized in Fig. 4, where the red region represented the best performance by DLA.

In this experiment, Fig. 4 shows that, the best combination of m_1, m_2 is $m_1 = 29, m_2 = 50$, with the corresponding accuracy 95.82%. It is worthwhile to note that when the parameters are chosen to $m_1 = 10, m_2 = 30$, the recognition rate can reach to 95.4%, which is a only rate decrease of less than 0.5% to 95.82 %($m_1 = 29, m_2 = 50$). Since m_1, m_2 could be chosen in a reasonable board region, and considering the computing cost, a sub-optimization combination $m_1 = 10, m_2 = 30$ is chosen in the following experiments for all other similar character sets.

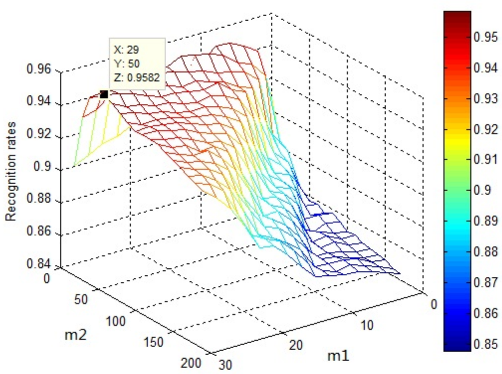

Figure 4. Recognition rate vs. m_1 and m_2 for DLA parameters optimization

C. *DLA evaluation experiments for SHCR*

To evaluation the performance of DLA in SHCR, the comparison experiments between DLA and LDA are carried out. In the experiments, we compare DLA and LDA in small training sample size setting (hereafter we denote them as DLA1 and LDA1 respectively) and (relatively) large training sample size setting (denoted as DLA2 and LDA2 respectively) for the ten similar characters sets listed in Table Ⅰ. Note that each class of character has 188 training and testing samples in total.

In DLA1 and LDA1 setting, we randomly select 30 of 188 samples per character for training (i.e. $n = 30$); and in DLA2 and LDA2 we set $n = 80$. In both the settings, the remaining samples are used for testing. Fig. 5 shows the recognition rate versus reduced dimensions for the ten similar character sets.

For comparison convenience, we arrange the experiment results in Table Ⅱ and Table Ⅲ for the ten similar character sets. Table Ⅱ lists the average recognition rates and the corresponding reduced dimensions for both DLA and LDA under two different data settings; whereas Table Ⅲ lists the best recognition rates.

TABLE II. AVERAGE RECOGNITION RATES(%)

Reduced Dimension	n=30		n=80	
	DLA1	LDA1	DLA2	LDA2
1	0.361	0.269	0.358	0.330
2	0.638	0.428	0.608	0.527
3	0.789	0.560	0.791	0.658
4	0.860	0.647	0.865	0.740
5	0.908	0.729	0.912	0.796
6	0.927	0.775	0.933	0.835
7	0.937	0.808	0.944	0.869
8	0.947	0.838	0.957	0.897
9	0.957	0.856	0.963	0.915

D. *Analysis of the Results*

From Fig. 5, Table Ⅱ and Table Ⅲ, the performance of DLA in SHCR can be analyzed in three aspects:

- In Fig.5 and Table Ⅱ, it is shown that in the same reduced dimensions, the recognition rates of both DLA1 and DLA2 are significantly higher than that of LDA1 and LDA2 respectively. It also shows when the recognition rate is in the same level, DLA have a better dimension reduction performance than LDA. For example, we can see from Table Ⅱ that when recognition rate reaches to over 0.85%, the DLA1 with reduced dimension of 4 outperforms LDA1 with reduced dimension of 9.
- In Fig. 5 and Table Ⅱ, it can be seen that in the same reduced dimensions, recognition rates in DLA1 and DLA2 have just a little variation; whereas the rates in LDA1 are consistently lower than LDA2. It demonstrates that DLA can maintain better recognition performance than LDA in SHCR when the training samples size is relatively small.
- From Table Ⅲ, it is shown that the best recognition rates in DLA1 and DLA2 are consistently higher than those in LDA1 and LDA2 for all similar character sets. It confirms us that DLA have better discriminant recognition performance than LDA for SHCR.

IV. CONCLUSION

In this paper, a manifold based subspace learning algorithm, Discriminative Locality Alignment (DLA), has been introduced for similar handwritten Chinese character recognition (SHCR). Comparing to the traditional widely used LDA subspace learning techniques, DLA has shown many competitive and attractive properties, and it is consistently superior to LDA. From the experiments, we can draw the following conclusions:

1) The discriminate information extraction and dimension reduction performance of DLA is very competitive in SHCR, for it can consistently achieve better recognition accuracies and better dimension reduction than LDA in the SHCR experiments.
2) In SHCR, DLA is a robust and promising manifold learning method that overcomes many computation problems including matrix singular problem, small sample size problem, and reduced dimension upper bound problem.
3) DLA is potentially useful for real world applications, for it can perform high recognition accuracy with a smaller projection matrix than that of LDA. That results in a much smaller storage cost with higher recognition performance, which could be very useful for many practical recognition applications.

ACKNOWLEDGMENT

This work is supported in part by NSFC (Grant no.U0735004 and 61075021), GDSTP (no. 2008A050200004, 2010B090400397, 2010A090100016).

REFERENCES

[1] T. Gao, C. Liu. High accuracy handwritten Chinese character recognition using LDA-based compound distances. Pattern Recognition 41(11) (2008) 3442-3451

[2] C. Liu, F.Yin, D. Wang, Q. Wang. Chinese Handwritten Recognition Contest 2010. CCPR 2010, pp. 1-5.

[3] H. Zhang, J. Yang, W. Deng, J. Guo. Handwritten Chinese character recognition using Local Discriminant Projection with Prior Information. ICPR 2008, pp. 1-4.

[4] C. Liu. High Accuracy Handwritten Chinese Character Recognition Using Quadratic Classifiers with Discriminative Feature Extraction. ICPR 2006, pp. 942-945.

[5] D. Wang, C. Liu, J. Yu, X. Zhou. CASIA-OLHWDB1: A Database of Online Handwritten Chinese Characters. ICDAR 2009, pp. 1206-1210 .

[6] L. Jin, K. Ding, Z. Huang. Incremental learning of LDA model for Chinese writer adaptation. Neural Computing, 2010, Vol. 73, pp. 1614–1623.

[7] R.A. Fisher The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics 1936, Vol. 7, pp. 179-188.

[8] H. Liu and X. Ding. Improve handwritten character recognition performance by heteroscedastic linear discriminant analysis. ICPR 2006, pp. 880 - 883 .

[9] Z. BAI and Q. HUO. A Study On the Use of 8-Directional Features For Online Handwritten Chinese Character Recognition. ICDAR 2005, pp. 232-236.

[10] T. Zhang, D. Tao, and J. Yang. Discriminative locality alignment. ECCV 2008 , pp. 725–738.

[11] T. Zhang, D. Tao, X. Li, and J. Yang. Patch Alignment for Dimensionality Reduction. IEEE Trans. Knowl. Data Eng. 2009, Vol. 21, 9, pp. 1299–1313.

[12] Z. Liu, L. Jin. A Static Candidates Generation Technique and its Application in Two-stage LDA Chinese Character Recognition. Chinese Control Conference 2007.

[13] Z. Zhang and H. Zha. Principal Manifolds and Nonlinear Di-mension Reduction via Local Tangent Space Alignment. SIAM J. Scientific Computing, 2005, Vol. 26, pp. 313-338.

[14] L. Jin, Y. Gao, G. Liu, Y. Li, K. Ding. A Comprehensive Online Unconstrained Chinese Handwritten Database and Benchmark Evaluation. IJDAR vol. 14, no.1, pp53-64,2011.

TABLE III. BEST RECOGNITION RATES(%)

		SET 1	SET 2	SET 3	SET 4	SET 5	SET 6	SET 7	SET 8	SET 9	SET 10
n=30	DLA1	0.981(9)	**0.977**(9)	0.943(9)	0.966(11)	0.968(14)	**0.950**(15)	0.938(10)	0.925(17)	0.959(9)	**0.971**(10)
	LDA1	0.923(9)	0.898(9)	0.805(9)	0.896(9)	0.853(9)	0.810(9)	0.791(9)	0.785(8)	0.876(9)	0.927(9)
n=80	DLA2	**0.986**(9)	0.973(9)	**0.954**(9)	**0.979**(12)	**0.970**(9)	0.944(9)	**0.943**(9)	**0.944**(9)	**0.967**(9)	0.969(13)
	LDA2	0.940(9)	0.935(9)	0.869(9)	0.950(9)	0.930(9)	0.886(9)	0.881(9)	0.874(9)	0.933(9)	0.947(9)

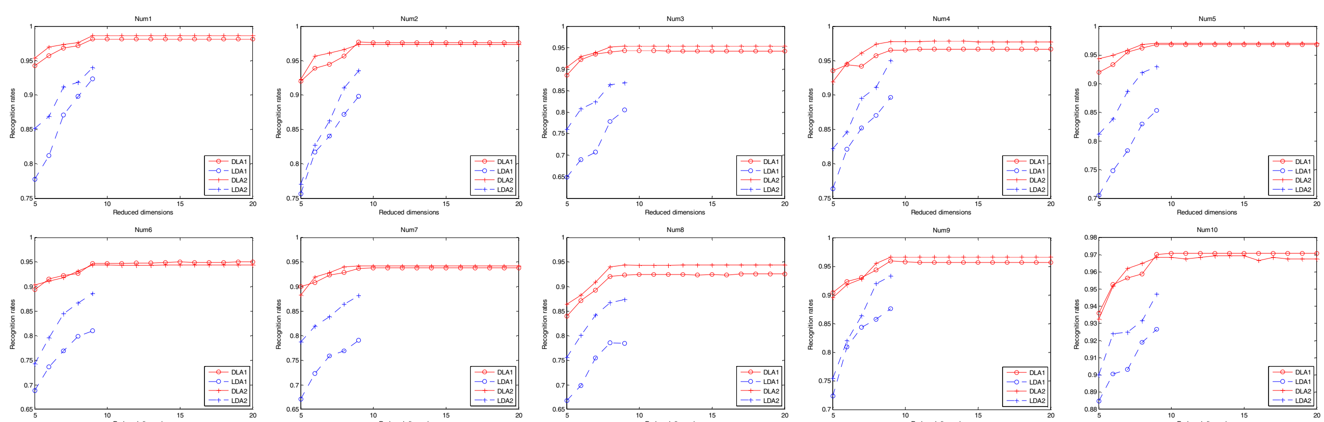

Figure 5. Recognition rate vs. diminsionality for ten similar character groups