Modified Two-class LDA Based Compound Distance for Similar Handwritten Chinese Characters Discrimination

Yunxue Shao, Chunheng Wang, Baihua Xiao, Rongguo Zhang, Linbo Zhang
The Key Laboratory of Complex Systems and Intelligence Science
Institute of Automation Chinese Academy of Sciences
95 Zhongguancun East Road, 100190, BEIJING, CHINA
{ yunxue.shao, chunheng.wang, baihua.xiao, rongguo.zhang, linbo.zhang }@ia.ac.cn

***Abstract*—This paper proposes a modified two-class LDA based compound distance for similar handwritten Chinese characters discrimination. First the definition of the Intersecting Subspace (IS) between two classes and the modified between-class scatter matrix is given. Then we prove that the modified between-class scatter matrix can supply additional information. Our experiments demonstrate that the additional information can be used to discriminate points in the IS and the proposed method outperforms the previous LDA based method.**

Keywords-intersecting subspace; similar handwritten Chinese character; modified two-class linear discriminant analyse;compound distance

I. INTRODUCTION

The problem of offline handwritten Chinese character recognition has been investigated by many researchers over a long time, and great improvements have been achieved [1-4]. Despite the success of existing methods, there are still rooms for improvement. One of the problems is the classification of similar characters. K.C.Leung and C.H.Leung [5] proposed the “critical region analysis” technique to tackle the problem of similar character classes. Bo Xu et al. [6] proposed an Average Symmetric Uncertainty based critical region selection method and they showed that the critical regions selected by their method contain more discriminative information than by the method proposed in [5]. The compound Mahalanobis function (CMF) method, proposed by Suzuki et al. [7], combines pair discrimination measures with class-wise Mahalanobis distance, this method is unique in that the pair discriminator has no extra parameters when using the MQDF as the baseline classifier. Tian-Fu Gao and Cheng-Lin Liu [8, 9] proposed a LDA-based compound distance method to discriminate similar character pairs, and they showed that under restrictive assumptions, the previous CMF is a special case of the LDA-based compound distance, and their experiments demonstrated that the LDA-based compound distance method outperforms the previous CMF methods.

The MQDF [10] has been widely used to handwritten Chinese character recognition with great success. In this paper, we use it as a baseline classifier and use the LDA-based compound distance method and the proposed modified two-class LDA based compound distance method (M2LDA) to discriminate similar pairs.

In the following sections, we first briefly review the MQDF and the LDA-based compound distance in Section2, and then in Section 3, the modified two-class LDA based compound distance is proposed and some implemental notes are given, followed by experimental results in Section 4 and conclusions in Section 5.

II. MQDF AND LDA-BASED COMPOUND DISTANCE

A. MQDF

For a class $\omega^{(c)}$, the quadratic discriminant function (QDF) of an n-dimensional feature vector is given as

$$g_0^{(c)}(x) = (x-\mu^{(c)})^T\{\Sigma^{(c)}\}^{-1}(x-\mu^{(c)}) + \log|\Sigma^{(c)}| - 2\log P(\omega^{(c)}) \tag{1}$$

Where $\mu^{(c)}$ and $\Sigma^{(c)}$ denote the mean vector and the covariance matrix for x in the class $\omega^{(c)}$, respectively, and $P(\omega^{(c)})$ is the a priori probability for the class $\omega^{(c)}$. Using the equation $\Sigma_M = \sum_{i=1}^{n}\lambda_i\varphi_i\varphi_i^T$, and omitting the subscript (c) and the term $P(\omega^{(l)})$ for simplicity, the QDF can be written as

$$g_0(x) = \sum_{i=1}^{n}\frac{1}{\lambda_i}\{\varphi_i^T(x-\mu_M)\}^2 + \log\prod_{i=1}^{n}\lambda_i \tag{2}$$

Where μ_M and Σ_M denote the maximum likelihood estimates of the mean and the covariance, respectively, and λ_i ($\lambda_i \ge \lambda_{i+1}$) and φ_i denote the *i*th eigenvalue and the eigenvector of the matrix Σ_M . By employing a kind of pseudo-Bayesian estimate of the covariance matrix $\Sigma_P = \Sigma_M + h^2 I$ instead of the maximum likelihood estimate Σ_M in the QDF, here I is the n-dimensional identity matrix and h^2 is an appropriate constant, an improved discriminant function MQDF1 is given as

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.137

$$g_1(x)=\sum_{i=1}^{n}\frac{1}{\lambda_i+h^2}\{\varphi_i^T(x-\mu_M)\}^2+\log\prod_{i=1}^{n}(\lambda_i+h^2) \tag{3}$$

Another modification of the discriminant function MQDF2 is derived by substituting h^2 for all of the eigenvalues λ_i, $i \geq k+1$ of Σ_M in QDF.

$$g_2(x)=\frac{1}{h^2}\left[\|x-\mu_M\|^2-\sum_{i=1}^{k}(1-\frac{h^2}{\lambda_i})\{\varphi_i^T(x-\mu_M)\}^2\right] + \log(h^{2(n-k)}\prod_{i=1}^{k}(\lambda_i) \tag{4}$$

B. LDA-based Compound Distance

The LDA-based discriminant vector [8, 9] is proposed for improving the accuracy of pair discrimination. The CMF [7] method projects the input feature vector onto an axis on the minor subspace of one class, which is based on the covariance of one class only. To improve the separability of two classes, they turn to estimate the discriminant axis using LDA. In LDA, the projection axis to maximize the Fisher criterion:

$$J(W)=tr((w^TS_Ww)^{-1}(w^TS_Bw)) \tag{5}$$

S_W and S_B denote the within-class scatter matrix and between-class scatter matrix, respectively. For two classes ω_i and ω_j, with means μ_i and μ_j, covariance matrices Σ_i and Σ_j, the within-class and between-class scatter matrices can be written as:

$$S_W=(\Sigma_i+\Sigma_j)/2 \tag{6}$$

$$S_B=(\mu_i-\mu_j)(\mu_i-\mu_j)^T \tag{7}$$

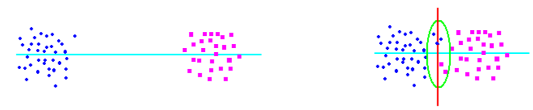
By LDA, the optimal discriminant vector is obtained as:

$$\begin{aligned} w &= S_W^{-1}(\mu_i-\mu_j) \\ &= ((\Sigma_i+\Sigma_j)/2)^{-1}(\mu_i-\mu_j) \\ &= \Sigma^{-1}(\mu_i-\mu_j) \\ &= \sum_{n=1}^{d}\frac{1}{\lambda_n}\varphi_n\varphi_n^T(\mu_i-\mu_j) \end{aligned} \tag{8}$$

Where λ_n and φ_n are the *n*th eigenvalue and eigenvector. The vector w calculated by (8) is then normalized to unit norm: $\tilde{w}=w/\|w\|$. The feature vector of input pattern is projected onto this axis for discriminating two classes.

III. MODIFIED TWO-CLASS LDA BASED COMPOUND DISTANCE

In paper [8, 9], the optimal discriminant vector is estimated by LDA which is proved useful in discriminating similar character pairs. In Fig.1, horizontal lines indicate the directions of the projection vectors learned by LDA. It can be seen that the projection direction is optimal in the whole space. However, as illustrated in Fig.1 (b), some mistakes may occur for the points in the middle of these two classes. This subspace is encircled by the green ellipse. We call this subspace as the Intersecting Subspace (IS). If a projection vector which lies along the vertical red line can be computed, it should be very useful to discriminate points in the IS. This is the motivation of the proposed method in this paper.



(a) (b)

Figure 1. LDA for two classes (a,b) and the intersecting subspace in which misclassification happens by LDA (b)

For two classes x and y, x_t and y_t are the *t*th sample in class x and y respectively, with means μ_x and μ_y, covariance matrices Σ_x and Σ_y, the within-class scatter matrix can be written as:

$$S_W=(\Sigma_x+\Sigma_y)/2 \tag{9}$$

Rather than using (7), the between-class scatter matrix is modified as:

$$S_B^m=\frac{1}{2}[\frac{1}{N}\sum_{t=1}^{N}(x_t-\mu_y)(x_t-\mu_y)^T + \frac{1}{N}\sum_{t=1}^{N}(y_t-\mu_x)(y_t-\mu_x)^T] \tag{10}$$

The modified between-class scatter matrix should supply additional information for discriminating the points in the IS while keeping the information in LDA. Theorem 1 tells us that the principle eigenvector is very close to $(\mu_x-\mu_y)$ which is used by LDA for computing the projection vector.

Theorem 1: *Suppose every $x_{t,i}$ and $y_{t,i}$ satisfy :*

$$\frac{\|x_{t,i}-\mu_{x,i}\|}{\|\mu_{x,i}-\mu_{y,i}\|}\to 0 \quad and \quad \frac{\|y_{t,i}-\mu_{y,i}\|}{\|\mu_{x,i}-\mu_{y,i}\|}\to 0$$

And suppose S_B^m can be rewritten as:

$$S_B^m=\sum_{i=0}^{n}\lambda_i\varphi_i\varphi_i^T,\ \lambda_0>>\lambda_1>...>\lambda_n,$$

Then we have:

$$\sqrt{\lambda_0}\varphi_0\to(\mu_x-\mu_y)$$

In theorem 1, "$\rightarrow$" means "be close to", and $x_{t,i}$, $y_{t,i}$, $\mu_{x,i}$ and $\mu_{y,i}$ represent the *i*th element in x_t, y_t, μ_x and μ_y respectively. $S_B(r,c)$ and $S_B^m(r,c)$ represent the element lying in the *r*th row and *c*th column of S_B and S_B^m respectively.

Proof: S_B can be written as:

$$\begin{aligned} S_B &= (\mu_x-\mu_y)(\mu_x-\mu_y)^T \\ &= \frac{1}{2}[(\frac{1}{N}\sum_{t=1}^{N}(x_t-\mu_y))(\frac{1}{N}\sum_{t=1}^{N}(x_t-\mu_y))^T \\ &\quad +(\frac{1}{N}\sum_{t=1}^{N}(y_t-\mu_x))(\frac{1}{N}\sum_{t=1}^{N}(y_t-\mu_x))^T] \end{aligned}$$

The element lying in the *i*th row and *j*th column of S_B is:

$$\begin{aligned} S_B(i,j) &= \frac{1}{2}\{\frac{1}{N}\sum_{t=1}^{N}[(x_{t,i}-\mu_{y,i})(\mu_{x,j}-\mu_{y,j})] \\ &\quad +\frac{1}{N}\sum_{t=1}^{N}[(y_{t,i}-\mu_{x,i})(\mu_{y,j}-\mu_{x,j})]\} \end{aligned}$$

Similarly:

$$\begin{aligned} S_B^m(i,j) &= \frac{1}{2}\{\frac{1}{N}\sum_{t=1}^{N}[(x_{t,i}-\mu_{y,i})(x_{t,j}-\mu_{y,j})] \\ &\quad +\frac{1}{N}\sum_{t=1}^{N}[(y_{t,i}-\mu_{x,i})(y_{t,j}-\mu_{x,j})]\} \end{aligned}$$

Since $\forall x_{t,i}$ and $\forall y_{t,i}$, we have

$$\frac{\|x_{t,i}-\mu_{x,i}\|}{\|\mu_{x,i}-\mu_{y,i}\|}\rightarrow 0 \qquad \frac{\|y_{t,i}-\mu_{y,i}\|}{\|\mu_{x,i}-\mu_{y,i}\|}\rightarrow 0$$

Then we get

$$\begin{aligned} x_{t,j}-\mu_{y,j} &= x_{t,j}-\mu_{x,j}+\mu_{x,j}-\mu_{y,j}\rightarrow \mu_{x,j}-\mu_{y,j} \\ y_{t,j}-\mu_{x,j} &= y_{t,j}-\mu_{y,j}+\mu_{y,j}-\mu_{x,j}\rightarrow \mu_{y,j}-\mu_{x,j} \end{aligned}$$

Then

$$S_B^m(i,j)\rightarrow S_B(i,j)$$

As it is assumed that

$$S_B^m=\sum_{i=0}^{n}\lambda_i\varphi_i\varphi_i^T,\ \lambda_0>>\lambda_1>...>\lambda_n$$

Let $S_0=\lambda_0\varphi_0\varphi_0^T$, then

$$S_0(i,j)\rightarrow S_B^m(i,j)\rightarrow S_B(i,j).$$

At last we have

$$\sqrt{\lambda_0}\varphi_0\rightarrow(\mu_x-\mu_y)$$

Theorem 1 indicates that the direction of $S_W^{-1}(\sqrt{\lambda_0}\varphi_0)$ is close to the optimal direction $S_W^{-1}(\mu_x-\mu_y)$. This result will be demonstrated in our experiments, too. We hope that directions along with $S_W^{-1}\varphi_i, i=1,2,...,n$ can help us to discriminate points in the IS better. As we can see in our experiments, they really can supply additional information for discriminating points in the IS.

IV. IMPLEMENTATION NOTES

For classes ω_i and ω_j, we calculate the S_W and S_B^m by (9) and (10) respectively. S_W and S_B^m can be rewritten as

$$S_W=\sum_{t=0}^{n}\gamma_t\phi_t\phi_t^T \qquad S_B^m=\sum_{t=0}^{n}\lambda_t\varphi_t\varphi_t^T \tag{11}$$

Where γ_i and ϕ_i are the *i*th eigenvalue and eigenvector of S_W and λ_i and φ_i are the *i*th eigenvalue and eigenvector of S_B^m. The principle projection vector w_0 is computed by

$$w_0=\sum_{t=0}^{d}{}^{1}\!/\!_{\gamma_t}\,\phi_t\phi_t^T\varphi_0 \tag{12}$$

The minor projection vectors are computed by

$$w_s=\sum_{t=0}^{d}{}^{1}\!/\!_{\gamma_t}\,\phi_t\phi_t^T\varphi_s,\ s=1,...,T_p \tag{13}$$

w_0 and w_s are then normalized to unit norm:

$$\tilde{w}_0=w_0/\|w_0\| \qquad \tilde{w}_s=w_s/\|w_s\| \tag{14}$$

Define the distance function $f_0(x,\omega_i)$ as:

$$f_0(x,\omega_i)=(\hat{x}_0-\hat{\mu}_{i,0})^2/\hat{\delta}_i^2 \tag{15}$$

Where

$$\hat{x}_0=x^T\tilde{w}_0 \qquad \hat{\mu}_{i,0}=\mu_i^T\tilde{w}_0 \tag{16}$$

And $\widehat{\delta}_i^2$ is the variance in one-dimensional projected subspace for class ω_i.
If the following formula satisfies, then x belongs to the IS between classes ω_i and ω_j.

$$\frac{|f_0(x,\omega_i)-f_0(x,\omega_j)|}{|f_0(\omega_i,\omega_j)|}<\theta \tag{17}$$

Where θ is a threshold. The compound distance functions for classes ω_i and ω_j are computed by

$$\begin{cases} f(x,\omega_i)=(1-\beta)*f_{mqdf}(x,\omega_i)+\beta*f_{m2lda}(x,\omega_i) \\ f(x,\omega_j)=(1-\beta)*f_{mqdf}(x,\omega_j)+\beta*f_{m2lda}(x,\omega_j) \end{cases} \tag{18}$$

If $f(x,\omega_i)<f(x,\omega_j)$ then x belongs to class ω_i, otherwise x belongs to class ω_j.
The distance function $f_{m2lda}(x,\omega_i)$ is computed by:

$$f_{m2lda}(x,\omega_i)=\begin{cases} (1-\alpha)*f_0(x,\omega_i)+\alpha*f_\xi(x,\omega_i), x\in IS \\ f_0(x,\omega_i), \text{otherwise} \end{cases} \tag{19}$$

The distance function $f_\xi(x,\omega_i)$ is computed by:

$$f_\xi(x,\omega_i)=\sum_{t=1}^{\xi}(\widehat{x}_t-\widehat{\mu}_{i,t})^2 \tag{20}$$

Where ξ is the number of minor projection directions used for discriminating points in the IS and

$$\widehat{x}_t=x^T\tilde{w}_t \quad \widehat{\mu}_{i,t}=\mu_i^T\tilde{w}_t \tag{21}$$

V. Experiment Results

We evaluate our methods on the CASIA database. The CASIA database, which is collected by the institute of automation, Chinese academy of sciences, contains 3755 Chinese characters, 300 samples per class. We choose 290 samples per class for training and the remaining 10 samples for testing. We use the line-density normalization method in character image normalization step [11]. Then 8-direction gradient direction features [12] are extracted on the normalized image. The extracted feature vector is compressed to 256 dimensions by LDA. Just for experiment, we select 3849 similar pairs by the error rate on the training set using the linear discriminant function (LDF) classifier. S_p will be used to represent the similar pair set.

For each input pattern x, the MQDF gives two top-rank candidate classes ω_i and ω_j. The MQDF distances from x to ω_i and ω_j are d_i and d_j, respectively. If $|d_i-d_j|\le T_c$, where T_c is a threshold, then we check if ω_i and ω_j is a similar pair in set S_p, if it is then we discriminate ω_i and ω_j with the LDA-based methods proposed in [8, 9] and the method proposed in this paper. Table 2 shows the number of similar pairs with different T_c.

TABLE I. Numbers of Similar Pairs With Different Threshold Tc

T_c	20	40	60	80	100
N	867	1559	2219	2974	3731

Table 2 compares some elements in the projection vector w computed by (8) and the projection vector w_0 computed by (12) between class 0 and class 2391. Table 3 shows the test accuracies using w and w_0 as the projection vector with different T_c. From table 2 and table 3, we can see that w and w_0 are very close both in elements and test accuracies. At this point, theorem 1 is proved by experiment results..

TABLE II. Some Elements in the Projection Vector w and w0.

Index	1	2	…	255	256
w	-0.197588	0.853617	…	0.122322	-0.098102
w_0	0.197768	-0.857342		-0.121990	0.097299

TABLE III. Test Accuracies Using w and w0 With Dfferent Tc

Tc	20	40	60	80	100
w	0. 971931	0. 974993	975393	0. 975526	0. 975366
w_0	0. 971957	0. 974967	975393	0. 975499	0. 975313

Fig.2 shows the test accuracies of the proposed method with different θ in (17) ($\alpha=0.7,\beta=0.8,T_c=60$). We can see that the best performance reaches when θ =0.2 for different ξ.

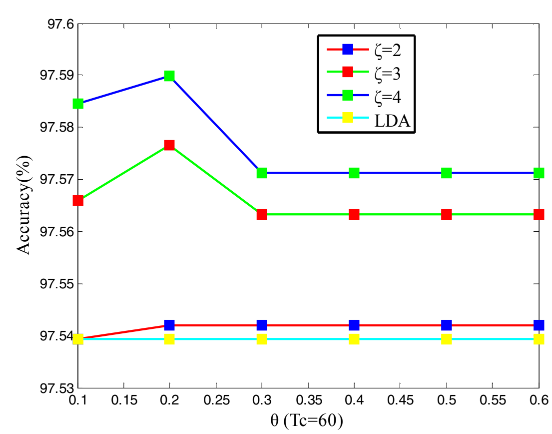


Figure 2. Test accuracies of the proposed method with different θ

Fig.3 shows the test accuracies of the proposed method with different ξ in (20) ($\alpha=0.7, \beta=0.8, T_c=60$) . We can see that test accuracies do not increase too much when ξ =4 for θ =0.2 and ξ =6 for θ =0.3.

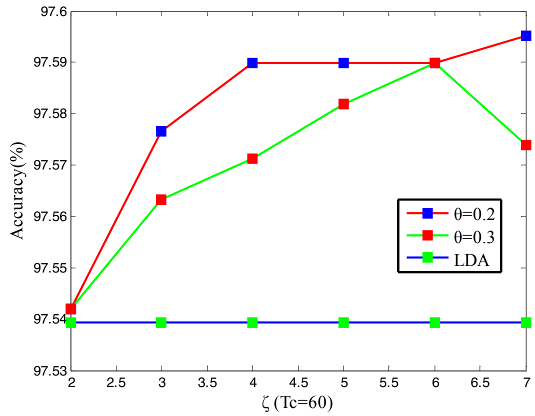


Figure 3. Test accuracies of the proposed method with different ξ

Fig.4 shows the test accuracies of the LDA based method and the proposed method (denoted by M2LDA) with different T_c for θ =0.2 and ξ =4. We can see that the proposed method outperforms the previous LDA based method.

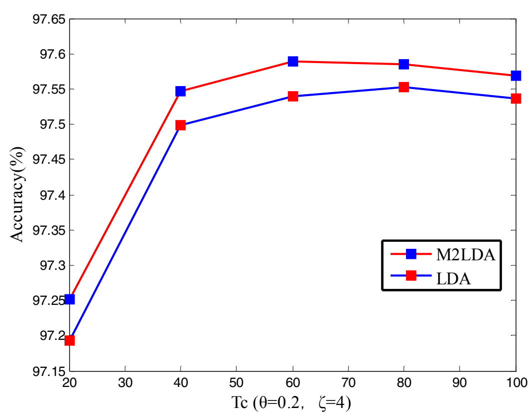


Figure 4. Test accuracies of the LDA-based method and the proposed method with different Tc.

VI. CONCLUSIONS

This paper proposes a modified two-class LDA based compound distance for similar handwritten Chinese characters discrimination. The modified between-class scatter matrix supplies additional information for discriminate points in the intersecting subspace. Our experimental results demonstrate that the proposed method outperforms the previous LDA-based method. In our future work, how to use the additional information more efficiently will be concerned.

ACKNOWLEDGMENT

This work is supported by National Science Foundation of China (NSFC) under grants no.60802055, no.60835001 and no.60933010.

REFERENCES

[1] J. Tsukumo, H. Tanaka, Classification of hand printed Chinese characters using non-linear normalization and correlation methods, in: Proceedings of the Ninth International Conference on Pattern Recognition, Roma, Italy, 1988, pp. 168–171.

[2] C.-L. Liu and K. Marukawa, Pseudo two dimensional shape normalization methods for handwritten Chinese character recognition, Pattern Recognition 38 (12) (2005), pp. 2242–2255.

[3] C.-L. Liu, High accuracy handwritten Chinese character recognition using quadratic classifiers with discriminative feature extraction, in: Proceedings of the 18th International Conference on Pattern Recognition, vol. 2, Hong Kong, 2006, pp. 942–945.

[4] K.-C. Leung, C.-H. Leung, Recognition of handwritten Chinese characters by combining regularization, Fisher's discriminant and distorted sample generation, in: Proceedings of the 10th International Conference on Document Analysis and Recognition, Barcelona, 2009, pp. 1026–1030.

[5] K.C. Leung, C.H. Leung, Recognition of handwritten Chinese characters by critical region analysis, Pattern Recognition, Volume 43, Issue 3, March 2010, Pages 949-961, ISSN 0031-3203, DOI: 10.1016/j.patcog.2009.09.001.

[6] Bo Xu, Kaizhu Huang, Cheng-Lin Liu, "Similar Handwritten Chinese Characters Recognition by Critical Region Selection Based on Average Symmetric Uncertainty," icfhr, pp.527-532, 2010 12th International Conference on Frontiers in Handwriting Recognition, 2010

[7] M. Suzuki, S. Ohmachi, N. Kato, H. Aso and Y. Nemoto, A discrimination method of similar characters using compound Mahalanobis function, Trans. IEICE Jpn J80-D-II (10) (1997), pp. 2752–2760.

[8] T.-F. Gao, C.-L. Liu, LDA-based compound distance for handwritten Chinese character recognition, in: Proceedings of the 9th ICDAR, Curitiba, Brazil, 2007, pp. 904–909.

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

[10] F. Kimura, K. Takashina, S. Tsuruoka and Y. Miyake, Modified quadratic discriminant functions and the application to Chinese character recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 9 (1) (1987), pp. 149–153.

[11] Takahiko Horiuchi , Ryoji Haruki , Hiromitsu Yamada , Kazuhiko Yamamoto, Two-dimensional Extension of Nonlinear Normalization Method using Line Density for Character Recognition, Proceedings of the 4th International Conference on Document Analysis and Recognition, p.511-514, August 18-20, 1997 .

[12] Cheng-Lin Liu, Kazuki Nakashima, Hiroshi Sako, Hiromichi Fujisawa, Handwritten digit recognition: investigation of normalization and feature extraction techniques, Pattern Recognition, Volume 37, Issue 2, February 2004, Pages 265-279, ISSN 0031-3203.