Locally Smoothed Modified Quadratic Discriminant Function

Xu-Yao Zhang Cheng-Lin Liu
National Laboratory of Pattern Recognition (NLPR)
Institute of Automation, Chinese Academy of Sciences, Beijing, P.R. China
xyz@nlpr.ia.ac.cn liucl@nlpr.ia.ac.cn

***Abstract*—Modified quadratic discriminant function (MQDF) is a state-of-the-art classifier for handwriting recognition. However, the big gap between accuracies on training and testing sets indicates that MQDF has a good capability to fit training data but the generalization performance is not promising. To solve this problem, we propose a new model called locally smoothed modified quadratic discriminant function (LSMQDF) by smoothing the covariance matrix of each class with its nearest neighbor classes. LSMQDF can be viewed as a regularization to avoid over-fitting. The covariance matrix estimated by local smoothing is more accurate and robust. LSMQDF can be also viewed as an extension of the global smoothing method, namely regularized discriminant analysis (RDA). Experiments on both offline and online Chinese handwriting databases demonstrate that: with local smoothing, the accuracy on training set is decreased (over-fitting avoided), and the accuracy on testing set is improved significantly and consistently (generalization improved).**

I. Introduction

The modified quadratic discriminant function (MQDF) [1] has been applied prevalently and successfully in handwritten Chinese character recognition [2] [3] over the past 25 years. However, the large number of free parameters of MQDF usually lead to over-fitting of the training data. To solve this problem, this paper proposes a local smoothing technique as a regularization to obtain a more robust and accurate estimation of the covariance matrix. The local smoothing can alleviate the over-fitting problem and improve the generalization performance of MQDF.

A. Quadratic Discriminant Function (QDF)

In Bayes decision theory, the quadratic discriminant function (QDF) is derived from the assumption of class-conditional Gaussian distribution

$$p(x|i) = \frac{\exp\left[-\frac{1}{2}(x-\mu_i)^\top \Sigma_i^{-1}(x-\mu_i)\right]}{(2\pi)^{\frac{d}{2}}|\Sigma_i|^{\frac{1}{2}}}, \tag{1}$$

where $\mu_i \in \mathbb{R}^d$ and $\Sigma_i \in \mathbb{R}^{d\times d}$ are the mean vector and covariance matrix of class i respectively. Considering a M-class classification problem, a pattern x is classified into the class of maximum a posterior (MAP) probability $x \in \arg\max_{i=1}^M p(i|x) = \frac{p(i)p(x|i)}{p(x)}$, where $p(i)$ is the prior probability and $p(x)$ is the mixture density function. When assuming equal prior probabilities, the MAP decision rule becomes $x \in \arg\max p(x|i)$ which is equivalent to $x \in \arg\min -\log p(x|i)$. Therefore QDF is defined as:

$$f_{\mathrm{QDF}}(x,i) = (x-\mu_i)^\top \Sigma_i^{-1}(x-\mu_i) + \log|\Sigma_i|, \tag{2}$$

which is actually a distance metric between x and class i:

$$x \in \text{class } \arg\min_{i=1}^{M} f_{\mathrm{QDF}}(x,i). \tag{3}$$

The performance of QDF is highly dependent on the computation of the inverse matrix Σ_i^{-1}. The covariance matrix Σ_i is usually singular and under-estimated due to the estimation errors and the sensitivity of the estimation to small sample size. Therefore QDF can not achieve high accuracy in real applications. To achieve an accurate and robust estimation of Σ_i^{-1}, many improvements have been proposed, among which the modified quadratic discriminant function (MQDF) [1] is the most effective one for large category Chinese handwriting recognition.

B. Modified Quadratic Discriminant Function (MQDF)

By utilizing eigen-decomposition algorithms, the covariance matrix can be diagonalized as $\Sigma_i = \Phi_i\Lambda_i\Phi_i^\top$, where $\Lambda_i = \mathrm{diag}[\lambda_{i1},\cdots,\lambda_{id}]$ with $\lambda_{ij} \in \mathbb{R}^+, j = 1,\cdots,d$ being the eigenvalues (ordered in decreasing order) of Σ_i, and $\Phi_i = [\phi_{i1},\cdots,\phi_{id}]$ with $\phi_{ij} \in \mathbb{R}^d, j = 1,\cdots,d$ being the ordered eigenvectors. Inserting the eigen-decomposed covariance matrix into QDF, we get $f_{\mathrm{QDF}}(x,i) = \sum_{j=1}^d \frac{1}{\lambda_{ij}}[\phi_{ij}^\top(x-\mu_i)]^2 + \sum_{j=1}^d \log\lambda_{ij}$. The minor eigenvalues are usually prone to be underestimated due to small sample size, therefore MQDF [1] replaces the minor eigenvalues ($\lambda_{ij}, j > k$) with a constant δ_i to stabilize the generalization performance

$$\begin{aligned} f_{\mathrm{MQDF}}(x,i) = &\sum_{j=1}^{k}\left(\frac{1}{\lambda_{ij}} - \frac{1}{\delta_i}\right)\left[\phi_{ij}^\top(x-\mu_i)\right]^2 \\ &+ \frac{1}{\delta_i}||x-\mu_i||^2 + \sum_{j=1}^{k}\log\lambda_{ij} + (d-k)\log\delta_i, \end{aligned} \tag{4}$$

where k denotes the number of principal axes. The above derivations utilize the properties of $\Phi_i^\top\Phi_i = I$ and $||x-\mu_i||^2 = \sum_{j=1}^d[\phi_{ij}^\top(x-\mu_i)]^2$. Compared with QDF, MQDF involves only the principal eigenvectors and eigenvalues,

1520-5363/13 $26.00 © 2013 IEEE
DOI 10.1109/ICDAR.2013.11

therefore, both the storage of parameters and the computational complexity are reduced. More importantly, higher classification accuracy can be obtained by setting the minor eigenvalues to be a class-independent constant as $\delta_1 = \delta_2 = \cdots = \delta_M = \alpha \frac{1}{Md} \sum_{i=1}^{M} \sum_{j=1}^{d} \lambda_{ij}$ and using cross-validation to select α from $[0, 1]$ on the training data.

Although MQDF is a state-of-the-art classifier for Chinese handwriting recognition, many improvements have been proposed in the past 25 years. We first give a summarization of the improvements on MQDF in Section II. After that, the motivation and definition of the newly proposed locally smoothed MQDF are described in Section III. The experimental results on both offline and online Chinese handwriting datasets are presented in Section IV, and the concluding remarks are drawn in Section V.

II. Related Works

Many improvements have been proposed for MQDF from different aspects, such as discriminative training, memory reduction, instance selection and generation, ensemble learning, and discriminative feature extraction.

Because MQDF is a generative model, Liu et al. [2] proposed to use the minimum classification error (MCE) [4] criterion for discriminative training of MQDF. After that, the sample separation margin criterion [5] and perceptron criterion [6] were also adopted for this purpose. The retraining of MQDF based on instance importance weighting [7], instance selection [8] and virtual instance generation (mirror image) [9] can be also viewed as some kinds of discriminative training. Discriminative training of MQDF can directly optimize the classification boundaries, therefore, high classification accuracy can be achieved especially when the size of training data is large.

The memory requirement of MQDF is usually very large (e.g. 120MB for a 3,755-class and 160-dimension problem). To reduce the memory requirement of MQDF, Long and Jin [10] proposed to use the vector quantization and split quantization techniques to compress the parameters (means, eigenvectors and eigenvalues) of MQDF. Modeling the inverse covariance matrices by the expansion of some tied basis matrices was proposed by Wang and Huo [11]. The precision constrained Gaussian model was combined with the minimum classification error (MCE) training criterion to simultaneously compress the parameters and improve the accuracy [12]. With parameter compression, the high accuracy MQDF classifier can be embedded into some handheld devices such as mobile phones and tablet computers.

There are also many other improvements of MQDF derived from different viewpoints. The kernel MQDF was proposed by Yang and Jin [13] to extend MQDF from original feature space to kernel space (implicit high-dimensional space). The ensemble learning methods such as cascade classifier training [14], boosting [15] and pairwise discrimination [16] were adopted to improve the accuracy of MQDF. The graphical lasso method was used to estimate a sparse inverse covariance matrix Σ_i^{-1} for QDF [17]. The normalization of the determinant of the covariance matrix was shown to achieve better classification accuracy [18]. The asymmetric Mahalanobis distance [19] can be viewed as an extension of the symmetric Gaussian distribution. Combining MQDF with the discriminative feature extraction [20] can further improve the performance.

III. Locally Smoothed MQDF

Although MQDF is a generative model, it has a very high accuracy on training set (over 99%). However, the accuracy on testing set is much lower (around 89% and 93% for offline and online Chinese handwriting recognition). This indicates that MQDF has a good capability to fit training data but the generalization performance is not promising.[1]

The covariance matrix estimated by maximum likelihood (ML) is usually under-estimated due to small sample size, and the large number of free parameters gives MQDF a strong memory of training samples (over-fitting). To improve the generalization performance of MQDF, we propose a new model of locally smoothed modified quadratic discriminant function (LSMQDF) by smoothing the covariance matrix of each class with its nearest neighbor classes. In the following sections, we first give a description of the ML estimation, and then present the details of LSMQDF. After that, we compare LSMQDF with a global smoothing method.

A. Class-wise ML Estimation

Given a training dataset $\{x_j^i \in \mathbb{R}^d\}$ ($i = 1, \cdots, M$ and $j = 1, \cdots, n_i$), where n_i is the number of training samples in class i, and M is the number of classes. Let x_j^i denotes the j-th training sample from class i. The function of the negative log-likelihood of the data from class i w.r.t. the mean μ and the covariance matrix Σ can be defined as:

$$\begin{aligned} \mathcal{NLL}(\mu, \Sigma, i) &= -\sum_{j=1}^{n_i} \log p(x_j^i | i) \\ &\propto \sum_{j=1}^{n_i} (x_j^i - \mu)^\top \Sigma^{-1} (x_j^i - \mu) + n_i \log |\Sigma| \, . \end{aligned} \tag{5}$$

The maximum likelihood (ML) estimation of μ_i and Σ_i is conducted class-by-class ($i = 1, \cdots, M$) as:

$$\{\mu_i, \Sigma_i\} = \arg\min_{\mu, \Sigma} \ \mathcal{NLL}(\mu, \Sigma, i) \, . \tag{6}$$

By solving this convex optimization problem, the ML estimation is:

$$\mu_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j^i \, , \tag{7}$$

$$\Sigma_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_j^i - \mu_i)(x_j^i - \mu_i)^\top \, . \tag{8}$$

[1]This explains why discriminative training of MQDF can only improve the testing accuracy slightly.

B. Locally Smoothed MQDF

The class-wise ML estimation is proven to be sensitive to small sample size (small n_i), and the estimation error in covariance matrix (8) is more serious than the estimation error in mean vector (7). To achieve a robust estimation, we propose a local smoothing method:

$$\widetilde{\Sigma_i} = \arg\min_{\Sigma} \ (1-\beta)\, \mathcal{NLL}(\mu_i, \Sigma, i) \ + \beta \frac{1}{K} \sum_{j \in \mathrm{KNN}(i)} \mathcal{NLL}(\mu_j, \Sigma, j)\,, \tag{9}$$

where $\mathrm{KNN}(i)$ denotes the K nearest neighbors of class i from the remaining classes (defined by the Euclidean distance between the class means). The first term of (9) is the ML estimation and the second term is a local smoothing of the covariance matrix with the nearest neighbor classes. The $\beta \in [0,1]$ is a tradeoff parameter. Because the estimation of the class-mean is accurate enough, μ_i is fixed as the ML estimation (7). The new covariance matrix $\widetilde{\Sigma_i}$ is estimated to maximize the likelihood of not only class i but also the neighboring classes. Therefore, both the samples from class i and the neighboring classes are used simultaneously to achieve a robust estimation. This is based on the assumption that the covariance matrices of neighboring classes are close to each other, which is satisfied in handwriting recognition, because similar characters usually have the same types of distortion (variation).

The optimization problem of (9) is also a convex problem which has a closed-form solution:

$$\widetilde{\Sigma_i} = \frac{(1-\beta) n_i \Sigma_i + \beta \frac{1}{K} \sum_{j \in \mathrm{KNN}(i)} n_j \Sigma_j}{(1-\beta) n_i + \beta \frac{1}{K} \sum_{j \in \mathrm{KNN}(i)} n_j}\,, \tag{10}$$

where Σ_i is the ML estimation (8). From this definition, we can find that $\widetilde{\Sigma_i}$ is a smoothing of Σ_i with the neighboring classes $\Sigma_j : j \in \mathrm{KNN}(i)$. Therefore, $\widetilde{\Sigma_i}$ should be more accurate than Σ_i especially when n_i is small. By using μ_i (7) and $\widetilde{\Sigma_i}$ (10) to train the MQDF classifier (Section I-B), much better generalization performance can be achieved, and we denote this method as LSMQDF. With local smoothing, some discriminative information of the training data is lost, but the new covariance matrix $\widetilde{\Sigma_i}$ will be more robust. Therefore, although the accuracy on training set is decreased (over-fitting avoided), the accuracy on testing set will be improved (generalization improved).

C. Local Smoothing V.S. Global Smoothing

LSMQDF can be viewed as an extension of the global smoothing method, namely regularized discriminant analysis (RDA) [21], which constrains the range of parameter values by interpolating the class covariance matrix with the common covariance matrix Σ_0 and the identity matrix I:

$$\widehat{\Sigma_i} = (1-\gamma)[(1-\beta)\Sigma_i + \beta \Sigma_0] + \gamma \delta_i^2 I, \tag{11}$$

where $\Sigma_0 = (\sum_{i=1}^{M} n_i \Sigma_i)/(\sum_{i=1}^{M} n_i), \delta_i^2 = \frac{1}{d}\mathrm{tr}(\Sigma_i)$, and $0 \leq \beta, \gamma \leq 1$. With appropriate values of β and γ empirically selected, RDA can improve the generalization performance, and some special cases are included: (i) original QDF: $\beta = \gamma = 0$; (ii) linear discriminant function (LDF): $\gamma = 0, \beta = 1$; and (iii) Euclidean distance: $\gamma = 1$. Improvements can be achieved when using μ_i (7) and $\widehat{\Sigma_i}$ (11) to train the MQDF classifier (Section I-B), and we still denote this method as RDA for simplification.

RDA is a global smoothing method, while LSMQDF is a local smoothing method. When the number of classes is large, global smoothing can not achieve any improvement, while local smoothing is still effective to improve the accuracy. This is because we can not assume all the covariance matrices to be close to each other, and contrarily, the neighboring classes are more likely to have the same covariance matrices. The superior performance of LSMQDF is also verified by experiments in the following sections.

IV. Experiments

We evaluated the performance of LSMQDF on two 3,755-class Chinese handwriting databases [22]: the offline handwriting database CASIA-HWDB1.1 and the online handwriting database CASIA-OLHWDB1.1. Both of them contain handwritten Chinese characters from 300 writers (240 for training and 60 for testing). Each writer has about 3,755 characters (one for each class). The extracted feature data can be downloaded from our website.[2]

A. LSMQDF for Offline Handwriting Recognition

For representing an offline character sample, we extract features from gray-scale character images (back-ground eliminated) using the normalization-cooperated gradient feature (NCGF) method [23]. The original feature dimensionality is 512 which is further reduced to 160 by Fisher linear discriminant analysis (FDA).

For the LSMQDF method, we set $K = 10$ and $\beta = 0.5$ for (10). Sensitive analysis of these parameters will be shown in the following sections. For the RDA method (11), we have $\Sigma_0 = I$ (the common covariance matrix is I in the FDA transformed subspace), and now the RDA can be simplified as $\widehat{\Sigma_i} = (1-\gamma)\Sigma_i + \gamma \delta_i^2 I$ and we report the best performance of RDA on testing set by searching γ from 0 to 1.

The accuracies of MQDF, RDA and LSMQDF on both training set and testing set with various numbers of principal eigenvectors are shown in Figure 1(a). We can find that: (i) with either global smoothing (RDA) or local smoothing (LSMQDF), the training accuracy is decreased (over-fitting avoided), and the testing accuracy is improved (generalization improved); (ii) LSMQDF outperforms RDA consistently and significantly on the testing set; and (iii) compared with the baseline MQDF, LSMQDF is effective to improve the

[2] http://www.nlpr.ia.ac.cn/databases/handwriting/Download.html

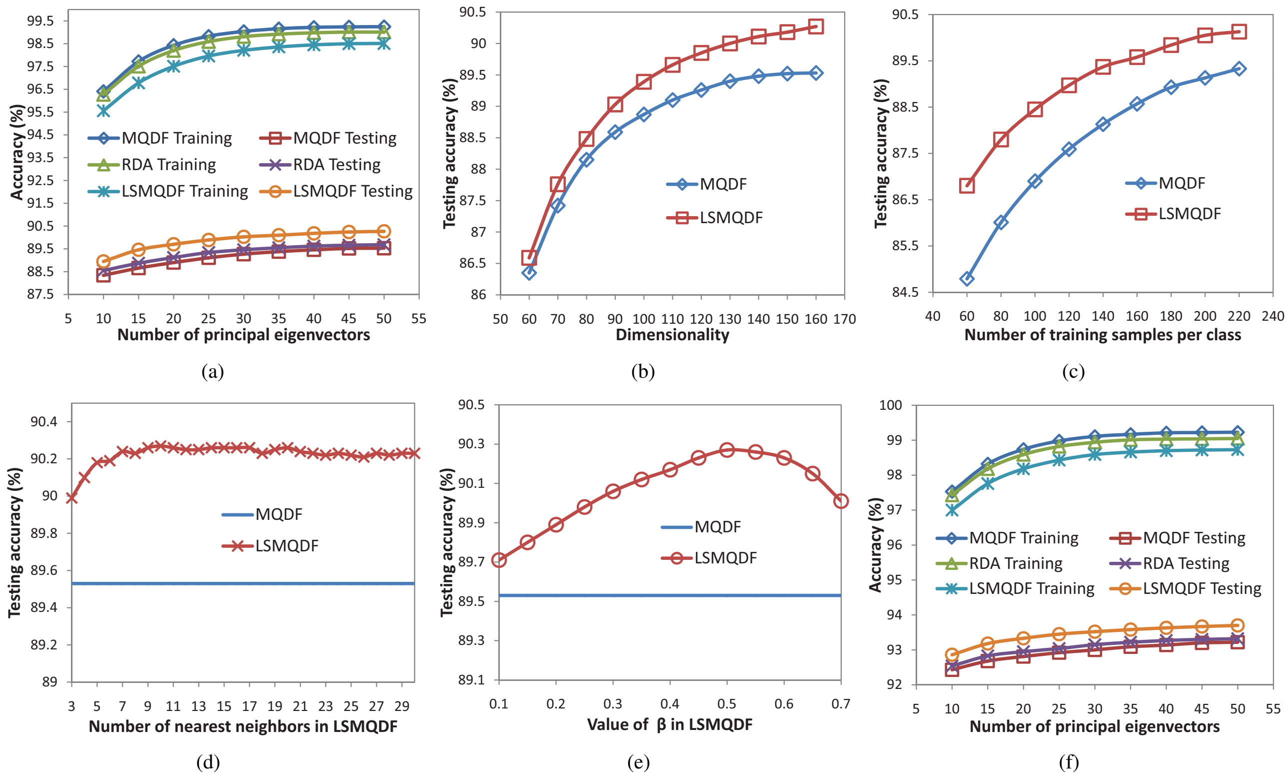

Figure 1. (a) Different models for offline handwriting recognition; (b) MQDF v.s. LSMQDF on different dimensions; (c) MQDF v.s. LSMQDF with different number of training samples; (d) LSMQDF with different numbers of K in (10); (e) LSMQDF with different values of β in (10); (f) Different models for online handwriting recognition.

testing accuracy, e.g., from 89.53% to 90.27% when the number of principal eigenvectors is 50.

B. Varying Dimensionality

By using FDA to reduce the dimensionality from 512 to different subspaces, we compare the performance of MQDF and LSMQDF ($K = 10$ and $\beta = 0.5$ for (10) and the number of principal eigenvectors is 50). The results are shown in Figure 1(b). When the dimensionality is small, the difference between MQDF and LSMQDF is not significant, but with the increasing of dimensionality, the difference becomes larger and larger. This is because the curse of dimensionality, i.e., in the high-dimensional space, the training data are not sufficient to get an accurate parameter estimation, and in this case, smoothing is an important tool for improving the generalization performance.

C. Varying the Size of Training Data

We also compare the performance of MQDF and LSMQDF with different number of training samples. In this case, we fix the dimensionality as 160, use $K = 10, \beta = 0.5$ for (10), and set the number of principal eigenvectors as 50. The results are shown in Figure 1(c). We can find that: LSMQDF can improve the testing accuracy consistently and significantly, especially when the number of training samples is small (e.g. the accuracy is improved from 84.79% to 86.80% when there are only 60 training samples per class).

D. On Selection of K

In this section, we evaluate the performance of LSMQDF by varying the number of nearest neighbors K in (10). Other parameters are fixed as the same of Section IV-C. The results are shown in Figure 1(d). When $K = 3$, the testing accuracy of LSMQDF is already higher than MQDF, and with the increasing of K, the performance is further improved. When $10 \leq K \leq 30$, the accuracy is nearly not changed. Therefore, the performance of LSMQDF is not sensitive to the selection of K.

E. On Selection of β

We also evaluate the performance of LSMQDF by varying the values of the tradeoff parameter β in (10). Other parameters are fixed as the same of Section IV-C. The results are shown in Figure 1(e). When $\beta = 0.1$, the accuracy of LSMQDF is higher than MQDF, because the information of neighboring classes is incorporated into the estimation of the covariance matrix. With the increasing of β, the performance is further improved. However, when $\beta > 0.5$, the accuracy is decreased due to the over-smoothing. Therefore, in practice, we should pay attention to the value of β, and $\beta = 0.5$ is shown to be a good choice in our experiments.

F. LSMQDF for Online Handwriting Recognition

In this section, we use LSMQDF for online handwriting recognition. For representing an online character sample, we use a benchmark feature extraction method [24]: 8-direction histogram feature extraction combined with pseudo 2D bi-moment normalization (P2DBMN). We also add the direction values of off-strokes (pen lifts) to real strokes with a weight of 0.5 [25]. The settings of different parameters are the same as Section IV-A.

The results are shown in Figure 1(f). Although MQDF can achieve much higher testing accuracy for online data (compared with offline data), LSMQDF is still effective in improving the generalization performance, e.g., from 93.22% to 93.70% when the number of principal eigenvectors is 50. The local smoothing of LSMQDF again outperforms the global smoothing method of RDA.

V. Conclusion

To improve the generalization performance of MQDF, we proposed a locally smoothed modified quadratic discriminant function (LSMQDF) by smoothing the covariance matrix of each class with its neighboring classes. The idea of local smoothing is simple and effective. In the future, we will explore the local smoothing from the theoretical viewpoint such as Bayesian learning [26]. Using the parameters of LSMQDF as initialization for discriminative training [2], and combining LSMQDF with improved dimensionality reduction methods [27] can further improve the accuracy.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (NSFC) Grant 60933010 and National Basic Research Program of China (973 Program) Grant 2012CB316302.

References

[1] F. Kimura, K. Takashina, S. Tsuruoka, and Y. Miyake, "Modified quadratic discriminant functions and the application to Chinese character recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1987.

[2] C.-L. Liu, H. Sako, and H. Fujisawa, "Discriminative learning quadratic discriminant function for handwriting recognition," *IEEE Trans. Neural Networks*, 2004.

[3] X.-Y. Zhang and C.-L. Liu, "Writer adaptation with style transfer mapping," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2013.

[4] B.-H. Juang and S. Katagiri, "Discriminative learning for minimum error classification," *IEEE Trans. Signal Processing*, 1992.

[5] Y. Wang and Q. Huo, "Sample separation margin based minimum classification error training of pattern classifiers with quadratic discriminant functions," *IEEE Int'l Conf. Acoustics Speech and Signal Processing*, 2010.

[6] T.-H. Su, C.-L. Liu, and X.-Y. Zhang, "Perceptron learning of modified quadratic discriminant function," *Int'l Conf. Document Analysis and Recognition*, 2011.

[7] Y. Wang, X. Ding, and C. Liu, "MQDF discriminative learning based offline handwritten Chinese character recognition," *Int'l Conf. Document Analysis and Recognition*, 2011.

[8] Y. Wang, X. Ding, and C. Liu, "MQDF retrained on selected sample set," *IEICE Trans. Information and Systems*, 2011.

[9] H. Liu and X. Ding, "Handwritten Chinese character recognition based on mirror image learning and the compound Mahalanobis function," *J. Tsinghua Univ. (Sci. & Tech.) (in Chinese)*, 2006.

[10] T. Long and L. Jin, "Building compact MQDF classifier for large character set recognition by subspace distribution sharing," *Pattern Recognition*, 2008.

[11] Y. Wang and Q. Huo, "Modeling inverse covariance matrices by expansion of tied basis matrices for online handwritten Chinese character recognition," *Pattern Recognition*, 2009.

[12] Y. Wang and Q. Huo, "Building compact recognizers of handwritten Chinese characters using precision constrained Gaussian model, minimum classification error training and parameter compression," *Int'l J. Document Analysis and Recognition*, 2011.

[13] D. Yang and L. Jin, "Kernel modified quadratic discriminant function for online handwritten Chinese characters recognition," *Int'l Conf. Document Analysis and Recognition*, 2007.

[14] Q. Fu, X. Ding, and C. Liu, "Cascade MQDF classifier for handwritten character recognition," *J. Tsinghua Univ. (Sci. & Tech.) (in Chinese)*, 2008.

[15] Q. Fu, X. Ding, and C. Liu, "A new adaboost algorithm for large scale classification and its application to Chinese handwritten character recognition," *Int'l Conf. Frontiers in Handwriting Recognition*, 2008.

[16] T.-F. Gao and C.-L. Liu, "High accuracy handwritten Chinese character recognition using LDA-based compound distances," *Pattern Recognition*, 2008.

[17] B. Xu, K. Huang, I. King, C.-L. Liu, J. Sun, and N. Satoshi, "Graphical lasso quadratic discriminant function and its application to character recognition," *Neurocomputing*, 2013.

[18] T. Kawatani, "Handwritten Kanji recognition with determinant normalized quadratic discriminant function," *Int'l Conf. Pattern Recognition*, 2000.

[19] N. Kato, M. Suzuki, S. Omachi, H. Aso, and Y. Nemoto, "A handwritten character recognition system using directional element feature and asymmetric Mahalanobis distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1999.

[20] C.-L. Liu, "High accuracy handwritten Chinese character recognition using quadratic classifiers with discriminative feature extraction," *Int'l Conf. Pattern Recognition*, 2006.

[21] J. Friedman, "Regularized discriminant analysis," *J. American Statistical Association*, 1989.

[22] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang, "CASIA online and offline Chinese handwriting databases," *Int'l Conf. Document Analysis and Recognition*, 2011.

[23] C.-L. Liu, "Normalization-cooperated gradient feature extraction for handwritten character recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2007.

[24] C.-L. Liu and X.-D. Zhou, "Online Japanese character recognition using trajectory-based normalization and direction feature extraction," *Int'l Workshop Frontiers in Handwriting Recognition*, 2006.

[25] K. Ding, G. Deng, and L. Jin, "An investigation of imaginary stroke techinique for cursive online handwriting Chinese character recognition," *Int'l Conf. Document Analysis and Recognition*, 2009.

[26] S. Srivastava, M. Gupta, and B. Frigyik, "Bayesian quadratic discriminant analysis," *J. Machine Learning Research*, 2007.

[27] X.-Y. Zhang and C.-L. Liu, "Evaluation of weighted Fisher criteria for large category dimensionality reduction in application to Chinese handwriting recognition," *Pattern Recognition*, 2013.