2009 10th International Conference on Document Analysis and Recognition

Recognition of Handwritten Chinese Characters by Combining Regularization, Fisher's Discriminant and Distorted Sample Generation

K.C. Leung and C.H. Leung*
Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong.
** email address: chleung@eee.hku.hk*

Abstract

The problem of offline handwritten Chinese character recognition has been extensively studied by many researchers and very high recognition rates have been reported. In this paper, we propose to further boost the recognition rate by incorporating a distortion model that artificially generates a huge number of virtual training samples from existing ones. We achieve a record high recognition rate of 99.46% on the ETL-9B database. Traditionally, when the dimension of the feature vector is high and the number of training samples is not sufficient, the remedies are to (i) regularize the class covariance matrices in the discriminant functions, (ii) employ Fisher's dimension reduction technique to reduce the feature dimension, and (iii) generate a huge number of virtual training samples from existing ones. The second contribution of this paper is the investigation of the relative effectiveness of these three methods for boosting the recognition rate.

1. Introduction

The problem of offline handwritten Chinese character recognition has been extensively studied by a lot of researchers, and very high recognition rates have been reported [1-20]. One difficulty often encountered is the lack of sufficient samples for training up the classifier. In most of the proposed systems, feature vectors of a few hundred dimensions are extracted from the input character. In order to have a non-singular class covariance matrix, the number of training samples must be greater than the feature dimension. Moreover, for statistical accuracy, the number of samples should be a lot more than this minimum. In practice, three techniques can be used to deal with this problem. The first is to regularize the class covariance matrix, as in the modified quadratic discriminant functions (MQDF) proposed by Kimura et al [1]. The second is to reduce the feature dimension using Fisher's dimension reduction technique [21]. The third method is to generate a huge number of artificial training samples by applying distortions to existing training samples [15-20]. The first two techniques are very popular and have been adopted in most of the reported systems including those with the highest recognition rates [2,4,5,10]. In this paper, we propose to apply all the three techniques together and we show that this can boost the recognition rate to a record high of 99.46% on the ETL-9B database. We also carried out a study on evaluating the relative effectiveness of these three techniques for boosting the recognition rate.

2. Regularization, Fisher's linear discriminant, and distorted sample generation

2.1 Regularization

Let n be the feature dimension, $\underline{x}$ be the unknown feature vector to be classified, $\underline{\mu}_j$ and Σ_j be the mean vector and covariance matrix respectively of class j, λ_i and $\underline{\varphi}_i$ be the i^{th} largest eigenvalue and eigenvector of Σ_j respectively. The quadratic discriminant function (QDF) is:

$$g_j(\underline{x}) = -\left(\underline{x}-\underline{\mu}_j\right)^T \Sigma_j^{-1}\left(\underline{x}-\underline{\mu}_j\right) - \log\left|\Sigma_j\right|$$
$$= -\sum_{i=1}^{n}\frac{1}{\lambda_i}\left[\underline{\varphi}_i^T\left(\underline{x}-\underline{\mu}_j\right)\right]^2 - \log\prod_{i=1}^{n}\lambda_i$$

To alleviate the problem of inaccurate estimation of the covariance matrix, two forms of modified QDFs, namely MQDF1 and MQDF2 [1], are often adopted, with MQDF1 given by:

$$g_j^{(1)}(\underline{x}) = -\sum_{i=1}^{n}\frac{1}{\lambda_i+h^2}\left[\underline{\varphi}_i^T\left(\underline{x}-\underline{\mu}_j\right)\right]^2 - \log\prod_{i=1}^{n}\left(\lambda_i+h^2\right)$$

where h^2 is a constant and $1 \leq k < n$.

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.48

2.2 Fisher's Linear Discriminant

The technique of Fisher's linear discriminant aims at reducing the dimension of the feature space while maximizing the separation of different classes of samples by finding some optimal projection planes $\underline{w}_i$. It can be shown that $\underline{w}_i$ are solutions to the generalized eigenvector problem [21]:

$$S_B \underline{w}_i = \lambda_i S_W \underline{w}_i \qquad i = 1, 2, \ldots$$

where S_B and S_W are the between-class and within-class scatter matrices respectively.

2.3 Distorted sample generation

Let $f(x,y)$ be an original handwritten character. By applying distortions to $f(x,y)$, additional samples can be obtained. Let $g(u,v)$ be the distorted character. The distortions are achieved via the mapping functions h_x and h_y such that $u = h_x(x,y)$ and $v = h_y(x,y)$ for all x and y. In this paper, we follow the mapping functions as described in [15] that take care of both shearing and local resizing. First we set:

$$u = h_x(x,y) = w_n(a_1, b_1(x)) + k_1 y + c_1, \text{ and}$$
$$v = h_y(x,y) = w_n(a_2, b_2(y)) + k_2 x + c_2,$$

where k_1 and k_2 are the shearing slopes, $a_1 \neq 0$ and $a_2 \neq 0$ are constants, c_1 and c_2 are constants to align the centroid back to the original position, b_1 and b_2 are functions to linearly scale the original coordinates to the interval [0,1], and w_n is a nonlinear warping function for producing local variations on the size of sub-patterns.

Two nonlinear warping functions w_1 and w_2 are implemented in order to produce more kinds of variations. The choice between w_1 and w_2 is selected randomly with a certain pre-defined probability. The two warping functions are plotted in Figure 1 and an example of distorted sample generation is illustrated in Figure 2.

$$w_1(a,t) = \frac{1-e^{-at}}{1-e^{-a}}$$

$$w_2(a,t) = \begin{cases} 0.5 w_1(a, 2t), & 0 \le t \le 0.5 \\ 0.5 + 0.5 w_1(-a, 2(t-0.5)), & 0.5 < t \le 1 \end{cases}$$

3. Normalization and feature extraction

The input character in the form of a 64x64 binary matrix to be recognized is first preprocessed to remove isolated noise pixels. The stroke boundaries are extracted and a 4-direction chain code is assigned to each boundary pixel.

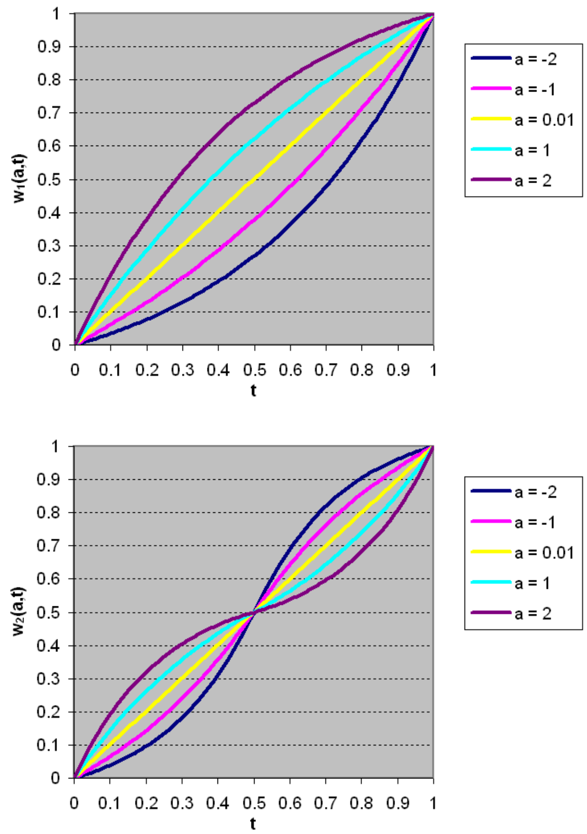


Figure 1: Warping functions $w_1(a,t)$ and $w_2(a,t)$.

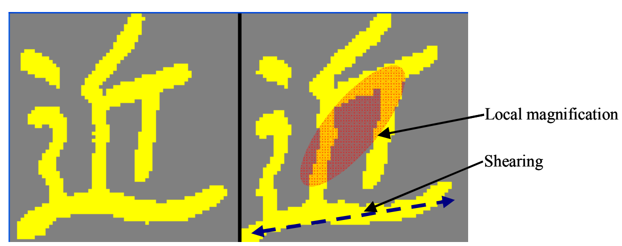
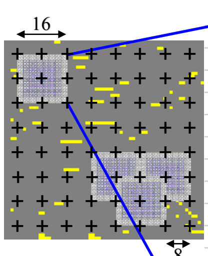


Figure 2: (Left) Genuine handwritten sample "近"; (Right) Artificially generated sample.

Nonlinear normalization is then applied to normalize the stroke distribution of the character. We adopt two nonlinear normalization methods: NLN-T [6] and NLN-Y [7], and extended them to 2D NLN-T and 2D NLN-Y respectively by the method in [8]. In implementing the NLN-T method [6], we set $\alpha = \beta = 0.22$ for all the stoke pixels. A deficiency in the original NLN-T method is that most of the white pixels outside the convex hull of the character (i.e., near the character frame boundary) generally possess a large value of $\alpha(i,j)$ and $\beta(i,j)$, leading to the formation of completely blank rows and columns in the normalized image. This is not desirable for recognition because the normalized characters tend to have large variations in size. To rectify this problem, we apply an edge penalty to scale down $\alpha(i,j)$ and $\beta(i,j)$ for these pixels by multiplying them with a constant less than 1.0. In our experiments, this constant is fixed at 0.2. The unified line density at a point, as needed in the 2D normalization process, is taken as $[\alpha(i,j) + \beta(i,j)]/2$. In implementing the NLN-Y method, we directly follow the method named "line density by inscribed circle" as described in [7]. For converting the 1D NLN-T and 1D NLN-Y to their 2D counterpart, the technique suggested by Horiuchi et al. [8] is implemented. We do not include the offset functions θ_x and θ_y for smoothing the line densities since we empirically find that this would produce adverse effects to the overall recognition rates. The parameter σ in the Gaussian mask for 2D normalization is set to 16.0 in all experiments.

Since nonlinear normalization distorts the original character, the stroke directions are distorted. Besides causing problems in discriminating between characters such as "千" and "干", the normalized character strokes often look very rugged instead of smooth. Hence we assign the direction codes in the original character to the stroke pixels in the normalized character [11]. The character is split into four sub-frames with each sub-frame containing pixels of one direction code only. Each sub-frame is then filtered by a 16x16 mask and sampled at an 8x8 sampling grid to give 64 feature values. The mask is of an improved design such that for each pixel, the contribution to each of its four neighboring sampling grid points is approximately inversely proportional to the Euclidean distance and the total contribution is 1 (Figure 3). This makes the extracted features more invariant to local variations of stroke positions. With 4 sub-frames, a 256-D feature vector is obtained. Variable transformation (by taking the square root) is then applied to each feature value to make the statistical distribution more Gaussian-like [17].



9	9	15	15	17	17	13	13	13	13	17	17	15	15	9	9
9	9	15	15	17	17	13	13	13	13	17	17	15	15	9	9
15	15	19	19	24	24	26	26	26	26	24	24	19	19	15	15
15	15	19	19	24	24	26	26	26	26	24	24	19	19	15	15
17	17	24	24	33	33	42	42	42	42	33	33	24	24	17	17
17	17	24	24	33	33	42	42	42	42	33	33	24	24	17	17
13	13	26	26	42	42	65	65	65	65	42	42	26	26	13	13
13	13	26	26	42	42	65	65	65	65	42	42	26	26	13	13
13	13	26	26	42	42	65	65	65	65	42	42	26	26	13	13
13	13	26	26	42	42	65	65	65	65	42	42	26	26	13	13
17	17	24	24	33	33	42	42	42	42	33	33	24	24	17	17
17	17	24	24	33	33	42	42	42	42	33	33	24	24	17	17
15	15	19	19	24	24	26	26	26	26	24	24	19	19	15	15
15	15	19	19	24	24	26	26	26	26	24	24	19	19	15	15
9	9	15	15	17	17	13	13	13	13	17	17	15	15	9	9
9	9	15	15	17	17	13	13	13	13	17	17	15	15	9	9

Figure 3: Improved 16x16 mask (scaled by 100).

4. Experimental results

The handwritten character samples from the ETL-9B database are used for evaluation of the algorithms. The database contains 3,036 character classes with 200 samples per class. Following the partitioning scheme in [10], for each character class, we take sample number 21 - 180 to form the training set while the remaining samples form the test set.

Regarding the generation of distorted character samples, the constant a in the warping functions controlling the extent of local resizing is randomly taken from the interval [-1.6, +1.6]. Shearing constants k_1 and k_2 are random numbers picked from [-0.17, +0.17] and [-0.20, +0.20] respectively. Finally, the probabilities of applying warping function w_1 and w_2 are fixed at 0.8 and 0.2 respectively.

We performed five sets of experiments to fully evaluate our recognition system:

- [Regularization] only
- [Fisher's linear discriminant] only
- [Distorted sample generation] only
- [Regularization + Fisher's linear discriminant]
- [Regularization + Fisher's linear discriminant + Distorted sample generation]

The Bayes classifier assuming Gaussian statistics was employed with the regularization method as given in MQDF1. The 2D NLN-T and 2D NLN-Y methods were used for nonlinear normalization. We found that after applying the edge penalty as explained above, 2D NLN-T performed better than 2D NLN-Y. The adoption of the direction code from the original character (before normalization) as well as using the improved 16x16 mask also helped to increase the recognition rate.

We first used the regularization method alone without the distortion model and Fisher's linear discriminant. The results are given in Table 1. The best recognition rate achieved is 99.19%.

Table 1: Recognition rates (%) obtained by using regularization only.

Regularization constant in MQDF1	Recognition rate (%) (2D NLN-T)
0.75	99.17
0.85	99.18
0.95	99.19
1.05	**99.19**
1.15	99.19
1.25	99.19
1.50	99.19
2.00	99.17

We then applied Fisher's discriminant alone. As the original feature dimension is 256 and the number of training samples per class is only 160, the reduced dimension should be within the range of 1 to 159 in order for Σ^{-1} in the QDF equation to exist. Table 2 summarizes our findings. Relatively low recognition rates are attained, implying that Fisher's linear discriminant by itself is not very effective. A Fisher's dimension of 60 gives the best recognition rate of 96.90%.

Table 2: Recognition rates (%) obtained by using Fisher's linear discriminant only.

Fisher dimension	Recognition rate (%) (2D NLN-T)
20	92.77
40	96.81
60	**96.90**
80	96.10
100	94.26
120	89.69
140	74.63

The third experiment is on the effectiveness of applying distorted sample generation alone. We generated additional training samples from each genuine training sample. As summarized in Table 3, the recognition rate increases with the number of generated samples. With a total of 16,000 (= 100x160) training samples per class, the recognition rate reaches 99.37%, showing that distorted sample generation is the most effective technique among the three methods.

Table 3: Recognition rates (%) obtained by using distorted sample generation only.

Distorted samples per genuine sample	Recognition rate (%) (2D NLN-T)
4	98.50
9	99.09
24	99.29
49	99.35
74	99.36
99	**99.37**

Next, we combined the regularization method with Fisher's linear discriminant. In the Fisher's discriminant, we also regularized S_B and S_W such that for each matrix, we added the 201st largest eigenvalue of the matrix to all the diagonal elements. The nonlinear normalization method adopted was either 2D NLN-Y or 2D NLN-T. From the results in Table 4, it is observed that combined technique performs better than either of the methods alone. Furthermore, using the modified S_B and S_W in Fisher's linear discriminant also leads to some improvement. A recognition rate of 99.31% is achieved, which is comparable to those attained by Liu using discriminative training [4], and Gao and Liu using LDA compound distance [5].

Table 4: Recognition rates (%) obtained by using regularization and Fisher's discriminant.

		Original Fisher's equation		Fisher's equation with modified S_B and S_W	
Fisher dimension	Regularization constant in MQDF1	2D NLN-Y	2D NLN-T	2D NLN-Y	2D NLN-T
160	0.55	98.98	99.24	98.99	99.24
	0.75	98.99	**99.26**	99.02	99.27
	0.95	**98.99**	99.26	**99.02**	**99.29**
	1.15	98.99	99.24	99.02	99.28
	1.50	98.97	99.22	99.02	99.27
200	0.55	99.00	99.25	98.99	99.25
	0.75	99.01	**99.28**	99.00	99.29
	0.95	**99.01**	99.26	**99.02**	**99.31**
	1.15	99.00	99.24	99.01	99.31
	1.50	98.97	99.23	99.01	99.29

Finally, we combined regularization, Fisher's linear discriminant, and distorted sample generation together. In the Fisher's discriminant, we regularized each of S_B and S_W by adding the 217th largest eigenvalue of the matrix to its diagonal elements. The results are shown in Table 5. The highest recognition rate is 99.46%.

Table 5: Recognition rates (%) obtained by using regularization, Fisher's discriminant and distorted sample generation.

		Distorted sample per genuine sample			
		24		74	
Fisher dimension	Regularization constant in MQDF1	2D NLN-Y	2D NLN-T	2D NLN-Y	2D NLN-T
176	0.10	99.26	99.37	99.31	99.41
	0.20	**99.28**	99.38	**99.31**	99.43
	0.25	99.28	99.39	99.30	99.43
	0.30	99.28	**99.41**	99.30	**99.44**
	0.40	99.28	99.40	99.30	99.44
	0.50	99.28	99.40	99.29	99.43
216	0.10	99.28	99.40	99.32	99.43
	0.20	**99.31**	99.42	**99.33**	99.46
	0.25	99.30	**99.42**	99.32	**99.46**
	0.30	99.30	99.42	99.32	99.45
	0.40	99.30	99.42	99.31	99.45
	0.50	99.30	99.41	99.31	99.44

5. Discussions and conclusions

The improved feature extraction mask, the modified nonlinear normalization method, and Fisher's equation with regularized S_B and S_W help to give more stable statistical estimations in training the classifier. Moreover, by combining regularization, Fisher's discriminant and distorted sample generation, we have achieved a record high recognition rate of 99.46% on the ETL-9B database. To our knowledge, this is the highest recognition rate reported in the literature so far. We have also compared the relative effectiveness of these three methods in boosting the recognition rate. Our experiments show that distorted sample generation is the most effective, followed by regularization and Fisher's discriminant.

6. Acknowledgement

We sincerely thank the AIST of Japan for permitting the use of the ETL-9B database for this research.

7. References

[1] F. Kimura, K. Takashina, S. Tsuruoka, Y. Miyake, Modified Quadratic Discriminant Functions and the Application to Chinese Character Recognition, IEEE Trans. Pattern Anal. Mach. Intell. 9 (1) (1987) 149–153.

[2] N. Kato, M. Suzuki, S. Omachi, H. Aso, Y. Nemoto, A Handwritten Character Recognition System using Directional Element Feature and Asymmetric Mahalanobis Distance, IEEE Trans. Pattern Anal. Mach. Intell. 21 (3) (1999) 258–262.

[3] C.-L. Liu, H. Sako and H. Fujisawa, Performance Evaluation of Pattern Classifiers for Handwritten Character Recognition, International Journal on Document Analysis Recognition 4 (2002) 191–204.

[4] 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.

[5] T.-F. Gao, C.-L. Liu, High Accuracy Handwritten Chinese Character Recognition using LDA-based Compound Distances, Pattern Recognition 41 (11) (2008) 3442–3451.

[6] J. Tsukumo, H. Tanaka, Classification of Handprinted Chinese Characters using Non-linear Normalization and Correlation Methods, Proceedings of the 9th International Conference on Pattern Recognition, Roma, Italy, 1988, pp. 168–171.

[7] H. Yamada, K. Yamamoto, T. Saito, A Nonlinear Normalization Method for Handprinted Kanji Character Recognition—Line Density Equalization, Pattern Recognition 23 (9) (1990) 1023–1029.

[8] T. Horiuchi, R. Haruki, H. Yamada, K. 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, Ulm, Germany, 1997, pp. 511–514.

[9] C.-L. Liu, H. Sako, H. Fujisawa, Handwritten Chinese Character Recognition: Alternatives to Nonlinear Normalization, Proceedings of the 7th International Conference on Document Analysis and Recognition, Edinburgh, Scotland, 2003, pp. 524–528.

[10] C.-L. Liu and K. Marukawa, Pseudo Two dimensional Shape Normalization Methods for Handwritten Chinese Character Recognition, Pattern Recognition 38 (2005) 2242–2255.

[11] M. Hamanaka, K. Yamada, J. Tsukumo, Normalization-cooperated Feature Extraction Method for Handprinted Kanji Character Recognition, Proceedings of the Third International Workshop on Frontiers of Handwriting Recognition, Buffalo, NY, 1993, pp. 343–348.

[12] J.-X. Dong, A. Krzyzak, C.-Y. Suen, An Improved Handwritten Chinese Character Recognition System using Support Vector Machine, Pattern Recognition Letters 26 (2005) 1849–1856.

[13] Q. Sun, L.-L. Wang, S.-H. Lim, G. DeJong, Robustness through Prior Knowledge: using Explanation-based Learning to Distinguish Handwritten Chinese Characters, International Journal on Document Analysis and Recognition 10 (2007) 175–186.

[14] H.-C. Fu, Y.-Y. Xu, Multilinguistic Handwritten Character Recognition by Bayesian Decision-based Neural Networks, IEEE Trans. Signal Process. 46 (10) (1998) 2781–2789.

[15] C.-H. Leung, Y.-S. Cheung, K.-P. Chan, A Distortion Model for Chinese Character Generation, Proceedings of the 1985 IEEE International Conference on Systems, Man and Cybernetics, Tucson, Arizona, 1985, pp. 38–41.

[16] J. Mao and K.M. Mohiuddin, Improving OCR Performance using Character Degradation Models and Boosting Algorithm, Pattern Recognition 18 (1997) 1415–1419.

[17] F. Kimura, T. Wakabayashi, S. Tsuruoka, Y. Miyake, Improvement of Handwritten Japanese Character Recognition using Weighted Direction Code Histogram, Pattern Recognition 30 (8) (1997) 1329–1337.

[18] G. Chen, H.-G. Zhang, J. Gao, Learning Pattern Generation for Handwritten Chinese Character using Pattern Transform Method with Cosine Function, Proceedings of the 5th International Conference on Machine Learning and Cybernetics, Dalian, 2006, pp. 3329–3333.

[19] H. Miyao, M. Maruyama, Virtual Example Synthesis Based on PCA for Off-line Handwritten Character Recognition, Proceedings of the 7th International Workshop on Document Analysis Systems, Nelson, New Zealand, 2006, pp. 96–105.

[20] T. Varga, H. Bunke, Generation of Synthetic Training Data for an HMM-based Handwriting Recognition System, Proceedings of the 7th International Conference on Document Analysis and Recognition, Edinburgh, Scotland, 2003, pp. 618–622.

[21] R.O. Duda, P.E. Hart and D.G. Stork, Pattern Classification, 2nd Edition, Wiley, 2001.