2013 12th International Conference on Document Analysis and Recognition

Style Consistent Perturbation for Handwritten Chinese Character Recognition

Fei Yin, Ming-Ke Zhou, Qiu-Feng Wang, Cheng-Lin Liu
National Laboratory of Pattern Recognition (NLPR)
Institute of Automation, Chinese Academy of Sciences
95 Zhongguancun East Road, Beijing 100190, P.R. China
E-Mail: {fyin, mkzhou,wangqf, liucl}@nlpr.ia.ac.cn

***Abstract*—Perturbation-based recognition is effective to recover the deformation of handwritten characters and improve the recognition performance by generating multiple distortions and selecting a distortion that best restores character deformation. Considering that the characters in a field undergo similar deformation under a consistent style, we proposed style consistent perturbation for handwritten character recognition. By generating multiple distortions for the characters in a field, each distortion style is evaluated at the field level and the uniform distortion style of maximum recognition confidence is selected to give the final result. To overcome the slight deviation from uniform style, we also propose to search the neighborhood distortions from the optimal uniform distortion for higher confidence. The experiments of handwritten Chinese character recognition on multi-writer data show that style consistent perturbation in very short fields outperforms individual character recognition, and neighborhood distortion search yields further improvement.**

Keywords—Style consistent; perturbation; field classification; handwritten Chinese character recognition.

I. Introduction

Offline handwritten Chinese character recognition (HCCR) is still a challenge. Although great efforts have been made during the past 40 years [1], and high accuracies of over 98% have been reported on regularly written databases [2,3]), recent studies on unconstrained handwriting show that the accuracy is still low, e.g., the very recently reported test accuracy 89.55% on a 3,755-class dataset CASIA-HWDB1.1 using a state-of-the-art recognizer [4]. The ICDAR2011 Chinese Handwriting Recognition Competition reported the highest accuracy 92.18% by training with large sample set [5]. Low accuracies have been reported on another handwriting database HIT-MW [6].

HCCR is difficult due to the large category set and the divergent writing styles (Fig. 1). After exploitation of state-of-the-art character normalization, feature extraction and classification methods, the recognition performance is now reaching a bottleneck. Further improvements are made possible by training with huge sample set (including distorted samples [7-9]) and exploiting the writing style consistency [10,11]. The superiority of deep neural networks [12] is partly due to training with huge number of distorted samples. On the other-hand, perturbation-based recognition [13,14] improves the performance by generating multiple distortions on the test pattern with the expectation that at least one of the distortions is similar to standard writing. However, perturbation can also increase the confusion between the distorted pattern and other classes, thus does not guarantee improved performance.

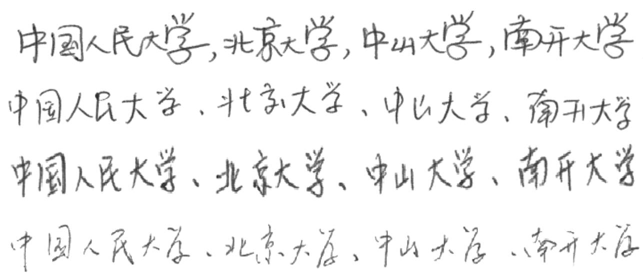


Fig. 1. Chinese character strings written by four persons, one line per person.

The perturbation method distorts different patterns in the same field (document, string, or word) individually. Consider the style consistency of written characters in the same field [10,11], we propose style consistent perturbation for handwritten character recognition. As shown in Fig. 1, the characters written by the same person show similar style. Particularly, when a character is written multiple times by the same person, it appears in similar shape. The basic principle of style consistent perturbation is that the characters in the same field under similar distortion. In implementation, we design multiple distortion styles. Under each style, the test characters in a field are distorted in the same way and all the distorted character are recognized. The distortion style that has the highest confidence of recognition is selected as the mostly likely distortion and its recognition result is taken as the final result. The key in style consistent perturbation is that multiple characters undergo the same distortion, unlike the ordinary perturbation that distorts different characters independently.

We evaluated the performance of style consistent perturbation on the Chinese handwriting dataset CASIA-HWDB1.1 [15], where the samples are stored in writer-specific files. Using style consistent perturbation on multiple samples of the same writer is shown to outperform individual recognition. To overcome the writer's slight deviation from uniform style, we propose to search neighborhood distortions in style consistent perturbation to further improve the performance.

1520-5363/13 $26.00 © 2013 IEEE
DOI 10.1109/ICDAR.2013.210

II. RELATED WORKS

The proposed style consistent perturbation method was motivated by perturbation-based character recognition and style consistent pattern field classification. We briefly review the related works below.

A. Perturbation-based character recognition

Perturbation can restore the deformation of handwritten characters, so it was often used as a powerful tool for synthesizing samples. A straightforward use is to synthesize distorted samples for improving the generalization performance of classifiers by training with large sample set, such as for deep neural networks [12]. Specific techniques have been proposed for distorting Chinese characters [8,9]. They have shown effectiveness in improving the performance of HCCR.

On the other hand, perturbation-based recognition method synthesizes distorted samples during classification stage. The distortion is aimed to restore the deformation of character shape caused by writing style variation. By generating multiple distortions on the test pattern, it is hoped at least one distortion is similar to the standard shape and gets high confidence in classification. Yasuda et al. proposed a perturbed correlation method [13], where the original character is slightly distorted by two-dimensional affine transformation with six parameters, and the distortion with best correlation with standard templates gives the recognition result. Ha and Bunk proposed a perturbation-based method for handwritten numeral recognition [14], where the numeral images were distorted by four parameterized geometric transformations, and the decision of classification was made by weighted-voting on the outputs of multiple distortions. These methods consider the distortions of different characters independently.

B. Style consistent filed classification

By field classification, multiple patterns in a field are classified simultaneously and the style consistency can be explored to improve the performance. Sarkar and Nagy [10] proposed a style mixture model which assumes that distribution of a field is the mixture of a fixed number of styles. Veeramachaneni and Nagy [11] proposed a field classification model which assumes Gaussian field-class-conditional density and uses only second-order statistics to reduce the number of parameters. The above two methods are hard to be used in large category set Chinese character recognition due to the high computational complexity. Zhang et al. [16] proposed a field Bayesian model for field classification, which learns a style normalized transformation (SNT) for each field and transforms the patterns of different fields to a uniform style space. For good performance, it needs a large number of patterns in a field to learn the SNT.

III. STYLE CONSISTENT PERTURBATION-BASED CLASSIFICATION

Unlike the existing perturbation-based recognition methods, the proposed method distorts multiple patterns in a field under the same distortion style and selects the style of maximum field classification confidence. In the following, we present the mathematical formulation and the classification techniques.

A. Mathematical Foundation

We consider the problem of classifying a field of L isogenous patterns $\mathbf{x}_1,\cdots,\mathbf{x}_L$, each $\mathbf{x}_i$ represents one pattern and belongs to one of C classes: $c_i \in \{1,\cdots,C\}$. For simplicity, we can define the concatenation of patterns $f=(\mathbf{x}_1,\cdots,\mathbf{x}_L)$ as field pattern, and the concatenation of the classes $c=(c_1,\cdots,c_L)$ as field class.

From the viewpoint of Bayes' theorem, field classification is to assign the field pattern to the field class of maximum a posteriori probability (MAP), which takes the form

$$\begin{aligned} p(c\mid f) &= \frac{p(c)p(f\mid c)}{p(f)} \\ &= \frac{p(c_1,\cdots,c_L)p(\mathbf{x}_1,\cdots,\mathbf{x}_L\mid c_1,\cdots,c_L)}{p(\mathbf{x}_1,\cdots,\mathbf{x}_L)} \end{aligned} \tag{1}$$

Obviously, for computing the a posteriori probability, we need the field class a priori probability $p(c_1,\cdots,c_L)$ and field-class-conditional probability $p(\mathbf{x}_1,\cdots\mathbf{x}_L\mid c_1\cdots c_L)$, which are not trivial to formulate, however.

Although the spatial context and linguistic context are widely used in text (character string) recognition, we herein focus on the style context, i.e., the style consistency in a field. We ignore the linguistic context by assuming that the patterns in field are linguistically independent:

$$p(c_1,\cdots c_L)=p(c_1)\cdots p(c_L) \tag{2}$$

Then, how to define field-class-conditional probability becomes the key problem. Some researchers assume Gaussian mixture density for pattern fields, but the number of field classes increases exponentially with the length of field (L), and this makes the computation intractable for large category set and large field length. Zhang et al. [15] proposed a strategy to overcome the exponential complexity of field class density model by assuming independence of patterns conditioned on uniform field style transformation. We incorporate this assumption into style consistent perturbation and formulate the field-class-conditional density as:

$$p(\mathbf{x}_1,\cdots,\mathbf{x}_L\mid c_1,\cdots,c_L;T_k)=\prod_{i=1}^{L}p(T_k(\mathbf{x}_i)\mid c_i) \tag{3}$$

where T_k denotes a perturbation (transformation) style applied to the patterns in a field simultaneously. This model takes into account the style consistency while avoiding the exponential complexity of field class. The probability $p(T_k(\mathbf{x}_i)\mid c_i)$ can be given by a classifier.

B. Classification Strategy

For using the field-class-conditional density model (3) in field classification, we design multiple distortion styles T_k, k=1,…,K. Under each distortion style, we obtain the conditional density for hypothesized field classes according to (3) and in turn the a posteriori probability according to (1). Assuming equal a priori probabilities as in (2), the field class of test patterns is decided by the MAP rule choosing the optimal distortion style:

$$\begin{aligned}\left\langle c, T_{opt} \right\rangle &= \arg\max_k \prod_{i=1}^{L} \max_{c_i} p(T_k(\mathbf{x}_i) \mid c_i) \\ &= \arg\max_k \sum_{i=1}^{L} \max_{c_i} \log p(T_k(\mathbf{x}_i) \mid c_i)\end{aligned} \tag{4}$$

Formula (4) indicates that we can predict the field class and the optimal perturbation model simultaneously given the class-conditional probabilities output by a classifier on distorted patterns. In the ideal case, a uniform perturbation style T_{opt} gives optimal shape restoration. But in practice, the writing style on different characters is not strictly consistent, which may deviate slightly from a uniform style. We thus propose a variation of style consistent perturbation which allows each pattern in the field deviates from a uniform style.

To allow each pattern in a field distort freely from a uniform style is equivalent to the ordinary perturbation method that distort patterns individually. Our proposed method is thus a compromise between style consistent perturbation and individual perturbation: it searches the similar distortion styles from the optimal uniform distortion style after the uniform style has been decided by (4). To do this, we define a neighborhood of style: the neighbors of a uniform style T_{opt} are the distortion styles that are close to the uniform style in the character feature space or have similar shape transformation parameters. We denote the neighborhood as $N(T_{opt})$. While the optimal uniform distortion style is decided in field classification, in neighborhood search, the patterns are decided individually. Since the neighborhood is constrained to be close from the uniform optimal style, the classification result is different from that of individual perturbation. The neighborhood search rule is formulated as

$$c_i^{'} = \arg\max_{k \in N(T_{opt})} \max_{c_j} \log p(T_k(\mathbf{x}_i) \mid c_j) \tag{5}$$

According to (4), if we set the length of field as L=1, the classification degenerates into the ordinary perturbation-based classification. In neighbor search according to (5), if the neighbor is identical to the optimal uniform distortion style, it is equivalent to style consistent perturbation as in (4); if the neighborhood contains all the distortion styles, the classification is equivalent to ordinary individual perturbation. So, the neighborhood search method is a compromise between individual perturbation and style consistent perturbation.

C. Confidence Transformation

The above classification model requires the classifier gives the class-conditional probability $p(T_k(\mathbf{x}_i) \mid c_i)$. We replace this conditional probability with the a posteriori probability, which is estimated using a confidence transformation method. There have been variable methods for transforming classifier outputs into a posteriori probabilities. We adopt the evidence combination method in [17], which have shown promise in handwriting recognition [18]. This method first transforms each class output measure into sigmoidal binary probability, then combines the binary probabilities into multi-class probabilities using the Dempster-Shafter theory of evidence. The resulting probabilities are

$$P(c_j \mid \mathbf{x}) = \frac{\exp\left[-a \cdot d_j(\mathbf{x}) + b\right]}{1 + \sum_{i=1}^{C} \exp\left[-a \cdot d_i(\mathbf{x}) + b\right]}, \quad j = 1, 2, \ldots, C \tag{6}$$

where C is the number of defined classes, $d_j(\mathbf{x})$ is the output score of class c_j, α and β are the confidence parameters estimated by empirical loss minimization.

Consider the relationship between conditional probability and a posteriori probability:

$$P(\mathbf{x} \mid c_j) = \frac{p(\mathbf{x}) P(c_j \mid \mathbf{x})}{P(c_j)} \propto \frac{P(c_j \mid \mathbf{x})}{P(c_j)} \tag{7}$$

and assume equal a priori probabilities $P(c_j)$, we can use $p(c_i \mid T_k(\mathbf{x}_i))$ to replace $p(T_k(\mathbf{x}_i) \mid c_i)$ in style consistent perturbation-based classification rules (4) and (5).

IV. CHARACTER RECOGNITION SYSTEM

We apply the proposed style consistent perturbation method to handwritten Chinese character recognition (HCCR). Fig. 2 shows the flowchart of the recognition system. Multiple character images of a field (written by the same person) are input into the character recognizer, which performs character normalization, feature extraction and classification. Each character undergoes multiple distortions. The classifier outputs of multiple characters under the same distortion are fused to give the field classification decision of this distortion, and the field decisions of multiple distortions are compared to decide the optimal uniform style as well as the optimal field class. From the optimal uniform style, neighborhood distortion search is performed to fine tune the character classes in the field.

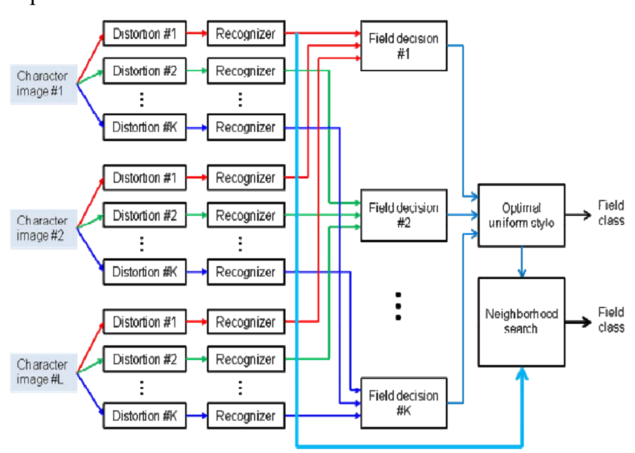


Fig. 2. Flowchat of style consistent perturbation-based recognition.

For generating distortions, we use the distortion function proposed in [9].

$$\begin{cases} u = w_n(d_1, b_1(y)) + k_1 b_2(x) + c_1 \\ v = w_n(d_2, b_2(x)) + k_2 b_1(y) + c_2 \end{cases} \quad (8)$$

where c_1 and c_2 are constants, and b_1 and b_2 are functions to linearly scale coordinates to the interval [0,1]. In our system, Perturbation parameter set $P = \{P_1, \cdots, P_i, \cdots, P_n\}$ are chosen from free parameters $\{d_1, d_2, k_1, k_2\}$ and free functions $\{w_1, w_2\}$, so that $P_i = \{d_1^i, d_2^i, k_1^i, k_2^i, w_n^i\}$. According to (8), the pixel coordinates (x,y) of original image are mapped to coordinates (u,v) in image distortion.

Our system was implemented in C++ code and ran on a linux server with 2*Xeon 5060 CPU, 128G RAM and 2*Tesela C2050 GPU cards.

V. EXPERIMENTAL RESULTS

We evaluated the performance of perturbation-based recognition on an offline handwritten Chinese character dataset CASIA-HWDB1.1 [15]. This dataset contains character samples of 3,755 classes written by 300 writers and the samples are stored in writer-specific files. Two classifiers, nearest prototype classifier (NPC) [19] and modified quadratic discriminant function (MQDF) [20] were used for character classification.

A. Experimental Setting

The dataset CASIA-HWDB1.1 contains character samples of 300 writers (no.1001-1300). It is divided into a training set of 240 writers (no.1001-1240) and a test set of 40 writers (no.1241-1300). Each writer has a set of Chinese character samples of 3,755 classes (GB2312-80 level-1 set). There are 897,758 training samples and 223,991 samples in total.

The character recognizer uses normalization–cooperated gradient feature (NCGF) [21]. The obtained 512D feature vector is reduced to 160D by Fisher linear discriminant analysis (FLDA), and then input into the classifier (MQDF and NPC with one prototype per class). The classifier parameters were estimated on the training sample set. For NPC training, we use the learning method called log-likelihood of hypothesis margin (LOGM) [19]. For the MQDF classifier, the mean and covariance matrix of each class are estimated by maximum likelihood, and after diagonal decomposition of covariance, we retain 50 principal eigenvectors for each class. The unified minor eigenvalue is proportional to a hyperparameter, which is selected by cross-validation (holding 1/5 of training samples for validation, and after selecting the hyperparameter, re-estimating classifier parameters on the whole training set).

For confidence parameter estimation, we first trained a classifier on 4/5 of training samples, and the classifier outputs on the held 1/5 of training samples were used to estimate the confidence parameters.

Using the distortion function (8), we generated 181 distortion styles by selecting perturbation parameters from the parameters set {d_1={-0.09, -0.03, 0.03, 0.09}, d_2={-0.09, -0.03, 0.03, 0.09}, k_1={-0.08, 0, 0.08}, k_2={-0.01, 0, 0.01}, w_1} and the set {d_1={-0.06, 0.06}, d_2={-0.06, 0.06}, k_1={-0.08, 0, 0.08}, k_2={-0.01, 0, 0.01}, w_2}.

For the neighborhood of each distortion style, we selected the distortion styles that are close in the character feature space according to the Euclidean distance between the means of the same class in different styles. The number of neighbors was selected from {10,20,30,40,50}.

The classifiers were trained using the original samples only and remain unchanged in perturbation-based recognition.

B. Performance and Analysis

For each classifier, we compare the character-level accuracies on fields of test samples of variable length, and in distortion neighborhood search, using variable size of neighborhood. The results of classifiers MQDF and NPC are shown in Table I and Table II, respectively. In the tables, the column "Base" gives the baseline performance (given by classification without perturbation), the 3rd columns gives the results of ordinary perturbation (OP), the 4th column gives the results of style consistent perturbation (SCP), and the 5-7th columns give the results of SCP with neighbor search of neighborhood size *n* (SCP+*n*).

The results of MQDF show that perturbation always improves the recognition performance. However, the SCP gives lower accuracy than the ordinary individual perturbation. By distortion neighborhood search with SCP, the accuracy is improved significantly and the resulting accuracy is higher than that of ordinary perturbation. Compared to the baseline, the accuracy is improved by nearly 1%, i.e., the error reduction rate is nearly 10%.

TABLE I. TEST ACCURACIES (%) USING MQDF CLASSIFIER.

Field length	Base	OP	SCP	SCP+10	SCP+20	SCP+30
5	89.55	90.20	89.87	90.28	90.37	90.41
10			89.96	90.30	90.43	90.46
20			89.91	90.34	90.43	90.50
30			89.90	90.35	90.45	90.50

TABLE II. TEST ACCURACIES (%) USING NEAREST PROTOTYPE CLASSIFIER (NPC)

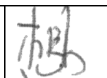
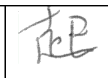
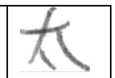
Field length	Base	OP	SCP	SCP+30	SCP+40	SCP+50
5	85.30	88.01	87.15	88.30	88.35	88.39
10			86.88	88.34	88.40	88.49
20			86.70	88.31	88.37	88.43

The results of NPC show similar tendency with those of MQDF. Compared to the baseline performance, the SCP with neighborhood search resulted error reduction rate of over 20%.

Our results show that SCP with field length 5 gives significant performance improvement and the performance only improves slightly when the field length increases from 5. Compared to the style normalization-based field classification method in [16], our SCP method achieves similar error reduction in very short fields (the field length in [16] is over 1,000).

Table III shows fives character samples of the same writer that baseline classifier and perturbation, SCP gives different results. In the table each recognition method gives three top-rank classes on a sample. On the five samples, baseline recognition gives only one correct top-rank result, ordinary perturbation (PO) gives three correct results, while SCP and SCP+30 gives four and five correct results, respectively. By the baseline classifier, three mis-recognized samples have the correct class in the second rank, which is promoted to the top rank by SCP.

TABLE III. RECOGNITION RESULTS OF A FIVE-CHARACTR FIELD BY DIFFERENT METHODS. EACH METHOD SHOWS THREE TOP-RANK CLASSES.

					
Base	棍**想**撼	超**起**迎	未尘束	部**都**郡	**太**大木
OP	**想**拙棍	褪超迎	求水束	**都**郡部	**太**木大
SCP	**想**热棍	枉**起**征	**来**束未	**都**部郡	**太**大木
SCP+30	**想**棍抽	**起**超迎	**来**束未	**都**部郡	**太**大木

VI. CONCLUSION

A style-consistency perturbation method is proposed for improving the recognition performance of handwritten Chinese character. This method explores the style consistency of the characters in a field to force them undergo the same distortion. The distortion neighborhood search method also allows slight style deviation from a uniform distortion style. The experimental results in handwritten Chinese character recognition show that the proposed method improves character recognition accuracy significantly compared to the baseline classifier and ordinary individual perturbation.

In the future work, the performance of style consistent perturbation-based recognition can be further improved via several ways: using better distortion models, training classifier with distorted samples, and fusing the classification output of multiple distortions.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (NSFC) under grants 61175021 and 60933010.

REFERENCES

[1] H. Fujisawa, "Forty years of research in character and document recognition: An industrial perspective", *Pattern Recognition*, 41(8): 2435-2446, 2008.

[2] H. Liu and X. Ding, "handwritten character recognition using gradient feature and quadratic classifier with multiple discrimination schemes," *Proc. 8th ICDAR*, pp. 19-23, *2005.*

[3] T.-F. Gao and C.-L. Liu, "High accuracy handwritten Chinese character recognition using LDA-based compound distance," *Pattern Recognition*, 41(11): 3442-3451, 2008.

[4] C.-L Liu, F. Yin, D.-H Wang and Q.-F Wang, "Online and offline handwritten Chinese character recognition: Benchmarking on new databases," *Pattern Recognition*, 46(1): 155-162 , 2013.

[5] C.-L. Liu, F. Yin, Q.-F. Wang and D.-H. Wang, "ICDAR 2011 Chinese Handwriting Recognition Competition," *Proc. 11th ICDAR*, pp. 1464-1469, 2011.

[6] T.-H. Su, T.-W. Zhang, D.-J. Guan and H.-J. Huang, "Off-line recognition of realistic Chinese handwriting using segmentation-free strategy," *Pattern Recognition*, 42(1): 167-182, 2009.

[7] T. Varga, H, Bunke, "Generation of synthetic training data for an HMM-based handwriting recognition System", *Proc. 7th ICDAR*, 2003, pp.618-622.

[8] G. Chen, H.-G. Zhang and J. Guo, "Learning pattern generation for handwritten Chinese character using pattern transform method with cosine function", *Proc. 5th ICMLC*, pp.13-16, 2006.

[9] K.-C. Leung and C.-H. Leung, "Recognition of handwritten Chinese characters by combining regularization, Fisher's discriminant and distorted sample generation", *Proc. 10th ICDAR*, 2009, pp.1026-1030.

[10] P. Sarkar and G. Nagy, "Style consistent classification of isogenous patterns", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(1): 88-98, 2005.

[11] S. Veeramachaneni and G. Nagy, "Style context with second-order statistics", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(1): 14-22, 2005.

[12] D. Ciresan, U. Meier and J. Schmidhuber, "Multi-column deep neural networks for image classification," *IEEE Conf. on CVPR*, pp. 3643-3649, 2012.

[13] M. Yasuda, K. Yamamoto and H. Yamada, "Effect of the perturbation correlation method for optical character recognition," *Pattern Recognition*, 30(8): 1315-1320, 1997.

[14] T. M. Ha and H. Bunk, "Off-line, handwritten numeral recognition by perturbation method," *IEEE Trans. Patern Analysis and Machine Intelligence*, 19(5): 535-539, 1997.

[15] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang, "CASIA online and offline Chinese handwriting databases," *Proc. 11th ICDAR*, pp. 37-41, 2011.

[16] X.-Y Zhang, K.-Z. Huang and C,-L. Liu, "Pattern field classification with style normalized transform," *Proc. 22th IJCAI*, vol. 2, pp. 1621-1626, 2011.

[17] C.-L. Liu, "Classifier combination based on confidence transformation", *Pattern Recognition*, 38(1): 11-28, 2005.

[18] Q.-F. Wang, F. Yin and C.-L. Liu, "Improving handwritten Chinese text recognition by confidence transformation," *Proc. 11th ICDAR*, pp. 518-522, 2011.

[19] X.-B. Jin, C.-L. Liu, X. Hou, "Regularized margin-based conditional log-likelihood loss for prototype learning", *Pattern Recognition*, 43(7): 2428-2438, 2010.

[20] F. Kimura, K. Takashina, S. Tsuruoka, Y. Miyake, "Modified quadratic discriminant functions and the application to Chinese character recognition", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(1): 149-153, 1987.

[21] C.-L Liu, "Normalization-cooperated gradient feature extraction for handwritten character recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(8): 1456-1469, 2007.