Writer Code Based Adaptation of Deep Neural Network for Offline Handwritten Chinese Text Recognition

Zi-Rui Wang and Jun Du
University of Science and Technology of China
Hefei, Anhui, P. R. China
Email: cs211@mail.ustc.edu.cn, jundu@ustc.edu.cn

***Abstract*—Recently, we propose deep neural network based hidden Markov models (DNN-HMMs) for offline handwritten Chinese text recognition. In this study, we design a novel writer code based adaptation on top of the DNN-HMM to further improve the accuracy via a customized recognizer. The writer adaptation is implemented by incorporating the new layers with the original input or hidden layers of the writer-independent DNN. These new layers are driven by the so-called writer code, which guides and adapts the DNN-based recognizer with the writer information. In the training stage, the writer-aware layers are jointly learned with the conventional DNN layers in an alternative manner. In the recognition stage, with the initial recognition results from the first-pass decoding with the writer-independent DNN, an unsupervised adaptation is performed to generate the writer code via the cross-entropy criterion for the subsequent second-pass decoding. The experiments on the most challenging task of ICDAR 2013 Chinese handwriting competition show that our proposed adaptation approach can achieve consistent and significant improvements of recognition accuracy over a high-performance writer-independent DNN-HMM based recognizer across all 60 writers, yielding a relative character error rate reduction of 23.62% in average.**

***Keywords*—Offline handwritten Chinese text, adaptation, writer code, DNN-HMM.**

I. Introduction

Historically, Chinese character recognition has been extensively studied [1], [2], [3], [4]. In the mobile internet era, the research on the robust recognition of handwritten Chinese characters in an unconstraint manner has become increasingly popular due to the application demand. In terms of the task complexity [5], [6], the offline handwritten Chinese text recognition is the most challenging task due to the lack of trajectory information and the free writing style, which is also the topic of this study.

The research efforts for offline handwritten Chinese text recognition can be divided into two categories, namely oversegmentation-based and segmentation-free approaches. The former one [7], [8], [9], [10] often builds several modules first including character oversegmentation, character classification, modeling the linguistic and geometric contexts, and then incorporate them for calculating the score for path search. The most effective one in [10] is to integrate multiple contexts for recognition of handwritten Chinese text line, which achieves the best recognition performance on the ICDAR 2013 Chinese handwriting competition task. Compared with the diversified oversegmentation-based approaches, there are not many segmentation-free approaches. One early approach to the text line modeling [11] adopts the Gaussian mixture model based hidden Markov model (GMM-HMM). Another recent approach [12] uses multidimensional long-short term memory recurrent neural network (MDLSTM-RNN), which is inspired by well verified LSTM-RNN approaches [13] for the recognition of handwritten western languages with a small set of character classes. The MDLSTM-RNN approach is quite flexible as connectionist temporal classification (CTC) technique [14] is adopted to avoid the explicit segmentation. And it can achieve the comparable recognition accuracy with the best oversegmentation-based approach [10]. More recently, we propose a novel segmentation-free approach [15] using deep neural network based hidden Markov model (DNN-HMM) [16], [17] for offline handwritten Chinese text recognition, which can yield the best reported results on the ICDAR 2013 competition task.

However, in real applications the recognition accuracy is not always satisfactory due to the different writing styles, especially the cursive and continuous writing styles. To address this problem, one possible solution is the normalization strategy to solve the shape variation, including the simple linear normalization, nonlinear normalization [18] or other normalization strategy such as the aspect ratio adaptive normalization (ARAN) [19]. Another solution is writer adaptation which aims to improve the recognition performance and user experience of a single writer by using the corresponding data samples to be recognized itself via an unsupervised adaptation strategy, or by using a small of adaptation data samples with labels collected from the target writer via a supervised adaptation strategy. For the past several decades, many writer adaptation approaches designed based on different models have been investigated. For example, the writer adaptive structures are incorporated with neural network based classifiers [20], [21]. In [22], a support vector machine (SVM) based classifier with a biased regularization is adopted for personalization. Furthermore, the writer adaptation via maximum likelihood linear regression (MLLR) or maximum a posteriori (MAP) criterion is conducted for an HMM based recognition system for cursive German script [23]. For online Chinese handwritten

2167-6445/16 $31.00 © 2016 IEEE

DOI 10.1109/ICFHR.2016.99

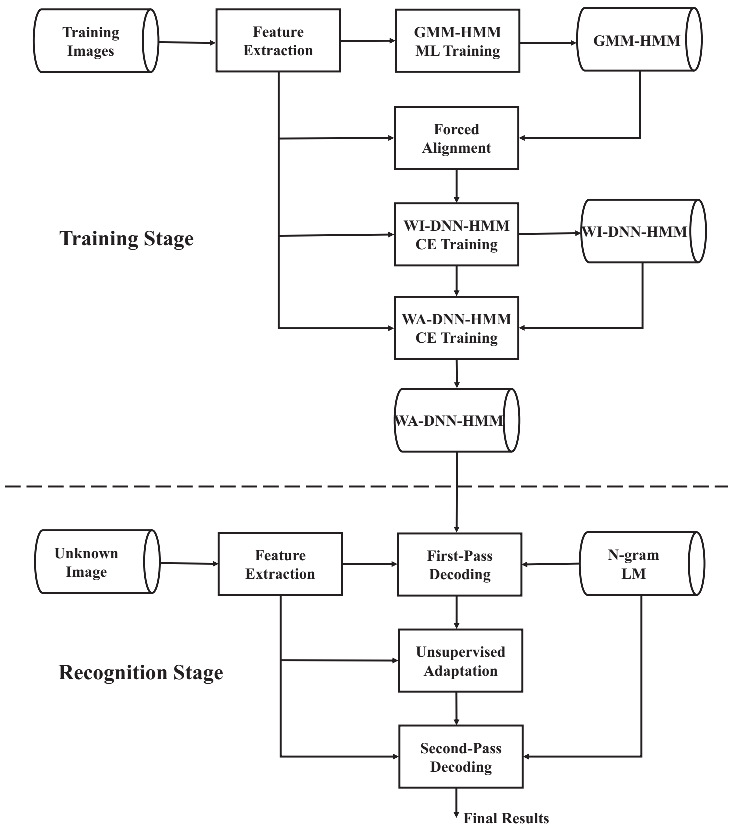

Fig. 1. A block diagram of the proposed system.

character recognition, one type of the adaptation methods is to use a linear feature transformation for adapting the writing styles via different criteria, e.g., style transfer mapping (STM) [24] or discriminative linear regression (DLR) [25], [26], which are verified to be effective incorporated with prototype-based classifier, DNN-based classifier, and convolutional neural network (CNN) based classifier.

In this study, for the most challenging task, namely the offline handwritten Chinese text recognition, we design a novel writer code based adaptation on top of a high-performance DNN-HMM based recognizer to further improve the accuracy. The main idea is motivated by the recent work of speaker adaptation in the speech recognition area [27], [28], [29]. The writer adaptive structures are implemented by linking the new layers to the writer-independent DNN. These new layers are driven by a writer-dependent vector, namely the writer code, which adapts the DNN-HMM based recognizer with the writer information. In the training stage, the writer-aware layers are jointly learned with the conventional DNN layers using back-propagation algorithm [30] in an alternative manner. In the recognition stage, with the initial recognition results from the first-pass decoding using the writer-independent DNN, an unsupervised adaptation is performed to generate the writer code via the cross-entropy criterion for the subsequent second-pass decoding. The experiments on the ICDAR 2013 Chinese handwriting competition task show that our proposed adaptation approach can yield consistent and significant improvements of recognition accuracy over a best reported DNN-HMM based recognizer.

The rest of this paper is organized as follows. In Section II we introduce the overall system architecture. In Section III we elaborate the writer-independent DNN-HMM recognizer. In Section IV we describe the proposed writer coder based adaptation DNN-HMM in detail. In Section V we report experimental results and finally we summarize our work in Section VI.

II. System Overview

The overall flowchart of our proposed system is shown in Fig. 1. In the training stage, we build three HMM systems, namely GMM-HMM [31], WI-DNN-HMM and WA-DNN-HMM. WI-DNN-HMM refers to the writer-independent DNN-HMM system while WA-DNN-HMM represents writer-adaptive DNN-HMM system. First, the gradient-based features on a frame-level are extracted, which is followed by principal component analysis (PCA) transformation to obtain to a lower dimensional feature vector. Then the parameters of GMM-HMMs for all character classes are trained using

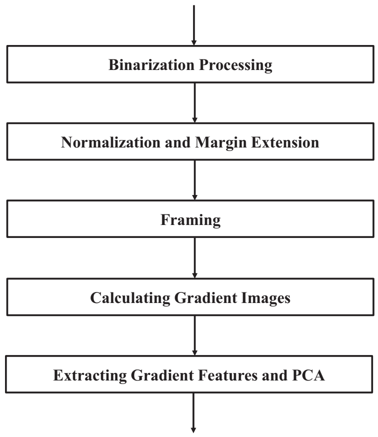

Fig. 2. The procedure of feature extraction.

maximum likelihood (ML) estimation. With the well-trained GMM-HMMs, the state-level forced-alignment is conducted to obtain the frame-level labels used for the subsequent DNN training. As for the WI-DNN-HMM system, the WI-DNN model is learned using the cross-entropy (CE) criterion via the back-propagation algorithm. On top of the WI-DNN, the writer adaptive structures based on the writer code are attached, which are optimized using the same CE criterion alternatively with the writer-independent layers. Finally, the WA-DNN model for the WA-DNN-HMM system is generated with both writer-independent and writer-code-driven layers. In the recognition stage, after the feature extraction of the unknown handwritten text lines, the first-pass decoding results are generated by using a weighted finite-state transducer (WFST) [32], [33] based decoder integrating both WI-DNN-HMM character model and n-gram language model (LM). Based on the recognition results, the unsupervised writer adaptation is performed to update the writer coder of WA-DNN, which is followed by the second-pass decoding to generate final recognition results. The details of WI-DNN-HMM and WA-DNN-HMM are elaborated in the following sections.

III. The WI-DNN-HMM based Recognizer

According to our recent work [15], the WI-DNN-HMM based recognizer has demonstrated its superiority over the other conventional approaches for recognition of offline handwritten Chinese text line. In this section, three important modules, namely feature extraction, HMM, and WI-DNN are briefly reviewed as follows.

A. Feature Extraction

The procedure of feature extraction, illustrated in Fig. 2, consists of the following steps.

- **Step1**: **Binarization Processing**
  The Otsu's method [34] is used for binarization of the original text line image.
- **Step2**: **Normalization and Margin Extension**
  First, we estimate the height of the text line while keeping the aspect ratio. Then the margin is extended to accommodate the text area for the sliding windows in the next step. Finally, the center line is calculated.
- **Step3**: **Framing**
  Along the centre line, each frame, represented by a 64×32 sliding window from the left to right, with a frame shift of 3 pixels, is scanned across the text line.
- **Step4**: **Calculating Gradient Images**
  For each frame, 8-directional gradient images can be generated according to [35] via the Sobel operator.
- **Step5**: **Extracting Gradient Features and PCA**
  Based on all 8 gradient images, a 256-dimensional feature vector can be obtained by spatial sampling and blurring techniques in [36], followed by the PCA transformation [37] to generate a lower 50-dimensional feature vector fed to the subsequent character modeling.

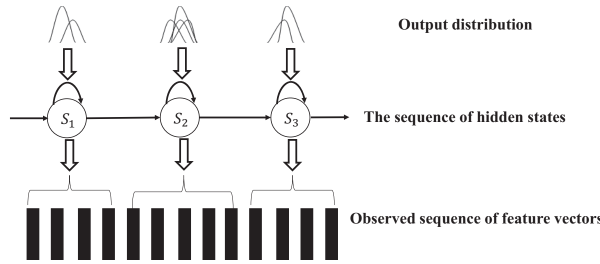

Fig. 3. A typical HMM representing one character class.

B. HMM

To model a sequence of Chinese characters, we adopt a left-to-right HMM to model each character class as a basic unit. Accordingly, the text line is modeled by a sequence of HMMs. A typical HMM [38] is illustrated in Fig. 3 for one handwritten character class with a set of invisible states. For the feature sequence extracted from one handwritten character sample, each frame is supposed to be assigned to one underlying state. For each state, an output distribution describes the statistical property of the observed feature vector, which is represented by a GMM for GMM-HMM system. For DNN-HMM system, a DNN, elaborated in the next section, is adopted to calculate the state posterior probability which can be easily converted to the observation probability given each state. Furthermore, in our implementation, the HMM topology for DNN-HMM system is copied from that of GMM-HMM system, including the state prior probabilities and transition probabilities.

C. WI-DNN

The WI-DNN is a writer-independent DNN directly trained using the data samples of all writers, as shown in Fig. 4. In this study, WI-DNN aims to model the posterior probability

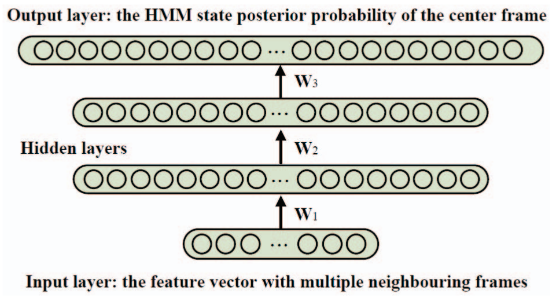

Fig. 4. The architecture of WI-DNN.

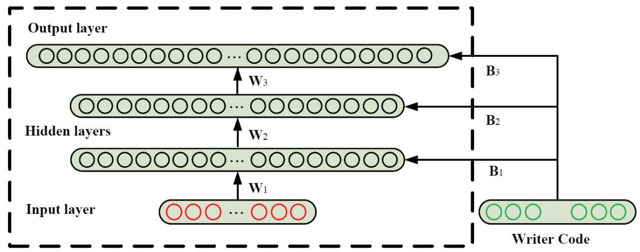

Fig. 5. The architecture of WA-DNN.

of the HMM states for all character classes using the input feature vector with multiple neighbouring frames. The activation function of each hidden layer is the logistic sigmoid function. After the random initialization of all the weight parameters, a supervised fine-tuning is conducted. For this multiclass classification task, output unit j converts its total input z_j into a class probability p_j by using the "softmax" non-linearity:

$$p_j = \frac{\exp(z_j)}{\sum_k \exp(z_k)} \tag{1}$$

where k is an index over all classes. Then the objective function is the cross-entropy between the target probabilities $\bar{p}$ and the outputs of the softmax p:

$$C = -\sum_j \bar{p}_j \log p_j \tag{2}$$

where the target probabilities, taking values of one or zero, are the supervised information provided to train the DNN classifier. As our task involves large training samples, the objective function is optimized using back-propagation procedure with stochastic gradient descent in mini-batch mode. For more details of WI-DNN-HMM training, the readers can refer to [15].

IV. Writer Code based Adaptation

To perform the writer adaptation, we design a writer-adaptive structure incorporated with WI-DNN, namely WA-DNN as illustrated in Fig. 5. The structure in the dashed box is exactly the WI-DNN with the corresponding parameters fixed during the WA-DNN training. For the newly added structure, the input is a K-dimensional vector named as the writer code to represent the information of a specific writer. For each of the hidden layers and output layer in WI-DNN, a new link is created from the writer code. For the l-th layer, the new relationship between the input vector $\boldsymbol{x}_{l-1}$ (actually the output of the $(l-1)$-th layer) and the output vector $\boldsymbol{y}_l$ before applying the activation or softmax function can be expressed as:

$$\boldsymbol{y}_l = \boldsymbol{W}_l\boldsymbol{x}_{l-1} + \boldsymbol{b}_l + \boldsymbol{B}_l\boldsymbol{w}(c), \quad 1 \le l \le L+1 \tag{3}$$

where $\boldsymbol{W}_l$ and $\boldsymbol{b}_l$ are the weight matrix and bias vector of l-th layer in the WI-DNN. $\boldsymbol{B}_l$ is a weight matrix linked from the writer coder $\boldsymbol{w}(c)$ of c-th writer. L is the number of hidden layers. Obviously, if $\boldsymbol{w}(c)$ is set to a zero vector, WA-DNN degenerates into WI-DNN.

In the training stage, by adopting the same CE criterion as in Eq. (2), the linking matrices $\boldsymbol{B}$ and the writer codes are simultaneously optimized with the random initialization. With the back-propagation algorithm, the gradients of the objective function C with respect to the linking matrices are calculated as:

$$\frac{\partial C}{\partial B_{lmk}} = \frac{\partial C}{\partial y_{lm}} \times \frac{\partial y_{lm}}{\partial B_{lmk}} = \delta_{lm} w_k(c) \tag{4}$$

where B_{lmk} is the $(m,k)^{\text{th}}$ element of $\boldsymbol{B}_l$ while y_{lm} is the m^{th} component of $\boldsymbol{y}_l$. δ_{lm} is the error signal of the m-th node in the l-th layer. $w_k(c)$ is the k-th component of the writer coder $\boldsymbol{w}(c)$. Similarly, the gradients of the objective function C with respect to the writer codes can be derived:

$$\frac{\partial C}{\partial w_k(c)} = \frac{1}{L+1}\sum_l \sum_m \delta_{lm} B_{lmk} \tag{5}$$

where the scaling of gradients by the number of layers is a good strategy to control the dynamic range. After the training procedure, only the linking matrices are stored as parameters and the learned writer codes for all writers are useless for the unsupervised adaptation with an unseen writer.

In the recognition stage, with the first-pass recognition results of all samples from a specific writer, the unsupervised adaptation is conducted to only update the writer code with the CE criterion by fixing both $\boldsymbol{W}$ and $\boldsymbol{B}$.

V. Experimental Results and Analysis

The experiments are conducted on the public CASIA-HWDB database [39]. The training set consists of HWDB1.0, HWDB1.1, HWDB2.0, HWDB2.1, and HWDB2.2 datasets. HWDB1.0 and HWDB1.1 are offline isolated handwritten Chinese character datasets while HWDB2.0-HWDB2.2 are offline handwritten Chinese text datasets. In total, there are 3,980 classes (Chinese characters, symbols, garbage) with 4,091,599 samples. Here "garbage" classes represent the short blank model between characters and the long blank model at the beginning or end of the text line. As for the evaluation set, the ICDAR 2013 competition set with 60 writers is adopted [6]. The 50-dimensional feature vector is directly used for GMM-HMM system while an augmented version of 7 frames is fed to DNN-HMM system.

For GMM-HMM system, each character class is modeled by a left-to-right HMM with 5 states. For each state, a GMM with

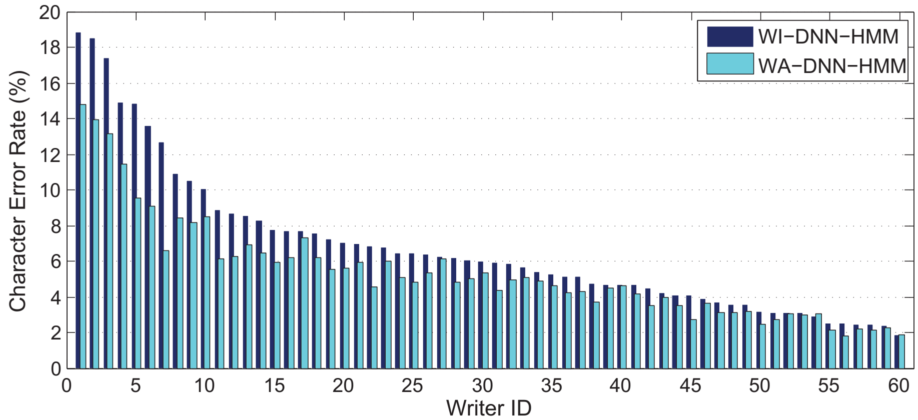

Fig. 6. Performance comparison of WI-DNN-HMM and WA-DNN-HMM with K=1000 on the test set of all 60 writers.

TABLE I
PERFORMANCE COMPARISON OF WI-DNN-HMM AND WA-DNN-HMM WITH DIFFERENT DIMENSIONS OF WRITER CODE ACROSS ALL THE WRITERS ON THE TEST SET.

System	CER	Relative CER Reduction
WI-DNN-HMM	6.73%	-
WA-DNN-HMM (K=100)	6.01%	10.70%
WA-DNN-HMM (K=200)	5.82%	13.52%
WA-DNN-HMM (K=300)	5.62%	16.49%
WA-DNN-HMM (K=500)	5.36%	20.36%
WA-DNN-HMM (K=800)	5.17%	23.18%
WA-DNN-HMM (K=1000)	5.14%	23.62%

40 Gaussian mixtures is used. The total number of Gaussians is $3,980 * 5 * 40 = 796,000$. For WI-DNN-HMM system, the input size of DNN is 350 while the output size is $3,980 * 5 = 19,900$ corresponding to the number of states of all classes. 6 sigmoidal hidden layers with 2048 nodes for each layer are used. The mini-batch size is 256. The initial step size is set to 0.008 which is halved after each iteration if the loss of cross-validation set is reduced. 16 iterations are conducted. For WA-DNN-HMM, the mini-batch size is 128 and the learning rate is set to a fixed value of 0.4.

For the decoding, a 3-gram language model is generated by using different text sources [15]. The Kaldi tool [40] is adopted for both training and testing. The evaluation measure is the character error rate (CER), which is the ratio between the total number of substitution/insertion/deletion errors and the total number of character samples in the evaluation set.

Table I lists a performance comparison of WI-DNN-HMM and WA-DNN-HMM with different dimensions of writer code across all the writers on the test set. First, the WI-DNN-HMM system yields the best reported recognition results on the competition set without adaptation techniques. Second, all WA-DNN-HMM systems using writer coder based adaptation can achieve the relative CER reductions of more than 10% over WI-DNN-HMM system. However, with the increase of the dimension K from 100, the CER is significantly reduced and saturated at K=1000, yielding a 1.59% absolute error reduction (a 23.62% relative error reduction) in average. By considering that we conduct the unsupervised adaptation without any supervised information and the baseline WI-DNN-HMM system already achieves a high recognition performance, our proposed writer adaptation approach seems quite effective.

Fig. 6 gives a performance comparison of WI-DNN-HMM and WA-DNN-HMM with K=1000 on the test set of all 60 writers sorted by the recognition accuracy of WI-DNN-HMM system. Clearly, significant and consistent improvements of recognition accuracy, especially for the cases with high CERs, can be observed for most of the writers, which indicates that our approach is robust for different writing styles. There are only two exceptions on the No.54 and No.60 writers, which seem reasonable as it is quite challenging for the unsupervised adaptation approach to generate additional gains over the systems with 2%-3% CER. While the CERs of WI-DNN-HMM system range from 18.87% to 1.90% for different writers, our proposed WA-DNN-HMM system achieves better CERs from 14.09% to 1.75%. For the No.7 writer, almost a half of the error rate is reduced with the CER from 12.68% to 6.50%.

VI. CONCLUSION

In this study, we propose an effective adaptation approach by incorporating the writer-coder-driven layers with the conventional DNN model for offline handwritten Chinese text recognition. And on the ICDAR 2013 competition task, the average CER of the proposed approach is close to 5% making a new milestone. As for the future work, more types of neural networks, e.g. CNN and LSTM, with more objective functions will be investigated to further demonstrate the effectiveness of the adaptation approach.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grants No. 61305002, National Key Technology Support Program under Grants No. 2014BAK15B05, the “Strategic Priority Research Program” of the Chinese Academy of Sciences under Grant No. XD-B02070006, MOE-Microsoft Key Laboratory of USTC. The

authors would like to thank Mr. Zhi-Ying Huang for the contributions on implementation.

REFERENCES

[1] R. Casey and G. Nagy, "Recognition of printed Chinese characters," *IEEE Trans. on Electronic Computers*, Vol. 15, No. 1, pp.91-101, 1966.

[2] C.-L. Liu, S. Jaeger, and M. Nakagawa, "Online recognition of Chinese characters: the state-of-the-art," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 2, pp.198-213, 2004.

[3] R.-W. Dai, C.-L. Liu, and B.-H. Xiao, "Chinese character recognition: history, status and prospects," *Frontiers of Computer Science in China* Vol. 1, No. 2, pp.126-136, 2007.

[4] H. Fujisawa, "Forty years of research in character and document recognition–an industrial perspective," *Pattern Recognition*, Vol. 41, No. 8, pp.2435-2446, 2008.

[5] C.-L. Liu, F. Yin, Q.-F. Wang, and D.-H. Wang, "ICDAR 2011 Chinese handwriting recognition competition," *Proc. ICDAR*, 2011, pp.1464-1469.

[6] F. Yin, Q.-F. Wang, X.-Y. Zhang, and C.-L. Liu, "ICDAR 2013 Chinese handwriting recognition competition," *Proc. ICDAR*, 2013, pp.1464-1470.

[7] C.-H. Wang, Y. Hotta, M. Suwa, and S. Naoi, "Handwritten Chinese address recognition," *Proc. IWFHR*, 2004, pp.539-544.

[8] Q. Fu, X.-Q. Ding, T. Liu, Y. Jiang, and Z. Ren, "A novel segmentation and recognition algorithm for Chinese handwritten address character strings," *Proc. ICPR*, 2006, pp.974-977.

[9] N.-X. Li and L.-W. Jin, "A Bayesian-based probabilistic model for unconstrained handwritten offline Chinese text line recognition," *Proc. IEEE SMC*, 2010, pp.3664-3668.

[10] Q.-F. Wang, F. Yin, and C.-L. Liu, "Handwritten Chinese text recognition by integrating multiple contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 34, No. 8, pp.1469-1481, 2012.

[11] 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*, Vol. 42, No. 1, pp.167-182, 2009.

[12] R. Messina and J. Louradour, "Segmentation-free handwritten Chinese text recognition with LSTM-RNN," *Proc. ICDAR*, 2015, pp.171-175.

[13] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for improved unconstrained handwriting recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 5, pp.855-868, 2009.

[14] A. Graves, S. Fernandez, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," *Proc. ICML*, 2006, pp.369-376.

[15] J. Du, Z.-R. Wang, J.-F. Zhai, and J.-S. Hu, "Deep neural network based hidden markov model for offline handwritten Chinese text recognition," *Submitted to Proc. ICPR 2016*.

[16] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, Vol. 313, No. 5786, pp.504-507, 2006.

[17] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Processing Magazine*, Vol. 29, No. 6, pp.82-97, 2012.

[18] J. Tsukumo and H. Tanaka, "Classification of handprinted Chinese characters using nonlinear normalization and correlation methods," *Proc. ICPR*, 1988, pp.168-171.

[19] C.-L. Liu, M. Koga, H. Sako, and H. Fujisawa, "Aspect ratio adaptive normalization for handwritten character recognition," *Proc. ICMI*, 2000, pp.418-425.

[20] N. Matić, I. Guyon, J. Denker, and V. Vapnik, "Writer adaptation for on-line handwritten character recognition," *Proc. ICDAR*, 1993, pp.187-191.

[21] J. C. Platt and N. P. Matić, "A constructive RBF network for writer adaptation," *Proc. NIPS*, 1997.

[22] W. Kienzle and K. Chellapilla, "Personalized handwriting recognition via biased regularization," *Proc. ICML*, 2006.

[23] A. Brakensiek, A. Kosmala, and G. Rigoll, "Comparing adaptation techniques for on-line handwriting recognition," *Proc. ICDAR*, 2001, pp.486-490.

[24] X.-Y. Zhang and C.-L. Liu, "Writer adaptation with style transfer mapping," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 7, pp.1773-1787, 2013.

[25] J. Du, J.-S. Hu, B. Zhu, S. Wei, and L.-R. Dai, "Writer adaptation using bottleneck features and discriminative linear regression for online handwritten Chinese character recognition," *Proc. ICFHR*, 2014, pp.311-316.

[26] J. Du, J.-F. Zhai, J.-S. Hu, B. Zhu, S. Wei, and L.-R. Dai, "Writer adaptive feature extraction based on convolutional neural networks for online handwritten Chinese character recognition," *Proc. ICDAR*, 2015, pp.841-845.

[27] G. Saon, H. Soltau, D. Nahamoo, and M. Picheny, "Speaker adaptation of neural network acoustic models using i-vectors," *Proc. ASRU*, 2013, pp.55-59.

[28] O. Abdel-Hamid and H. Jiang, "Fast speaker adaptation of hybrid NN/HMM model for speech recognition based on discriminative learning of speaker code," *Proc. ICASSP*, 2013, pp.7942-7946.

[29] S.-F. Xue, O. Abdel-Hamid, H. Jiang, L.-R. Dai, and Q.-F. Liu, "Fast adaptation of deep neural network based on discriminant codes for speech recognition," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, Vol. 22, No. 12, pp.1713-1725, 2014.

[30] D. E. Rumelhart, J. L. McClelland, and PDP Research Group, *Parallel Distributed Processing*, MIT Press, 1987.

[31] M. Gales and S. Young, "The application of hidden Markov models in speech recognition," *Foundations and Trends in Signal Processing*, Vol. 1, No. 3, pp.195-304, 2008.

[32] M. Mohri, F. Pereira, and M. Riley, "Weighted finite-state transducers in speech recognition," *Computer Speech and Language*, Vol. 20, No. 1, pp.69-88, 2002.

[33] C. Allauzen, M. Riley, J. Schalkwyk, W. Skut, and M. Mohri, "OpenFst: a general and efficient weighted finite-state transducer library," *Proc. CIAA*, 2007, pp.11-23.

[34] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 9, No. 1, pp.62-66, 1979.

[35] C.-L. Liu, "Normalization-cooperated gradient feature extraction for handwritten character recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 29, No. 8, pp.1465-1469, 2007.

[36] Z.-L. Bai and Q. Huo, "A study on the use of 8-directional features for online handwritten Chinese character recognition," *Proc. ICDAR*, 2005, pp.262-266.

[37] A. C. Rencher, *Methods of Multivariate Analysis*, Wiley, New York, 2002.

[38] L. E. Baum and J. A. Eagon, "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology," *Bulletin of American Mathematical Society*, Vol. 73, No. 2, pp.360-363, 1967.

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

[40] D. Povey, A. Ghoshal, et al., "The kaldi speech recognition toolkit," *Proc. ASRU*, 2011.