ICDAR 2013 Chinese Handwriting Recognition Competition

Fei Yin, Qiu-Feng Wang, Xu-Yao Zhang, Cheng-Lin Liu
National Laboratory of Pattern Recognition (NLPR)
Institute of Automation of Chinese Academy of Sciences, Beijing 100190, China
E-mail: {fyin, wangqf, xyz, liucl}@nlpr.ia.ac.cn

***Abstract*—This paper describes the Chinese handwriting recognition competition held at the 12th International Conference on Document Analysis and Recognition (ICDAR 2013). This third competition in the series again used the CASIA-HWDB/OLHWDB databases as training set, and all the submitted systems were evaluated on closed datasets to report character-level correct rates. This year, 10 groups submitted 27 systems for five tasks: classification on extracted features, online/offline isolated character recognition, online/offline handwritten text recognition. The best results (correct rates) are 93.89% for classification on extracted features, 94.77% for offline character recognition, 97.39% for online character recognition, 88.76% for offline text recognition, and 95.03% for online text recognition, respectively. In addition to the test results, we also provide short descriptions of the recognition methods and brief discussions.**

Keywords—Chinese handwriting recognition competition; isolated character recongition; handwritten text recognition; offline; online; CASIA-HWDB/OLHWDB database.

I. INTRODUCTION

Despite the tremendous works in the past 40 years, researches on Chinese handwriting recognition, including online (stroke trajectory-based) and offline (image-based) recognition of both isolated characters and continuous texts, is still a challenge. To stimulate the research in this field, the National Laboratory of Pattern Recognition (NLPR), Institute of Automation of Chinese Academy of Sciences (CASIA), released large databases of free handwriting: CASIA-HWDB (offline) and CASIA-OLHWDB (online) [1], and organized two competitions at Chinese Conference on Pattern Recognition 2010 (CCPR 2010) [2] and 11th International Conference on Document Analysis and Recognition (ICDAR 2011) [3]. Significant improvements can be observed in these past competitions.

The recognition tasks of the third competition in 2013 include those four tasks evaluated in 2011: online and offline isolated character recognition, online and offline handwritten text recognition. In addition, a new task of classification on extracted features was launched to evaluate classification algorithms on standard feature data [4]. Targeting the five tasks, 27 systems were submitted by 10 groups, three of which participated in the competition 2011 as well, while seven are new. Again, the recognition performance was evaluated on closed datasets. Executable software systems were sent to the organizer and tested in same environment. Standard training datasets were recommended but the participants are free to use any training data. Accompanying the submitted system, the participants were requested to submit a brief description of their methods.

For reference, the first Chinese handwriting recognition competition in 2010 evaluated only online/offline isolate character recognition and received nine systems submitted by four groups [2]. The second competition in 2011 evaluated both isolated character recognition and handwritten text recognition, and received 25 systems submitted by eight groups [3].

In the following, we first describe the databases and evaluation protocol in Section 2; Section 3 describes the recognition methods of the submitted systems; Section 4 presents the evaluation results and Section 5 provides concluding remarks.

II. DATABASES AND EVALUATION PROTOCOL

For promoting the research of Chinese handwriting recognition, we released the large databases CASIA-HWDB/OLHWDB [1], which are free for academic research. The competition participants were encouraged to use the released databases for training their recognition systems, and can use any additional private or open datasets and distorted samples for enhancement. We reserve the un-open test datasets for evaluating the submitted systems in competition, and rank the systems according to the character-level correct rate.

A. CASIA Databases

The databases CASIA-HWDB and CASIA-OLHWDB contain offline/online handwritten characters and continuous texts written by 1,020 persons using Anoto pen on papers, such that the online and offline data were produced concurrently. Either the (offline) HWDB or the (online) OLHWDB contain six datasets: three for isolated characters (DB1.0–1.2) and three for handwritten texts (DB2.0–2.2). The datasets of isolated characters contain about 3.9 million samples of 7,356 classes (7,185 Chinese characters and 171 symbols), and the datasets of handwritten texts contain about 5,090 pages and 1.35 million character samples. All the data has been segmented and annotated at character level, and each dataset is partitioned into standard training and test subsets. More details of the databases can be found in [1].

B. Test Datasets

The test datasets which are unknown to all participants were collected for the Competition 2010 [2]. They were written by 60 writers who did not contribute to the released CASIA-HWDB/OLHWDB databases.

For evaluating isolated character recognition, we confine the character set to the 3,755 Chinese characters in the level-1 set of GB2312-80, which has been popularly tested in Chinese character recognition research. The handwritten text data was produced by hand-copying natural language texts on un-formatted pages. The texts in the test dataset are different from those in the databases CASIA-HWDB/OLHWDB. The characters in the texts are mostly within the set of 7,356 classes of the isolated character datasets (DB1.0-DB1.2) in CASIA-HWDB/OLHWDB. Table 1 shows the statistics of the test datasets, we can see that the online and offline data of concurrently written texts have slightly different numbers of character samples because of some data loss in the digital ink or scanned images. In this table, "#outlier" denotes the number of characters beyond the 7,356 classes.

Table 1. Statistics of the closed test datasets.

	Isolate characters		Continuous texts	
	online	offline	online	offline
#writer	60	60	60	60
#class	3,755	3,755	1,375	1,385
#text line			3,432	3,432
#sample	224,590	224,419	91,576	91,563
#Chinese	224,590	224,419	81,049	81,025
#symbol	0	0	10,487	10,502
#outlier	0	0	40	36

C. Performance Evaluation

In classification on extracted feature data and isolated character recognition, the recognition systems read extracted features or isolated character samples and output the classification results (top-rank class and top 10 classes) for each sample. The results are compared with the ground-truth to judge whether they are correct or not. The systems are ranked according to the correct rate, i.e., the percentage of correctly classified samples over the test samples:

$$CR = N_C / N_I , \tag{1}$$

Where N_C is the number of correctly recognized samples, and N_I is the total number of test samples. We report the top-rank correct rate as well as the acumulated correct rate of top 10 classes.

For continuous text recognition, we provide handwritten pages with text lines segmented. The recognition systems output the result (text transcription, a character string) for each text line. We compare the output character string of each text line with its ground-truth by error-correcting string matching to count how many characters are correctly recognized. A correct rate (CR) and an accurate rate (AR) [5,6] are calculated over all the text lines in the test dataset:

$$\begin{aligned} CR &= (N_t - D_e - S_e) / N_t , \\ AR &= (N_t - D_e - S_e - I_e) / N_t , \end{aligned} \tag{2}$$

where N_t is the total number of characters in the ground-truth texts, the numbers of substitution errors (S_e), deletion errors (D_e) and insertion errors (I_e) are obtained by error-correcting string matching by dynamic programming (DP). The accurate rate AR takes into account the inserted characters, and can be negative if the text lines are seriously over-segmented.

III. Participating Systems

In the following, we give brief descriptions of the submitted recognition systems provided by the developers.

A. Classification on Extracted Feature Data (TASK 1)

This task received registrations from five groups, and finally, three groups submitted three systems.

HIT: The School of Software at Harbin Institute of Technology (HIT) submitted a system, contributed by Tonghua Su, Songze Li, and Yu Ran, based on the method [7]. The original 512D feature vector is first reduced to 160D by linear discriminant analysis (LDA), and classified by a modified quadratic discriminant function (MQDF) [8] classifier with 50 principal eigenvectors per class. The MQDF classifier is further discriminatively trained using fast Perceptron learning with the margin regularization set as 0.05. The released feature data of HWDB1.0-1.1/OLHWDB1.0-1.1 [1] was used for training in the offline/online track.

SCUEC: The Information Processing Laboratory for Minority Language, College of Computer Science, South-Central University for Nationalities submitted a system, contributed by Xiaoxiao Li, Yi Yang and Zongxiao Zhu. The original 512D feature vector is first reduced to 128D by LDA, and classified by a MQDF classifier with 100 (for offline) or 128 (for online) principal eigenvectors per class. For improving the speed of recognition, a Euclid distance based classifier is applied to select 128 candidates for the MQDF classifier. The released feature data of HWDB1.0-1.1/OLHWDB1.0-1.1 was used for training in the offline/online track.

THU: The Department of Electronic Engineering of Tsinghua University (THU) submitted a system, contributed by Yanwei Wang, Changsong Liu and Xiaoqing Ding, based on MQDF [8] and Compound Mahalanobis Function (CMF) [9]. The original 512D feature is firstly transformed by Box-Cox transformation [10] and then reduced to 200D subspace by heteroscedastic linear discriminant analysis (HLDA) [11]. The MQDFs are trained with maximum likelihood rule on HWDB1.0-1.1 and OLHWDB1.0-1.1 respectively for offline and online tasks.

B. *Offline Isolate Character Recognition (Task 2)*

This task received registration from eight groups, and finally, five groups submitted seven systems.

HIT: The School of Software at Harbin Institute of Technology (HIT) submitted a system, contributed by Tonghua Su, Songze Li, Yu Ran, Tong Wei, Heng Zhang and Jianjun Fu, based on the method [7]. The system normalizes each gray-scale character image with the bi-moment normalization method [12]. Then gradient feature vector (512D) is extracted and reduced to 160D by LDA, and classified by a MQDF-like model trained by perceptron learning algorithm. The classifier structure is the same as our classifier in Task 1. The character samples (3755 classes) in CASIA-HWDB1.0 and CASIA-HWDB1.1 were used in training.

SCUEC: The Information Processing Laboratory for Minority Language, College of Computer Science, South-Central University for Nationalities submitted a system, contributed by Xiaoxiao Li, Yi Yang and Zongxiao Zhu. The character image is transformed to 64×64 by a nonlinear normalization method based on line density equalization. The 512D feature vector is extracted and reduced to 128D by LDA, and classified by a MQDF classifier with 57 principal eigenvectors per class, with acceleration by Euclid distance based candidate selection. 2,682,887 samples in CASIA-HWDB1.0-1.1 were used for training. About 60,000 mis-classified training samples were used to re-train a second MQDF classifier. The final decision is given by the winner of these two MQDF classifiers.

Fujitsu: The Fujitsu R&D Center Co., Ltd, Beijing, China, submitted a system, contributed by Chunpeng Wu, Wei Fan, Yuan He and Jun Sun. It is a high-performance GPU implementation based on the voting of four convolutional neural network (CNN). In each CNN, there are ten traditional convolutional layers (only two layers are fully connected), for the spatial pooling layer, not sub-sampling operator but max-pooling operator is used, and the output is a soft-max layer. The response of each neuron is truncated nonlinearity. The training set of CASIA-HWDB1.1 was used for training, and the test set of CASIA-HWDB1.1 was used for validating. In preprocessing, the character images are binarized and resized to 40x40 pixels and placed in the center of a 48x48 image with linear moment normalization. During training, each image is also normalized, randomly distorted using elastic deformations. All parameters of a CNN are randomly initialized using a 0-1 uniform distribution, and only the standard forward propagation is used. The training is stopped if no significant improvement is observed on the validation set.

IDSIAnn: The Dalle Molle Institute for Artificial Intelligence (IDSIA), Switzerland, submitted two systems based on a Multi-column Deep Neural Network (MCDNN) [13] with several DNN (columns) with different architectures (details in the Appendix), implemented by Dan Ciresan. Eight nets were trained separately, four on all data of CASIA-HWDB1.1, four only on its training set; seven with affine and elastic, one only with affine distortions. The eight nets were trained in parallel on four GTX 580 graphics cards for up to two weeks. Each character image is rescaled to 40x40 pixels and placed in the center of a 48x48 pixel image and finally feds it directly to the DNN input. One system combines the first four DNN; the other combines all eight.

THU: The Department of Electronic Engineering of Tsinghua University (THU) submitted a system, contributed by Yanwei Wang, Changsong Liu and Xiaoqing Ding, based on cascade classifiers. The gradient feature (588D) is extracted on gray-scale image and reduced to 200D by heteroscedastic linear discriminant analysis (HLDA)[11]. The main classifier is MQDF trained by reweighting the training samples [14]. The slave MQDF classifier is used to correct the classification errors of the main MQDF. The MQDFs are trained on HWDB1.0-1.1 and character samples extracted from the text data of HWDB2.0-2.2.

C. *Online Isolate Character Recognition (Task 3)*

This task received registrations from eight groups, and finally, six groups submitted nine systems.

HIT: The School of Software at Harbin Institute of Technology (HIT) submitted a systems, contributed by Tonghua Su, Songze Li, Yu Ran and Jianjun Fu based on the method [7]. The system normalizes each character sample with nonlinear normalization. Then gradient feature vector (512D) is extracted and reduced to 160D by LDA, and classified by a MQDF-like model trained by perceptron learning on the character samples (3755 classes) in CASIA-HWDB1.0 and CASIA-HWDB1.1.

Faybee: The Faybee Ltd., China, submitted a system based on the Euclidean distance classifier, implemented by Jinyun Hu. For feature extraction, strokes direction feature and virtual strokes direction feature are combined, and then the 1024D feature vector is reduced to 128D by LDA. The class prototypes were trained using LVQ2.1 on the training samples of CASIA-OLHWDB1.0 and CASIA-OLHWDB1.1.

TUAT: The Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology (TUAT) submitted a system based a two-level cascade of coarse classifiers [15], contributed by Bilan Zhu and Masaki Nakagawa. After coarse classification, two sets of character candidate classes, which are given by a structural recognizer and an un-structural recognizer, are combined to get the final result by a MCE-based combination method [16]. The structural recognizer uses a MRF model to match the feature points with the states of each character class among candidates and obtain a similarity for each character class [17]. The un-structural recognizer extracts directional features: histograms of normalized stroke direction [18] (512D 8-directional element features) after pseudo 2D bi-moment normalization (P2DBMN) [19], then the extracted features are reduced to 160D by FLDA and classified by an MQDF classifier trained with the GB2312-80 level-1 samples in CACIA-OLHWDB1.0-1.2.

UWarwick: University of Warwick, UK, submitted a RPCNN system, contributed by Ben Graham. The RPCNN consists of two parts. Firstly, characters are encoded in the form of a three dimensional array. Secondly, the array is fed into a large convolutional neural network for classification. The encoding is motivated by the "signature" from the theory of differential equations driven by rough paths. Truncated at the second level, the signature characterises a segment of a path by six numbers--two corresponding to the displacement of the path segment and four corresponding to the curvature of the path. Given the pen strokes that make up a character, draw the strokes onto a 72x72 square grid to produce a binary array. Extend the array into an array of size 72x72x7 by adding six additional layers corresponding to the elements of the truncated signatures. Once the array has been constructed, it is placed in the middle of a larger (96x96x7) array and fed into a convolutional neural network with architecture: 150C3-MP2-300C2-MP2-450C2-MP2-600C2-MP2-750C2-MP2-900N-3755N. The system was trained using the CASIA OLHWDB1.0-1.2 datasets, extended by elastic distortions, using CUDA code running on a GeForce GTX 680 graphics card.

VO: The Vision Objects Ltd., France, submitted three systems, contributed by Zsolt Wimmer based on their MyScript technology. The system normalizes the digital ink by applying a B-spline approximation on the input stroke, and extracts the features integrating dynamic and static information. Dynamic features include such as the position, direction and curvature of the ink signal trajectory. Static features are computed from a bitmap representation of the ink and are typically based on projections and histograms. Finally the feature vector is fed into a simple Multilayer Perceptron (MLP) classifier. The training data includes the samples in CASIA-OLHWDB1.0-1.1, the GB1 samples in SCUT-COUCH2009 [20], some private data, as well as distorted samples. The three systems differ in the trade-offs between speed and recognition: the VO-1 is the fastest but slightly less accurate than the VO-3 which provides the highest accuracy. The VO-2 is a compromise between them.

USTC: The National Engineering Laboratory for Speech and Language Information Processing (NEL-SLIP) , University of Science and Technology of China (USTC) submitted two systems, contributed by Jun Du. They use 8-directional feature [21] for prototype-based classifier [22] and the line segment feature for GMM-HMM classifier, respectively. For the former, the trajectory of a character is first mapped to 64*64 image. Then after smoothing and nonlinear normalization, the 392D 8-directional raw feature is extracted and reduced to 96D by LDA. For the line segment feature, the key step is to determine the segmentation points by calculating the angle changes using three consecutive points [23] and finally 4-dimensional feature vector is formed. Both the prototype-based classifier and GMM-HMM classifier are discriminatively trained [22][24]. At the recognition stage, for each character, 50 candidates are generated based on prototype-based classifier. Then based on those 50 candidates, they use GMM-HMM classifier with line segment features to rescore. Finally, they sort the candidates based on those two classifiers. The difference of USTC-1 and USTC-2 is that all the samples of 3755 character classes (GB2312-80 level-1) from CASIA-OLHWDB 1.0-2.2 are used to train both USTC-1 and USTC-2, and an additional dataset (3751940 samples) is only added for USTC-2.

D. *Offline Handwritten Text Recognition (Task 4)*

This task received registration from five groups, and finally, three groups submitted four systems.

HIT: The School of Software at Harbin Institute of Technology (HIT) submitted two systems, contributed by Tonghua Su, Peijun Ma, Hongliang Dai, Qin Xu, Fangyun Sun, Songze Li and Jianjun Fu based on the method [25]. In both systems, HMMs and connected component analysis are utilized to generate character segments. Then, a MQDF classifier and a trigram language model (LM) are used to score the segment networks. The optimal path is identified by beam search. To train the character classifier, isolated character samples of 7,356 classes are extracted from CASIA-HWDB1.0-1.2 and CASIAHWDB2.0-2.2 and PL-MQDF is selected to boost the accuracy [7]. The trigram LM is estimated using Kneser-Ney discounting on a large corpus of People’s Daily (with 6900 tokens in over 100M characters). Differences between these two systems lie in: HIT-2 used both the training and test data of CASIA-HWDB for training the character classifier, while HIT-1 used the training data only; HIT-1 compressed the character classifier; HIT-2 used a simple geometric context to facilitate the path searching.

SCUEC: The Information Processing Laboratory for Minority Language, College of Computer Science, South-Central University for Nationalities submitted a system, contributed by Xiaoxiao Li, Bin Yang, Zongxiao Zhu. The original text line image is coarsely segmented into segments based on projection strategy firstly, then it merges the over-segmented characters and segments adhesive characters, and last, merges the leading or trailing characters. The segmented character is transformed to 64×64 size by a nonlinear normalization method based on line density equalization, then 512D feature is extracted and reduced to 128D by LDA. An MQDF classifier with 57 principal eigenvectors is used for classification with acceleration by Euclidean distance based candidate selection. The classifier was trained with the character samples in CASIA-HWDB1.0-1.2, and CASIA-HWDB2.0-2.2 were used to train the character segmenting methods.

THU: The Department of Electronic Engineering of THU submitted a system contributed by Yanwei Wang, Changsong Liu and Xiaoqing Ding. The system employs an over-segmentation-and-merging method [26]. Each segmentation path is scored by integrating the character recognition model, linguistic context and geometric information. The optimal segmentation path is found by dynamic programming search. The character recognition model is an MQDF classifier (3,957 classes, including 3,879 Chinese characters and 78 characters) trained on samples extracted from HWDB1.0-1.1 and HWDB2.0-2.2. The language model is a character bi-gram trained on a corpus of People’s Daily.

E. *Online Handwritten Text Recognition (Task 5)*

This task received registrations from six groups, and finally, three groups submitted four systems

TUAT: The Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology (TUAT) submitted a system contributed by Bilan Zhu and Masaki Nakagawa. The system is based on an over-segmentation-and-merging method. It applies the beam search strategy to search for the candidate lattice. During the search, the paths are evaluated in accordance with the path evaluation criterion proposed by Zhu et al. [27], which combines the scores of character recognition, linguistic context, and geometric features with the weighting parameters estimated by GA. The character recognition model is a combined recognizer of a structural recognizer (MRF recognizer) and an un-structural recognizer (P2DBMN-MQDF recognizer) (7,356 classes, including 7,184 Chinese characters and 172 symbols) trained on samples extracted from CASIA-HWDB1.0-1.2. The language model is a character tri-gram trained on a corpus of People's Daily.

VO: The three systems submitted by Vision Objects use three "experts" (segmentation, recognition, interpretation) collaborating through dynamic programming to process the digital ink and generate candidates at the character, word, and sentence level. The segmentation expert constructs a segmentation graph where each node corresponds to a character hypothesis and adjacency constraints between characters are handled by the node connections. The recognition expert (an MLP classifier handling 7,425 character classes) associates a list of character candidates with recognition scores to each node of the graph. The interpretation expert generates linguistic meaning for the different paths in the segmentation graph, using a word tri-gram language model based on a 130K word lexicon to evaluate the linguistic likelihood of the interpretation of a given path of the graph. Moreover, a global discriminant training scheme on the text level with automatic learning of all classifier parameters and meta-parameters of the recognizer is employed. The three systems differ in the trade-offs between speed and recognition, the VO-1 is the fastest but slightly less accurate than the VO-3 which provides the highest accuracy. The VO-2 is a compromise between them.

USTC: The National Engineering Laboratory for Speech and Language Information Processing (NEL-SLIP), University of Science and Technology of China (USTC) submitted one system, contributed by Jun Du. For training, all the character samples of 7363 character classes from CASIA-OLHWDB and an additional dataset were used. The classifier configuration is the same as that in Task 3. The overall system architecture is similar to [28]. For the input handwritten text, we first perform stroke segmentation and character over-segmentation to form the lattice, and then tri-gram language model combined with character classifiers are used for decoding to improve the performance of text recognition.

IV. Recognition Results

The submitted systems were evaluated on the competition test datasets, and each system loads the test samples from hard disk and output the recognition results in a result file of specified format [4]. All CPU based systems were executed on a personal computer with Intel Core i5-2400-3.1GHz CPU, 8G RAM, integrated graph card and MS Windows7 OS, while GPU based systems were executed on a Linux server with two Intel X5650-2.67GHz CPU, 64G RAM and two Tesla-C2050 card. For tasks 1-3, we also report the average processing time per character. For handwritten text recognition, we report the average time per text line. The number of characters per text line is about 26.68 (slightly different between online and offline data). As a measure of system complexity, we also show the size (number of bytes) of dictionary file, which stores the classifier parameters and context model parameters.

The evaluation results of five tasks are shown in Tables 2-6, for classification on extracted features, offline character recognition, online character recognition, offline text recognition and online text recognition, respectively. In each Table, the last column shows the dictionary size. Some systems are given the size of the executive file which embeds the dictionary.

For comparing with the state of the art, we also include some published results on the same test sets. Further, we provide human recognition results on isolated characters. This result was obtained by showing 10 confusing classes for each sample and letting the human operator to click a class. The online/offline character sample set was partitioned into six parts, each part recognized by at least three persons. For each part, we took the result of highest correct rate among the persons, because the high accuracy indicates that the person worked faithfully.

Table 2. Result of classification on extracted features (%).

	CR(1) online	CR(1) offline	CR(1) ave	CR(10) ave	Ave time	Dic size
HIT	**95.18**	**92.60**	**93.89**	**99.20**	2ms	120M
THU	95.03	92.21	93.62	99.19	6ms	242M
SCUEC	92.44	86.52	89.48	98.26	9ms	50M
Ref [1]	95.31	92.72				

Table 3. Results of offline character recognition (%).

System	CR (1)	CR (10)	Ave time	Dic size
Fujitsu	**94.77**	**99.59**	55ms*	2460M
IDSIAnn-2	94.42	99.54	315ms	349M
IDSIAnn-1	94.24	99.52	197ms	47.1M
HIT	92.62	98.99	4.6ms	120M
THU	92.56	99.13	8.2ms	198M
SCUEC	77.71	95.29	19.3ms	484M
Ref [1]	92.72			
Human	96.13			

*Time of execution on Linux GPU sever.

Table 4. Results of online character recognition (%).

System	CR (1)	CR (10)	Ave time	Dic size
UWarwick	**97.39**	**99.88**	355ms	37.8M
VO-3	96.87	99.67	15.3ms	87.6M*
VO-2	96.72	99.61	4.1ms	36M*
VO-1	96.33	99.61	1.6ms	10M*
HIT	95.18	99.39	2.3ms	120M
USTC-2	94.59	99.14	3.8ms	5.25M
USTC-1	94.25	99.06	2.0ms	3.19M
TUAT	93.85	99.24	5.3ms	96.2M
Faybee	92.97	98.87	0.5ms	4.48M
Ref [1]	95.31			
Human	95.19			

Table 5. Results of offline text recognition (%).

	CR	AR	Ave time	Dic size
HIT-2	**88.76**	**86.73**	1.2s	309M
HIT-1	86.15	83.58	0.64s	111M
THU	82.92	79.81	0.85s	102M
SCUEC	42.05	35.14	0.15s	442M
Ref[6]	90.22	89.28		

Table 6. Results of online text recognition (%).

	CR	AR	Ave time	Dic size
VO-3	**95.03**	**94.49**	1.72s	56M*
VO-2	94.94	94.37	1.23s	37.9M*
VO-1	93.11	92.57	0.72s	20.8M*
TUAT	88.49	87.66	1.42s	246M
USTC	82.20	81.57	0.25s	29.3M
Ref [29]	94.62	94.06		

*Size of executive file embedding dictionary.

In classification on extracted feature data (Task 1), the system of HIT yields the highest accuracy on both online and offline features, but the difference between the systems of HIT and THU is small. These results are slightly lower than the published results in [1].

In offline character recognition (Task 2), the system of Fujitsu yields the highest accuracy and accumulated accuracy. The IDSIAnn-2 and IDSIAnn-1 systems also yield competitive accuracies. The superior performance of CNN-based method is attributed to its complex neural network structure and discriminative training with large number of original samples and distorted samples. Both the HIT and THU systems use direction histogram features and discriminatively trained quadratic discriminant classifiers. The performance of Fujitsu and IDSIAnn are superior to the published results in [1], but are by far inferior to human recognition performance.

In online character recognition (Task 3), The UWarwick system yields the highest accuracy. Its superior performance is due to the fact that they effectively denote the characteristics of stroke trajectory with “signature” from the theory of differential equations, and use the complex structure convolutional neural network classifier trained with large number of multi-source samples and distorted samples. Another neural network classifier, VO, also performs competitively well. Among the systems using statistical classifiers, the HIT system performs best. The VO-1 systems show good tradeoff between performance and complexity. On the other hand, the high speed of Faybee systems is attributed to its simple classifier structure. The systems of UWarwick and VO outperform the published results in [1] and human recognition. Interestingly, the human recognition accuracy of online characters is lower than that of offline recognition. This can be explained that the display of online characters as still images is not as pleasing as that of offline samples.

The best results of both offline and online character recognition are much better than the best ones of Competition 2011 [3], where 92.18% of offline recognition and 95.77% of online recognition were achieved on the same test datasets as for 2013. Also, the best performing groups in 2011, IDSIAnn and VO, exhibited evident progress in 2013.

In offline text recognition (Task 4), the system HIT reports the highest CR and AR. Except SCUEC, other three participating systems take the character over-segmentation strategy and integrate the character recognition model, linguistic and geometric contexts. However, their performances are significantly lower than that of the published state-of-art method [6].

In online text recognition (Task 5), the systems of VO yield superior performance. They also adopt the character over-segmentation strategy, but implement the character classifier (neural network with discriminative training on large number of samples), context models, and combine the models with better implementation. They outperform the VO systems in Competition 2011 and a recent published method in [29].

Overall, the results of both isolated handwritten Chinese character recognition and handwritten text recognition have shown evident progress compared to the previous competitions. The research of handwritten Chinese text recognition has not been widely undertaken, but the competition results are still encouraging.

V. CONCLUSION

The Chinese Handwriting Recognition Competition 2013 attracted ten groups to participate and received 27 systems for five tasks: classification on extracted feature data (Task1), offline isolated character recognition (Task 2), online isolated character recognition (Task 3), offline handwritten text recognition (Task 4), and online handwritten text recognition (Task 5). The best results were yielded by the systems of HIT (Task 1), Fujitsu (Task 2), UWarwick (Task 3), HIT (Task 4) and VO (Task5), respectively. The submitted systems are variable in complexity in respect of dictionary size and processing time. Neural network based method have shown superiority in both isolated character recognition and handwritten text recognition. However, the high complexity in both training and testing of deep neural networks is hoped to be overcome in the future. We look forward to more participants in the future competitions and more researchers joining the research of Chinese handwriting recognition.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (NSFC) under grants 60933010, 61175021 and 61273269. We thank the volunteers contributed to human recognition and the competition participants for their active participation and feedbacks.

REFERENCES

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

[2] C.-L. Liu, F. Yin, D.-H. Wang, Q.-F. Wang, Chinese handwriting recognition contest 2010, Proc. 2010 Chinese Conf. Pattern Recognition, 2010, pp. 1–5.

[3] C.-L. Liu, F. Yin, Q.-F. Wang, D.-H. Wang, ICDAR 2011 Chinese Handwriting Recognition Competition, Proc. Int'l 11th Conf. Document Analysis and Recognition, 2011, pp. 1464-1469.

[4] http://www.nlpr.ia.ac.cn/events/CHRcompetition2013/competition/Home.html

[5] A. Vinciarelli, S. Bengio, H. Bunke, Offline recognition of unconstrained handwritten texts using HMMs and statistical language models, IEEE Trans. Pattern Anal. Mach. Intell., 26(6): 709-720, 2004.

[6] Q.-F. Wang, F. Yin, C.-L. Liu, Handwritten Chinese Text Recognition by Integrating Multiple Contexts. IEEE Trans. Pattern Anal. Mach. Intell. 34(8): 1469-1481, 2012.

[7] T.-H. Su, C.-L. Liu, X.-Y. Zhang, Perceptron learning of modified quadratic discriminant function, Proc.11th Int'l Conf. on Document Analysis and Recognition, 2011, pp. 1007-1011.

[8] F. Kimura, K. Takashina, S. Tsuruoka, et al. Modified quadratic discriminant functions and its application to Chinese character recognition. IEEE Trans. Pattern Anal. Mach. Intell, 9(1): 149~153, 1987

[9] H.-L. Liu, X.-Q. Ding. Handwritten character recognition using gradient feature and quadratic classifier with multiple discrimination schemes. Proc. 8th Int'l Conf. on Document Analysis and Recognition, 2005, pp. 19~23.

[10] R.V.D. Heiden, F.C.A. Gren, The Box–Cox metric for nearest neighbor classification improvement, Pattern Recognition 30 (2): 273–279, 1997.

[11] H.-L. Liu, X.-Q. Ding, Improve Handwritten Character Recognition Performance by Heteroscedastic Linear Discriminant Analysis Conference, Proc. Int'l Conf. on Pattern Recognition, pp. 880-883, 2006.

[12] C.-L. Liu, H. Sako, H. Fujisawa, Handwritten Chinese character recognition: alternatives to nonlinear normalization, Proc. 7th Int'l Conf. on Document Analysis and Recognition, 2003, pp. 524-528.

[13] D. Ciresan, U. Meier, J. Schmidhuber, Multi-column Deep Neural Networks for Image Classification, CVPR, 2012, pp. 3642-3649.

[14] Y.-W. Wang, X.-Q. Ding, C.-S. Liu, MQDF discriminative learning based offline handwritten Chinese character recognition. Proc. Int'l Conf. on Document Analysis and Recognition, 2011, pp. 1100-1104.

[15] B. Zhu and M. Nakagawa, A Coarse Classifier Construction Method from a Large Number of Basic Recognizers for On-line Recognition of Handwritten Japanese Characters, Proc. 11th ICDAR, 2011, pp.1090-1094.

[16] B. Zhu, J. Gao and M. Nakagawa, Objective Function Design for MCE-Based Combination of On-line and Off-line Character Recognizers for On-line Handwritten Japanese Text Recognition, Proc. 11th ICDAR, 2011, pp.594-599.

[17] B. Zhu and M. Nakagawa, On-line Handwritten Japanese Characters Recognition Using a MRF Model with Parameter Optimization by CRF, Proc. 11 th ICDAR, 2011, pp.603-609.

[18] C.-L. Liu and X.-D. Zhou, Online Japanese Character Recognition Using Trajectory-based Normalization and Direction Feature Extraction, Proc. 10th IWFHR, 2006, pp.217-222.

[19] C.-L. Liu and K. Marukawa, Pseudo Two-dimensional Shape Normalization Methods for Handwritten Chinese Character Recognition, Pattern Recognition, 38(12), pp. 2242-2255, 2005.

[20] L.-W. Jin, Y. Gao, et.al, SCUT-COUCH2009--A comprehensive online unconstrained Chinese handwriting database and benchmark evaluation, Int. J. Document Analysis and Recognition, 14(1): 53-56, 2011.

[21] Z.-L Bai, Q. Huo, Study on the use of 8-directional features for online handwritten Chinese character recognition, Proc. 8th ICDAR, pp.262-266, 2005.

[22] Q. Huo, Y. Ge, Z.-D. Feng, High performance Chinese OCR based on Gabor features, discriminative feature extraction and model training, Proc. ICASSP, pp.III-1517-1520.

[23] K. Parthasarathy, J. L. C. Seybold. Method and microprocessor for preprocessing handwriting having characters composed of a preponderance of straight line segments. In US Patent 5,740,273.

[24] Y. Ge, Q. Huo, Z.-D. Feng, ffline recognition of handwritten Chinese characters using Gabor features, CDHMM modeling and MCE training, Proc. ICASSP, pp.I–1053–1056, 2002.

[25] T.-L. Su, Chinese Handwriting Recognition: An Algorithmic Perspective, Springer, 2013.

[26] Q. Fu, X.-Q. Ding,T. Liu, J. Yan, Z. Ren. A novel segmentation and recognitionalgorithm for Chinese handwritten address character strings. Proc. ICPR, 2006: 974-977.

[27] B. Zhu, X.-D. Zhou, C.-L. Liu, M. Nakagawa, A Robust Model for On-line Handwritten Japanese Text recognition, IJDAR, 13(2), pp.121-131, 2010.

[28] X.-D. Zhou, J.-L. Yu, C.-L. Liu, T. Nagasaki, K. Marukawa, online handwritten Japanese character string recognition incorporating geometric context, Proc. 9th ICDAR, pp.48-52, 2007.

[29] X.-D. Zhou, D.-H. Wang, F. Tain, C.-L. Liu, M. Nakagawa, Handwritten Chinese/Japanese text recognition using semi-Markov conditional random field, IEEE Trans. Pattern Anal. Mach. Intell., in press, 2013.

APPENDIX

The DNN architecture is compactly described as follows. For convolutional layers we list the number of maps followed by ‘C’ and filter size. Max-pooling layers (MP) need only specify filter size. Fully connected layers are described by their numbers of neurons suffixed by ‘N’.

Table 7. DNN architecture and training data distortions; validation error on the undistorted training set.

DNN	Architecture	Maximum translation [%]	Maximum rotation [%]	Maximum scaling [%]	Elastic	Validation error [%]
0	150C3-MP2-250C2-MP2-350C2-MP2-450C2-MP2-1000N	10	10	10	6; 36	2.42
1	150C3-MP2-250C2-MP2-350C2-MP2-450C2-MP2-1000N	10	10	10	0; 0	1.89
2	300C3-MP2-300C2-MP2-300C2-MP2-300C2-MP2-1000N	10	10	10	6; 36	3.18
3	100C3-MP2-200C2-MP2-300C2-MP2-400C2-MP2-500N	10	10	10	6; 36	3.61
4	100C3-MP2-200C2-MP2-300C2-MP2-400C2-MP2-1000N	10	10	10	6; 36	2.39
5	100C3-MP2-200C2-MP2-300C2-MP2-400C2-MP2-1000N	10	10	10	6; 36	2.08
6	100C3-MP2-200C2-MP2-300C2-MP2-400C2-MP2-1000N	15	15	15	6; 36	3.65
7	100C3-MP2-200C2-MP2-300C2-MP2-400C2-MP2-500N	10	10	10	6; 36	3.50