ICDAR 2011 – Arabic Handwriting Recognition Competition

Volker Märgner
Technische Universität Braunschweig
Institute for Communications Technology (IfN)
Braunschweig, Germany
v.maergner@tu-bs.de

Haikal El Abed
Technische Universität Braunschweig
Institute for Communications Technology (IfN)
Braunschweig, Germany
elabed@tu-bs.de

***Abstract*—This paper describes the Arabic handwriting recognition competition held at International Conference on Document Analysis and Recognition (ICDAR) 2011. This fifth competition again used the IfN/ENIT-database with Arabic handwritten Tunisian town names. Today, more than 110 research groups from universities, research centers, and industry are working with this database worldwide. This year, 4 groups with 4 systems were participating in the competition. The systems were tested on known data (sets *d* and *e*) and on two data sets which are unknown to the participants (sets *f* and *s*). The systems were compared based on the most important characteristic: the recognition rate. A short description of the participating groups, their systems, and the results achieved are finally presented.**

***Keywords*-Off-line Text Recognition; Evaluation; Benchmarking, IfN/ENIT Database**

I. Introduction

Research on Arabic handwritten word and text recognition is still of great interest. Much works were done in recent years in this field. Especially since 2005, when the first competition took place at ICDAR conference [1], an improvement of published systems could be observed. This paper presents the results of the fifth competition of Arabic handwritten word recognition systems. The results of this third competition were presented during the ICDAR 2011 conference in Beijing, China. The competition was again carried out by the group at the Institute for Communications Technology (IfN) of Technische Universitaet Braunschweig, Braunschweig, Germany. In comparison to the competition in 2010, this year 4 groups with 4 systems participated in the competition: two groups were also participants in the last competition, while two other groups were now participating for the first time. The competition is again held as a closed competition, runtime versions of recognition systems were sent to the organizing group and tested in their environment. This year the test was performed on the same datasets (sets *f* and *s*) as those in the ICDAR 2007, ICDAR 2009, and ICFHR 2010 competitions [2], [3], [4].

The first competition on Arabic handwriting recognition was based on the IfN/ENIT-database, and the results were presented at the International Conference on Document Analysis and Recognition (ICDAR) 2005 [1]. Five groups submitted systems to this competition.

The second competition on Arabic handwriting recognition was organized in the same manner than the first with the only difference that the test set of the first competition (dataset e) was available for training too. The results again were presented at the ICDAR 2007 [2]. This competition compared 14 systems submitted from 9 groups (some groups delivered more than one system). A comparison with the 2005 tests show an improvement of more than 5% for the best systems. But of course there are still recognition errors of about 20% of the best systems.

The third competition on Arabic handwriting recognition was organized in the same manner than the first two competitions. The results were presented at the ICDAR 2009 [3], [5]. This competition compared 17 systems submitted from 7 groups. A comparison with the 2007 tests again show an improvement of more than 5% for the best systems.

The fourth competition on Arabic handwriting recognition was organized in the same manner than the first three competitions. The results were presented at the International Conference on Frontiers in Handwriting Recognition (ICFHR) 2010 [4]. This competition compared 6 systems submitted from 4 groups. All participating systems were based on Hidden Markov models (HMM) classifier and they have shown very high performances.

This paper is organized as follows: In Section II the database and the test sets are presented shortly. Section III presents the participating groups and gives a short description of the submitted systems. Section IV describes the tests and the results achieved by the different systems. Finally the paper ends with some concluding remarks.

II. IfN/ENIT-Database

A. The IfN/ENIT-Database

The IfN/ENIT-database was developed to advance the research and development of Arabic handwritten word recognition systems. Since the presentation of this database at the CIFED 2002 conference [6], more than 110 groups in about 35 countries are working today (i.e., at the beginning of 2011) with the IfN/ENIT-database, which is freely available (www.ifnenit.com) for non commercial research.

The database in version 2.0 patch level 1e (v2.0p1e) consists of 32492 Arabic words handwritten by more than 1000

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.287

Table I
NUMBER OF NAMES, CHARACTERS, AND PAWS APPEARING IN THE IFN/ENIT-DATABASE V2.0P1E (SETS *a* TO *e*) AND THE TEST SETS *f* AND *s*

	set	names	characters	PAWs
training sets	*a*	6537	51984	28298
	b	6710	53862	29220
	c	6477	52155	28391
	d	6735	54166	29511
	e	6033	45169	22640
test sets	*f*	8671	64781	32918
	s	1573	11922	6109

Table III
FREQUENCY OF NUMBER OF PAWS

PAWs	frequency in %		PAWs	frequency in %	
	set *f*	set *s*		set *f*	set *s*
1	4.69	4.32	6	9.11	8.96
2	16.58	15.13	7	3.16	3.50
3	25.82	25.30	8	2.24	2.67
4	23.11	23.67	>8	0.21	0.38
5	15.11	15.77			

writers. The words written are 937 Tunisian town/village names [1] (Table II). Each writer filled one to five forms with pre-selected town/village names and the corresponding post code. Ground truth was added to the image data automatically and verified manually.

B. The Test Datasets

The test datasets which are unknown to all participants were collected for the tests of the ICDAR 2007 competition [2]. The words are from the same lexicon as those of IfN/ENIT-database and written by writers, who did not contribute to the data sets before. For the test purpose, these data are separated into set *f* and set *s* (Table I).

Set *f* was collected in Tunisia, while set *s* was collected in the United Arab Emirates (UAE) at the University of Sharjah. Table III shows the frequency of PAWs (Parts of Arabic Words) within each name of the new test datasets *f* and *s*.

III. PARTICIPATING SYSTEMS

Eight groups have registered for the ICDAR 2011 Arabic handwriting recognition competition, finally 4 groups have submitted their system for evaluation. The following section gives a brief description of the systems submitted to the competition. Each system description was provided by the system's authors and edited (summarized) by the competition organizers. The descriptions vary in length due to the level of detail in the provided source information.

A. JU-OCR

The JU-OCR system is submitted by Gheith Abandah and Fuad Jamour, from the University of Jordan, Jordan.

JU-OCR is a recognition system for handwritten Arabic text. This system is intended to recognize unlimited vocabulary and is based on explicit grapheme segmentation. It segments a cursive word into the set of graphemes that forms it, then it recognizes each of the graphemes, and it maps the recognized grapheme into the letters that form the word.

JU-OCR uses the segmentation algorithm described in [7]. This algorithm is a rule-based algorithm that utilizes features extracted from the skeleton of Arabic sub-words to segment them into a set of graphemes. Most graphemes are forms of the Arabic letters. However, some graphemes are parts of letters (over segmentation), and some graphemes are vertical ligatures (under segmentation). Each grapheme has a main body, and some graphemes have secondary bodies. A body is a connected component that forms a blob recognizable by humans. Statistical and morphological features [8] are extracted from each body of the graphemes and passed to a Random Forest (RF) classifier [9] to recognize the body. We use the OpenCV implementation of the RF classifier. After each body is recognized, the bodies are combined to form graphemes. This combination is carried out through rules for what bodies combine to form graphemes. Finally, another set of rules are used to map graphemes into letters. These rules map some graphemes to one letter each, multiple graphemes to one letter, or one grapheme to multiple letters.

As this competition uses a limited vocabulary of 937 words, the authors have developed a string matching algorithm that finds the closest word match of a recognized grapheme sequence without mapping the graphemes to letters. The matching algorithm finds a weighted edit distance between the predicted grapheme sequence and the most probable grapheme sequence of each word in the dictionary. The word that has the minimum distance is the recognition result.

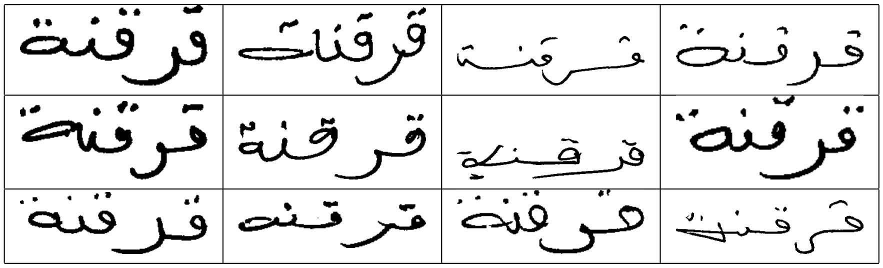
B. CENPARMI

The CENPARMI-OCR system is submitted form Muna Khayyat, Louisa Lam, and Ching Y. Suen, from the Computer Science and Software Engineering Department, Concordia University, Center for Pattern Recognition and Machine Intelligence (CENPARMI), Montreal, Quebec, Canada.

The CENPARMI-OCR uses three sets of features appropriate for Arabic handwriting, with each set of feature passed to one classifier. The confidence levels and classification results of the classifiers were used for the final classification result. The three sets of features are: Gradient Features [10], Gabor Features [11] and Fourier Features [12].

The authors used three different SVMs [13] each of which was trained on a different feature set. The three classifiers are divided into two groups: Primary and Secondary. The

Table II
EXAMPLES FROM THE IFN/ENIT-DATABASE: THE TUNISIAN ISLAND NAME KERKENNAH (قرقنة) WRITTEN BY 12 WRITERS.



former group consists of the classifier on which the gradient features were trained while the latter group consists of the two classifiers on which Gabor and Fourier features were trained.

The testing samples are tested on the three classifiers. The system verifies the result of the primary classifier. If the confidence value (posterior probability) of the primary classifier for a sample is below a pre-determined threshold the system verifies the classification results of the two secondary classifiers. If they agree on the class, then the sample would be assigned to this (common) class together with the higher confidence value from these two secondary classifiers.

C. RWTH-OCR

The RWTH-OCR Arabic Handwriting Recognition System for ICDAR 2011 competition is submitted by Patrick Doetsch, Philippe Dreuw, Mahdi Hamdani, Christian Plahl, and Hermann Ney from the RWTH Aachen University, Human Language Technology and Pattern Recognition, Aachen, Germany.

Without any preprocessing of the input images, the authors extract simple appearance-based image slice features X_t at every time step $t = 1, \cdots, T$ which are augmented by their spatial derivatives in horizontal direction $\Delta = X_t - X_{t-1}$.

Due to a character and position dependent length modeling of the 28 base Arabic characters [14], the authors finally model the Arabic words by 121 different character labels. The system described in [15] is used to generate an alignment of the features to the 121 labels.

The raw slice features X_t together with their corresponding state alignments are then processed by a hierarchical MLP framework originally described in [16].

A TRAP-DCT MLP network is based on the MLP framework originally described in [16]. The system uses a TRAP-DCT [17] preprocessing of the raw pixel input features. In order to incorporate temporal and spatial context into the features, the authors concatenate consecutive features in a sliding window, where the MLP outputs are later reduced by an LDA transformation.

The hierarchical system uses at the first level a spatio-temporal TRAP-DCT window to augment the 30-dimensional raw pixel input feature vectors to a 240-dimensional vector. The first level hierarchical network uses a single hidden layer with 750 nodes, and 121 output nodes, which are reduced by a log-LDA transformation to 96 components. The second network concatenates these features in addition to the raw features, and uses a window size of 5 consecutive log-LDA network features, and a window size of 9 consecutive raw input features to account for different spatio-temporal information. The 750-dimensional features (i.e. $96 \times 5 + 30 \times 9$ features) are forwarded to a single hidden layer with 1500 nodes, and finally reduced again by a log-LDA transformation to 36 components.

The hidden Markov model (HMM) based handwriting recognition system is Viterbi trained using the maximum-likelihood training criterion. For Gaussian mixture training in our base system, the authors perform supervised model training by iteratively re-estimating the emission model parameters and splitting of the mixtures. For the discriminatively trained model, the authors use 7 splits with up to 128 densities per mixture and 3 mixtures per character label, resulting in 70745 densities.

D. REGIM

The Research Group on Intelligent Machines (REGIM) at the Ecole Nationale d'Ingénieurs de Sfax (ENIS), University of Sfax, Tunisia, participated with a system submitted by Mahdi Hamdani, Tarek M. Hamdani, and Adel M. Alimi.

This system is based on HMMs [18] and it is an improved version of the work presented in [19]. The improvement consist on the optimization of the HMMs architectures (number of states) using Particle Swarm Optimization (PSO). The maximum likelihood is used as fitness function by the PSO. The used features are based on the transformation of the pixel values extracted from normalized images using Karhunen-Loéve-Transform. More details about the used features are presented in [20]. The results of single PSO-HMMs are improved using the combination methods described in [21].

IV. Tests and Results

We evaluated the performance of the 4 different Arabic handwriting recognition systems in two steps. In the first step, we used a subset and then the whole datasets *d* and *e* of the IfN/ENIT-database for a function check of the systems. In a second step, we used the test datasets *f* and *s*, unknown to all participants.

The most important results of our tests are shown in Table IV. For each test, the best result is marked in bold font. More details will be presented at ICDAR 2011 conference in Beijing.

A. Tests with known Data (sets d and e)

The comparison of the systems based on the results of sets *d* and *e*, which are part of the training set, shows 3 systems with a recognition rate better than 90% on set *d* and 85% on set *e*. It is interesting to see that the relative position of all systems is the same for sets *d* and *e*.

B. Main Tests (sets f, f_a, f_f, and f_g)

The most important test to compare the performance of different systems is of course the test using the new set *f*. The features of this set should be similar to sets *a* to *e*, as it was collected in the same country. As the distributions of words in all sets of the database are different, three subsets of set *f* are generated to make the word distribution of training and testing sets more similar: Set f_a (8290 names) limits the number of a name in the test set by the number of the name in the training set, set f_f (4319 names) approaches the distribution of the test set by that of the training set, and in set f_g (3393 names) the appearance of a name in the test set is limited to three.

Table IV shows some interesting results: (1) one system recognize more than 90% correctly, (2) the difference between set *f* and the f_x sets is about 1 to 3% (i.e., there is no strong dependency of the words statistic), (3) the loss of the systems compared to set *e* differs very much, however, one system shows even a better result on set *f* than on set *e*. The best system has a recognition rate of 11% higher than the second-best system, and the absolute value is again comparable to that in the competitions ICDAR 2007, ICDAR 2009, and ICFHR 2010.

C. Robustness Test (set s)

The test with data from the UAE is very interesting. Although all training data comes from Tunisia, the recognition rate on this set of one system is better than 80% and of 2 systems are better than 50%. This is a loss of about 10% of the best system compared to the recognition rate on set *f*, but it shows that the generalization ability of all systems is not too bad.

V. Conclusion

The competition results show that Arabic handwriting recognition systems in this fifth competition made no further progress. Only one of the participating systems shows a high accuracy. Details and specific features and classification approaches of the systems cannot be presented in this short paper. The system 3 (RWTH-OCR) is the winner of the ICDAR 2011 Arabic handwriting recognition competition.

VI. Acknowledgments

The organizers of the fifth Arabic handwriting recognition competition would like to thank all participants. Their contributions makes this series of competitions successful.

References

[1] V. Märgner, M. Pechwitz, and H. El Abed, "ICDAR 2005 – Arabic Handwriting Recognition Competition," in *Proceedings of the 8th International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1, 2005, pp. 70–74.

[2] V. Märgner and H. El Abed, "ICDAR 2007 – Arabic Handwriting Recognition Competition," in *Proceedings of the* 9^{th} *International Conference on Document Analysis and Recognition (ICDAR)*, vol. 2, 2007, pp. 1274–1278.

[3] ——, "ICDAR 2009 – Arabic Handwriting Recognition Competition," in *Proceedings of the 10th International Conference on Document Analysis and Recognition (ICDAR)*, 2009, pp. 1383–1387.

[4] ——, "ICFHR 2010 – Arabic Handwriting Recognition Competition," in *Proceedings of the 12th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2010, pp. 709–714.

[5] H. El Abed and V. Märgner, "ICDAR 2009 – Arabic Handwriting Recognition Competition," *International Journal on Document Analysis and Recognition*, vol. 14, no. 1, pp. 3–13, 2011.

[6] M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, and H. Amiri, "IfN/ENIT – Database of Handwritten Arabic Words," in *Proceedings of the Colloque International Francophone sur l'Ecrit et le Document (CIFED)*, 2002, pp. 127–136.

[7] G. Abandah and F. Jamour, "Recognizing Handwritten Arabic Script Through Efficient Skeleton-Based Grapheme Segmentation Algorithm," in *Proceedings of the 10th International Conference on Intelligent Systems Design and Applications (ISDA)*, 2010, pp. 977–982.

[8] G. Abandah and T. Malas, "Feature Selection for Recognizing Handwritten Arabic Letters," *Dirasat Engineering Sciences Journal*, vol. 37, no. 2, pp. 242–256, October 2010.

[9] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[10] M. Shi, Y. Fujisawa, T. Wakabayashi, and F. Kimura, "Handwritten Numeral Recognition Using Gradient and Curvature of Gray Scale Image," *Pattern Recognition*, vol. 35, no. 10, pp. 2051–2059, 2002.

Table IV
RECOGNITION RESULTS IN % OF CORRECT RECOGNIZED IMAGES ON REFERENCE DATASETS d AND e, NEW DATASETS f AND s, SUBSETS f_a, f_f, AND f_g. (G-ID: GROUP ID, S-ID: SYSTEM ID).

G-ID	S-ID	Approach	set d	set e	set f_a	set f_f	set f_g	set f			set s		
			top 1	top 1	top 1	top 1	top 1	top 1	top 5	top 10	top 1	top 5	top 10
JU-OCR	1	RF	75.49	63.75	64.96	66.77	67.64	63.86	80.18	84.65	49.75	66.86	72.46
CENPARMI-OCR	2	SVM	**99.90**	**99.91**	40.00	40.00	40.00	40.00	69.33	74.00	35.52	54.56	63.84
RWTH-OCR	3	HMM	99.67	98.61	92.35	92.20	92.81	**92.20**	95.73	96.15	**84.55**	91.99	93.52
REGIM	4	HMM	94.12	86.62	80.60	81.36	81.52	79.03	89.35	91.34	68.44	81.99	84.98
Results of the 3 best systems at ICFHR 2010 [4]													
UPV PRHLT	2	HMM	99.38	98.03	93.46	94.30	94.02	**92.20**	95.72	96.29	**84.62**	91.42	93.32
CUBS-AMA	4	HMM	89.97	80.80	81.75	83.35	83.55	80.32	88.26	88.96	67.90	78.58	79.87
RWTH-OCR	6	HMM	99.66	98.84	92.35	93.35	93.55	90.94	95.31	96.00	80.29	89.83	91.80
Results of the 3 best systems at ICDAR 2009 [5]													
MDLSTM	11	NN	99.94	99.44	94.68	95.65	96.02	**93.37**	96.46	96.77	**81.06**	88.94	90.72
Ai2A	8	HMM	97.02	91.68	90.66	91.92	92.31	89.42	95.33	95.94	76.66	88.01	90.28
RWTH-OCR	14	HMM	99.79	98.29	87.17	88.63	88.68	85.69	93.36	94.72	72.54	83.47	86.78
Results of the 3 best systems at ICDAR 2007 [2]													
Siemens	08	HMM	94.58	87.77	88.41	89.26	89.72	**87.22**	94.05	95.42	**73.94**	85.44	88.18
MIE	06	LDC	93.63	86.67	84.38	85.21	85.56	83.34	91.67	93.48	68.40	80.93	83.73
UOB-ENST	11	HMM	92.38	83.92	83.39	84.93	85.18	81.93	91.20	92.76	69.93	84.11	87.03

[11] J. Chen, H. Cao, R. Prasad, A. Bhardwaj, and P. Natarajan, "Gabor Features for Offline Arabic Handwriting Recognition," in *Proceedings of the 9th International Workshop on Document Analysis Systems (DAS)*, 2010, pp. 53–58.

[12] V. Lavrenko, T. M. Rath, and R. Manmatha, "Holistic Word Recognition for Handwritten Historical Documents," in *Proceedings of the First International Workshop on Document Image Analysis for Libraries (DIAL)*, January 2004, pp. 278 –287.

[13] J. Shawe-Taylor and N. Cristianini, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, 2000.

[14] P. Dreuw, S. Jonas, and H. Ney, "White-Space Models for Offline Arabic Handwriting Recognition," in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1–4.

[15] P. Dreuw, G. Heigold, and H. Ney, "Confidence-Based Discriminative Training for Model Adaptation in Offline Arabic Handwriting Recognition," in *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, 2009, pp. 596–600.

[16] F. Valente, J. Vepa, C. Plahl, C. Gollan, H. Hermansky, and R. Schlüter, "Hierarchical Neural Networks Feature Extraction for LVCSR system," in *Proceedings of the Interspeech*, 2007, pp. 42–45.

[17] H. Hermansky and S. Sharma, "TRAPs - Classifiers of Temporal Patterns," in *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP)*, 1998, pp. 1003–1006.

[18] M. Hamdani, H. El Abed, T. M. Hamdani, V. Märgner, and A. M. Alimi, "Improving a HMM-based Off-line Handwriting Recognition System using MME-PSO Optimization," in *Proceedings of the Document Recognition and Retrieval XVIII*, 2011.

[19] M. Hamdani, H. El Abed, M. Kherallah, and A. M. Alimi, "Combining Multiple HMMs using On-line and Off-line Features for Off-line Arabic Handwriting Recognition," in *Proceedings of the 10th International Conference on Document Analysis and Recognition (ICDAR)*, 2009, pp. 201–205.

[20] H. El Abed and V. Märgner, "Comparison of different preprocessing methods for offline recognition of handwritten Arabic words," in *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR)*, vol. 2, 2007, pp. 974 –978.

[21] ——, "Improvement of Arabic Handwriting Recognition Systems – Combination and/or Reject?" in *Proceedings of the Document Recognition and Retrieval XVI*, no. 7247–10, 2009.