ICDAR 2009 Arabic Handwriting Recognition Competition

Volker Märgner, Haikal El Abed
Technische Universitaet Braunschweig,
Institute for Communications Technology (IfN),
Schleinitzstrasse 22, 38106 Braunschweig, Germany
{v.maergner, elabed}@tu-bs.de

Abstract

This paper describes the Arabic handwriting recognition competition held at ICDAR 2009. This third competition (the first was at ICDAR 2005 and the second at ICDAR 2007) again used the IfN/ENIT-database with Arabic handwritten Tunisian town names. Today, more than 82 research groups from universities, research centers, and industry are working with this database worldwide. This year, 7 groups with 17 systems were participating in the competition. The systems were tested on known data and on two data sets which are unknown to the participants. The systems were compared based on the most important characteristic: the recognition rate. Additionally, the relative speed of the different systems was compared. A short description of the participating groups, their systems, and the results achieved are finally presented.

1. 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 [14], an improvement of published systems could be observed. This paper presents the results of the third competition of Arabic handwritten word recognition systems. The results of this third competition were presented during the ICDAR 2009 conference in Barcelona, Spain. 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 2007, this year 7 groups with 17 systems participated in the competition: one group was also a participant in the last competition, while all 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 as those in the 2007 competition [13].

This paper is organized as follows: In Section 2 the database and the test sets are presented shortly. Section 3 presents the participating groups and gives a short description of the submitted systems. Section 4 describes the tests and the results achieved by the different systems. Finally the paper ends with some concluding remarks.

2. Training and Test Sets

2.1. 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 [17], more than 82 groups in about 31 countries are working today (i.e., at the beginning of 2009) 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 writers. The words written are 937 Tunisian town/village names [14]. Each writer filled one to five forms with preselected town/village names and the corresponding post code. Ground truth was added to the image data automatically and verified manually.

2.2. The Test Datasets

The test datasets which are unknown to all participants were collected for the tests of the ICDAR 2007 competition [13]. 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 1).

Set *f* was collected in Tunisia, while set *s* was collected in the United Arab Emirates (UAE) at the University of

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.256

Table 1. Features of datasets f, s ,t and t_1

set	names	characters	PAWs
f	8671	64781	32918
s	1573	11922	6109
t	1000	7921	4252
t_1	100	821	412

Table 2. 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			

Sharjah. Table 2 shows the frequency of PAWs (Parts of Arabic Words) within each name of the new datasets f and s. The sets t and t_1 are subsets of sets a to f used to measure the processing time of the systems in the competition environment.

3. Participating Systems

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.

3.1. UOB-ENST

This system was submitted by Chafic Mokbel and Ramy Al-Hajj from the University of Balamand (UOB), Lebanon and Laurence Likforman-Sulem from Telecom ParisTech, France. The realization of the handwritten word recognition system is a HMM-based system without pre-segmentation.

This system participated as well in ICDAR 2005 and 2007 competitions. In this year 4 variants of the UOB-ENST system were presented: a basic variant similar to that presented at ICDAR 2005 [1] and two advanced systems that it better in handling the slanted handwriting [3]. The system is a HMM-based system, of analytic type without pre-segmentation. It uses the general purpose HMM (Hidden Markov Model) toolkit called HCM [16]. The development of the handwriting systems was carried within the PhD thesis of Ramy El-Hajj and in tight collaboration with ENST-Paris. The advanced version was developed to reduce the recognition errors coming from slanted handwriting and the erroneous positions of diacritical points and marks. The proposed system comprises two stages: the first stage is for recognition and classification based on the technique of slanted windows (with different angles) to extract the features, and the second stage comprises of a combined post-processing steps. Different combination methods were used and examined such as: majority vote rules and Borda count combination operator. In addition, a combination method based on an ANN with Multi-Layer Perceptron is used [2].

3.2. REGIM

The Research Group on the Intelligent Machines (REGIM) at Ecole Nationale d'Ingénieurs de Sfax (ENIS), University of Sfax, Tunisia participated with one system, submitted by Abdelkarim ElBaati, Monji Kherallah, Houcine Boubaker, Mahdi Hamdani, Adel M. Alimi, and Abdellatif Ennaji from LITIS, University of Rouen, France. This system is based on the restoration of the temporal order of the off-line trajectory of a word [6]. To benefit from dynamic information, a sampling operation by the consideration of trajectory curvatures is calculated. Studies showed that there is a correlation between the angular velocity $V_\sigma(t)$ and the curve $C(t)$. Moreover, they propose, in this process, to sample the rebuilt trajectory with fixed time interval (sampling step), by traversing it with a curvilinear velocity that checks the law of two thirds [6]. Subsequently the curvilinear velocity signal uses the beta-elliptical modeling, which was developed for on-line systems [11] to calculate features, for feature extraction. For recognition a HMM-based system using HTK is used [9].

3.3. MDLSTM

These systems were submitted by Alex Graves from Techische Universität München, München, Germany. This multilingual handwriting recognition system is based on a hierarchy of multidimensional recurrent neural networks [7]. It can accept either on-line or off-line handwriting data, and in both cases works directly on the raw input without any preprocessing or feature extraction. It uses the multidimensional Long Short-Term Memory network architecture [7], an extension of Long Short-Term Memory to data with more than one spatio-temporal dimension. The basic structure of the system, including the hidden layer architecture and the hierarchical subsampling method is described in [8].

3.4. LSTS

This system was submitted by Samia Snoussi-Maddouri from LSTS group at the Ecole Nationale d'Ingénieurs de Tunis (ENIT), Tunis, Tunisia. This system is called Transparent Neural Network (TNN), combining Global and Local Vision Modeling (GVM-LVM) of the word [19]. In the forward propagation movement, the GVM proposes a list of words containing structural features characterizing the presence of some letters in the word. Then, in the back-propagation movement, these letters are confirmed or not according to their proximity to corresponding printed letters. The correspondence between the letter shapes and the corresponding printed letters is performed by LVM using the correspondence of their normalized Fourier descriptors [18]. The particularities of the TNN-DF are that it does not use any training steps. It can be used for different languages or different lexicon by a simple change of the content of each layers.

3.5. A2iA

The A2iA Arab-Reader system was submitted by Fares Menasri and Christopher Kermorvant (A2iA SA, France), Anne-Laure Bianne (A2iA SA and Telecom ParisTech, France), and Laurence Likforman-Sulem (Telecom ParisTech, France). This system is a combination of two different word recognizers, both based on HMM. The first one is a Hybrid HMM/NN with grapheme segmentation [12]. It is mainly based on the standard A2iA word recognizer for Latin script, with several adaptations for Arabic script [15]. The second one is a Gaussian mixture HMM based on HTK, with sliding windows (no explicit pre-segmentation). The computation of features was greatly inspired by Al-Hajj works on geometric features for Arabic recognition [2]. The results of the two previous word recognition systems are combined so as to compute the final answer [2].

3.6. LITIS-MIRACL

This system was submitted by Yousri Kessentini (LITIS and MIRACL), Thierry Paquet (LITIS, University of Rouen, France), and AbdelMajid Benhamadou (MIRACL, University of Sfax, Tunisia). This word recognition system is based on a multi-stream segmentation free HMM. Two feature vector sequences are created using a sliding window, and they are simultaneously decoded according to the multi-stream formalism. One stream is composed of density features while the other is made of contour features [10].

3.7. RWTH-OCR

These systems were submitted by Philippe Dreuw, Stephan Jonas, Georg Heigold, David Rybach, and Hermann Ney from RWTH Aachen University, Human Language Technology and Pattern Recognition, Aachen, Germany. Without any preprocessing of the input images, simple appearance-based image slice features X_t at every time step $t = 1, \dots, T$ which are augmented by their spatial derivatives in horizontal direction $\Delta = X_t - X_{t-1}$, are extracted. In order to incorporate temporal and spatial context into the features, 7 consecutive features in a sliding window, which are later reduced by a PCA transformation matrix, are concatenated. The System-1 is a multi-pass system. The first-pass system is built using a modified maximum mutual information training criterion. The second-pass is automatically built using a novel unsupervised confidence based discriminative training criterion on the output of the first-pass system to automatically adapt the model to the unknown testing data [4]. System-2 is a HMM based handwriting recognition system, in which Viterbi is trained using the maximum-likelihood training criterion. A lexicon with multiple writing variants, where the white spaces between the pieces of Arabic words are explicitly modeled as proposed in [5], is used.

4. Tests and Results

We evaluated the performance of the 17 different Arabic handwriting recognition systems in three 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. In a third step, the speed performance of the systems was compared on two subsets *t* and t_1.

The most important results of our tests are shown in Table 3. For each test, the best result is marked in bold font. More details will be presented at ICDAR 2009 Conference.

4.1. 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 7 systems with a recognition rate better than 90% on set d and 83% on set e. Four systems have a recognition rate less than 70% on set d and three systems less than 60% on set e. It is interesting to see that the relative position of all systems is the same for sets *d* and *e*.

4.2. 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 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 3. 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. The average recognition time in ms per image on subsets t and t_1 is shown in the last two columns. (G-ID: Group ID, S-ID: System ID).

G-ID	S-ID	set d	set e	set f_a	set f_f	set f_g	set f			set s			time (**ms**)	
		top 1	top 1	top 1	top 1	top 1	top 1	top 5	top 10	top 1	top 5	top 10	set t	set t_1
UOB-ENST	1	92.52	85.38	83.57	84.77	85.09	82.07	89.74	91.22	69.99	81.44	84.68	812.69	841.25
	2	89.06	81.85	79.49	80.90	81.11	78.16	89.06	91.88	65.61	81.44	85.95	2365.48	2755.01
	3	89.84	83.52	80.89	82.15	82.17	79.55	90.60	92.16	67.83	83.47	86.65	2236.58	2754.08
	4	92.59	86.28	85.42	86.96	87.21	83.98	91.85	93.00	72.28	85.19	87.92	2154.48	2651.57
REGIM	5	79.52	63.53	58.81	59.27	60.42	57.93	73.43	78.10	49.33	65.10	71.14	1564.75	1712.15
Ai2A	6	93.90	87.25	86.73	88.54	89.36	85.58	92.57	94.12	70.44	82.01	84.87	1056,98	956,82
	7	94.92	82.21	83.53	84.86	84.67	82.21	91.24	92.47	66.45	80.52	83.13	**519,61**	1616,82
	8	97.02	91.68	90.66	91.92	92.31	**89.42**	95.33	95.94	**76.66**	88.01	90.28	2583,64	1585,49
MDLSTM	9	99.72	98.64	92.59	93.79	94.22	91.43	96.11	96.61	78.83	87.98	90.40	115.24	122.97
	10	99.60	97.60	92.58	94.03	94.40	91.37	96.24	96.61	78.89	88.49	90.27	**114.61**	122.05
	11	99.94	99.44	94.68	95.65	96.02	**93.37**	96.46	96.77	**81.06**	88.94	90.72	371.85	467.07
RWTH-OCR	12	99.91	98.71	86.97	88.08	87.98	85.51	93.32	94.61	71.33	83.66	86.52	17845.12	18641.93
	13	99.79	98.29	87.17	88.63	88.68	85.69	93.36	94.72	72.54	83.47	86.78		
	14	99.79	98.29	87.17	88.63	88.68	**85.69**	93.36	94.72	72.54	83.47	86.78		
	15	96.72	91.25	86.97	88.08	87.98	83.90	-	-	65.99	-	-	**542.12**	560.44
LITIS-MIRACL	16	93.04	85.46	83.29	84.51	84.35	82.09	90.27	92.37	**74.51**	86.14	88.87	143269.81	145157.23
LSTS	17	18.58	14.75	15.34	16.00	15.65	15.05	29.58	35.76	11.76	23.33	29.62	612.56	685.42
Results of the 3 best systems at ICDAR 2007														
Siemens	08	94.58	87.77	88.41	89.26	89.72	87.22	94.05	95.42	73.94	85.44	88.18	109.406	125.31
MIE	06	93.63	86.67	84.38	85.21	85.56	83.34	91.67	93.48	68.40	80.93	83.73	188.439	210.55
UOB-ENST	11	92.38	83.92	83.39	84.93	85.18	81.93	91.20	92.76	69.93	84.11	87.03	2172.55	2425.47

Table 3 shows some interesting results: (1) Three systems 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 the same and another system shows even a better results on set *f* than on set *e*. The best system has a recognition rate of 2% higher than the second-best system, and the absolute value is again much higher than that in the competition 2007. It is obvious that a further improvement of the systems performance since ICDAR 2007 competition is accomplished.

4.3. 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 9 systems are better than 70%. This is a loss of about 10% compared to the recognition rate on set *f*, but it shows that the generalization ability of these systems is not too bad.

4.4. Speed Tests (sets t and t_1)

The average processing time per image on the two test sets t (1000 images) and t_1 (100 images) respectively is shown in the last two columns of Table 3. A substantial difference in speed can be observed. The slowest system is more than 1000 times slower than the fastest one. An average processing time of 114 ms per image is a good result and it combines high speed with very good recognition results (second best recognition result on set f). The total processing time was 50 days, 16 h, 30 min, and 25.186 s.

5. Conclusions

The competition results show that Arabic handwriting recognition systems in this third competition made a remarkable further progress. Most of the participating systems show a very high accuracy and some also with a very high speed. Details and specific features of the systems cannot be presented in this short paper. The system 11 (MDLSTM) is the winner of this competition. The system 10 (MDLSTM) is the system with the shortest processing time.

References

[1] R. Al-Hajj, L. Likforman-Sulem, and C. Mokbel. Arabic handwriting recognition using baseline dependant features and hidden Markov modeling. In *8th ICDAR*, volume 2, pages 893–897, 2005.

[2] R. Al-Hajj, L. Likforman-Sulem, and C. Mokbel. Combining Slanted-Frame Classifiers for Improved HMM-Based Arabic Handwriting Recognition. *T-PAMI*, 31, 2009.

[3] R. Al-Hajj, C. Mokbel, and L. Likforman-Sulem. Reconnaissance de l'écriture arabe cursive: Combinaison de classifieurs MMCs á fenêtres orientées. In *CIFED*, pages 271–276, 2006.

[4] P. Dreuw, G. Heigold, and H. Ney. Confidence-based discriminative training for model adaptation in offline Arabic handwriting recognition. In *Proc. ICDAR*, 2009.

[5] P. Dreuw, S. Jonas, and H. Ney. White-space models for offline Arabic handwriting recognition. In *Proc. 19th Inter. Conf. on Pattern Recognition (ICPR)*, pages 1–4, 2008.

[6] A. Elbaati, M. Kherallah, H. El Abed, A. Ennaji, and A. M. Alimi. Arabic Handwriting Recognition Using Restored Stroke Chronology. In *ICDAR*, 2009.

[7] A. Graves. *Supervised Sequence Labelling with Recurrent Neural Networks*. PhD thesis, Fakultat für Informatik – Technische Universität München, 2007.

[8] A. Graves and J. Schmidhuber. Offline handwriting recognition with multidimensional recurrent neural networks. In *Advances in Neural Information Processing Systems 21*, 2009.

[9] 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 *ICDAR*, 2009.

[10] Y. Kessentini, T. Paquet, and A. Benhamadou. A Multi-Stream HMM-based Approach for Off-line Multi-Script Handwritten Word Recognition. In *Proc. 10th ICFHR*, pages 147–152, 2008.

[11] M. Kherallah, L. Haddad, A. M. Alimi, and A. Mitiche. Towards The Design Of Handwriting Recognition System By Neuro-Fuzzy And Beta - Elliptical Approaches. In *Proc. of the Artficial Intelligence Applications & Innovations (AIAI)*, pages 187–196, 2004.

[12] S. Knerr and E. Augustin. A neural network-hidden Markov model hybrid for cursive word recognition. In *Proc. Fourteenth International Conference on Pattern Recognition*, volume 2, pages 1518–1520 vol.2, 1998.

[13] V. Märgner and H. El Abed. ICDAR 2007 – Arabic Handwriting Recognition Competition. In *Proceedings of the* 9^{th} *ICDAR*, volume 2, pages 1274–1278, 2007.

[14] V. Märgner, M. Pechwitz, and H. El Abed. ICDAR 2005 – Arabic Handwriting Recognition Competition. In *8th ICDAR*, volume 1, pages 70–74, 2005.

[15] F. Menasri, N. Vincent, E. Augustin, and M. Cheriet. Un systéme de reconnaissance de mots arabes manuscrits horsligne sans signes diacritiques. In *Proc. CIFED*, pages 121–126, 2008.

[16] C. Mokbel, H. A. Akl, and H. Greige. Automatic speech recognition of Arabic digits over Telefone network. In *RTST*, 2002.

[17] M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, and H. Amiri. IFN/ENIT- database of handwritten Arabic words. In *CIFED*, pages 127–136, 2002.

[18] S. Snoussi Maddouri and H. Amiri. Une Méthode de Reconnaissance de Mots Manuscrits Arabes par Réseaux de Neurones Transparent. In *Conf. Inter. en Informatique*, 1999.

[19] S. Snoussi Maddouri and H. Amiri. Combination of local and global vision modelling for Arabic handwritten words recognition. In *IWFHR*, pages 128–135, August 2002.