ICDAR 2009 Online Arabic Handwriting Recognition Competition

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

Monji Kherallah, Adel M. Alimi
Ecole Nationale d'Ingènieurs de Sfax (ENIS),
Research Group on Intelligent
Machines (REGIM)
Sfax, Tunisia
monji.kherallah@enis.rnu.tn
adel.alimi@ieee.org

Abstract

This paper describes the Online Arabic handwriting recognition competition held at ICDAR 2009. This first competition uses the ADAB-database with Arabic online handwritten words. This year, 3 groups with 7 systems are participating in the competition. The systems were tested on known data (sets 1 to 3) and on one test dataset which is unknown to all participants (set 4). The systems are compared on the most important characteristic of classification systems, the recognition rate. Additionally, the relative speed of the different systems were compared. A short description of the participating groups, their systems, the experimental setup, and the performed results are presented.

1. Introduction

Automatic recognition of handwritten words can be classified into two different approaches. The first approach, offline, uses the images as input for the recognition steps. The second approach, online, uses the trace of a pen for the classification and recognition of the input information. This field remains a challenging task even though the latest improvements of recognition methods and systems are very promising. Especially for the automatic recognition of online Arabic handwritten words a lot of work has still to be done. Compared to English text where handwritten words and numbers have been publicly available for a long time (e.g. UNIPEN) the situation for Arabic today is quite different. In the case of Arabic handwritten words many papers use a specific, more or less small datasets of their own [18], The series of competition for Arabic handwriting recognition systems has shown a positif effect for the improvement of recognition systems. At the first Competition organized on the ICDAR 2005 [17] only 5 systems

have participated, in the second competition [16] more than 14 systems have participated and the performances are better. This improvement of the performance of recognition systems [3] motivates us to organize the first Online Arabic handwriting recognition competition. One important step to organize such competition was the collection of a new database of online Arabic handwritten text.

This paper is organized as follows. In Section 2 the new ADAB-database and the competition test sets are presented in some detail. Section 3 presents the participating groups and a short description of their systems. Section 4 describes the different tests and discuss the results performed with the different systems. Finally the paper ends with some concluding remarks.

2. The ADAB-database

The database ADAB (Arabic DAtaBase) was developed to advance the research and development of Arabic online handwritten text recognition systems. This database is developed in a cooperation between the Institute for Communications Technology (IfN) and the Ecole Nationale d'Ingènieurs de Sfax (ENIS), Research Group on Intelligent Machines (REGIM), Sfax, Tunisia.

The database in version 1.0 consists of 15158 Arabic words handwritten by more than 130 different writers , most of them selected from the narrower range of the l'Ecole Nationale d'Ingènieurs de Sfax (ENIS). The text written is from 937 Tunisian town/village names. We have developed special tools for the collection of the data and verification of the ground truth. These tools give the possibilities to record the online written data, to save some writer information, to select the lexicon for the collection, and re-write and correct wrong written text. Ground truth was added to the text information automatically from the selected lexicon and verified manually.

IEEE
computer
society

2.1. Training Data

The ADAB-database in version 1.0 is spilt in 3 sets. Details about the number of files, words, characters, and writers for each set 1 to 3 are shown in Table 1.

Table 1. Features of ADAB-datasets 1, 2 , and 3

set	files	words	characters	writers
1	5037	7670	40500	56
2	5090	7851	41515	37
3	5031	7730	40544	39
Sum	15158	23251	122559	132

2.2. New Datasets

New data which are unknown to all participants were collected for the processing of the ICDAR 2009 competition. Again the words are selected from the same lexicon, however written by new writers. These data are include into the set 4. Sets t, and t_1 are generated from all sets 1 to 3 to measure the processing time of the participant systems. Table 2 shows the size and the numbers of characters. Set 4 is collected in Braunschweig, with writers coming from Tunisia.

Table 2. Features of datasets 4, t and t_1

set	files	words	characters	writers
4	1562	2418	12648	24
t	450	609	3588	10
t_1	45	61	327	4

3. Participating Systems

The following section gives a brief description of the systems submitted to the competition. Each system description has been 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 source information provided. Some groups (2 industrial groups and 2 research groups) have decided after the first tests to not participate at this competition. It is the first time to compare online Arabic handwriting recognition systems and we hope that next time all participants will give detailed information about their participation, when it is clear that the competition is objective and useful for participants and for the research community.

3.1. VisionObjects

They have built a cursive Arabic handwriting recognition system for this competition based on MyScript® handwriting recognition technology. The overall system follows the following concepts:

- use of a modular and hierarchical recognition system,

- use of soft decisions (often probabilistic) and deferred decisions by means of considering concurrent hypotheses in the decision paths,

- use of complementary information at all stages of the recognition process, and

- use of global optimization criteria, making sure that the recognizer is trained in order to perform optimally on all levels.

The processing chain of the recognizer starts out with some of the usual preprocessing operations, such as ink smoothing and reference line detection. Then the on-line handwriting is pre-segmented into strokes and sub-strokes. The general idea is to over-segment the signal and let the recognizer decide later on where the boundaries between characters and words are. Here, specific techniques for processing diacritical marks have been employed to assure the proper association of letters and their diacritical marks. This segmentation stage is followed by the feature extraction stage. Feature sets use a combination of on-line and off-line information at various resolutions, including some higher level structural features. The feature sets are processed by a set of character classifiers, which use Neural Networks and other pattern recognition paradigms. The total number of characters classes is 150, which corresponds to the number of Arabic letters multiplied by the number of different shapes for each letter (initial, medial, final and isolated), plus some other symbols encountered in Tunisian cities like digits or the Latin letter 'V'. All the information accumulated in the various processing steps is then processed by dynamic programming on the word and sentence level in order to generate character, word, and sentence level candidates with corresponding confidence scores. A global discriminant training scheme on the word level with automatic learning of all classifier parameters and meta-parameters of the recognizer, is employed for the overall training of the recognizer. For the recognition process, a lexicon containing around 1000 Tunisian city names is employed.

We have designed the recognizer according to two different criteria. The first system (VisionObjects-1) provides the best accuracy whereas the second system (VisionObjects-2) is faster in exchange for a somewhat lower accuracy.

3.2. MDLSTM

The MDLSTM system is submitted by Alex Graves (TU Munich, Germany). It is a multilingual handwriting recognition system based on a hierarchy of multidimensional recurrent neural networks [7, 5]. It can accept either on-line or off-line handwriting data, and in both cases works directly on the raw input (i.e. the pixel values or the sequence of pen positions) with no preprocessing or feature extraction. It uses the multidimensional Long Short-Term Memory network architecture [7, 5], an extension of Long Short-Term Memory [11] to data with more than one spatio-temporal dimension. In the case of handwriting recognition the networks are either one-dimensional (for on-line data) or two-dimensional (for off-line images). The basic structure of the system, including the hidden layer architecture and the hierarchical subsampling method is described in [9]. However the exact parameters (e.g. size of hidden layers, size of subsampling blocks etc.) varied from experiment to experiment. In addition the choice of output layer and objective function used for training depend on the network task.

Connectionist Temporal Classification [6, 5] is an recurrent neural network output layer designed for labelling sequences of data whose segmentation is ambiguous or difficult to determine, e.g. speech signals or cursive handwriting. It trains the network to map directly from the input sequence to a probability distribution over output label sequences, and therefore does not require either pre-segmentation or post-processing.

For the online Arabic recognition competition two different recognizers were created. The first one ('MDLSTM-1') uses the online data to create off-line images, then transcribes these using the standard hierarchy of 2D RNNs described above. The second ('MDLSTM-2') transcribes the online data directly, using a hierarchy of 1D RNNs. 1D RNNs have previously proved effective for online handwriting [5, 8], but this is the first time a hierarchical structure has been used.

3.3. REGIM

REGIM-HTK The REGIM-HTK is submitted by Abdelkarim ElBaati, Monji Kherlallah, Houcine Boubaker, Mahdi Hamdani, and Adel M. Alimi. This system is based on the features extracted from the temporal order of the trajectory of a word [4].

A study made on the neuronal and muscular effect shows that the pen velocity decreases at the begin and end strokes and in significant angular variations of the curve. To benefit from dynamic information we make a sampling operation by the consideration of trajectory curvatures. Studies established initially by Lacquaniti and al. [15], then by Viviani and al. [19], showed the existence of a correlation between the angular velocity $V_\sigma(t)$ and the curve $C(t)$. This correlation is named the law of two thirds power. The authors propose, in this process, to sample with fixed time interval (sampling step), the rebuilt trajectory, by traversing it with a curvilinear velocity that checks the law of two thirds (correspondence between velocity and curvature radius) [4].

After that, the curvilinear velocity signal and use the beta-elliptical modeling [13] to calculate the features for each word. This approach is validated by HTK-based recognition system [10].

REGIM-CV The REGIM-CV system is submitted by Monji Kherlallah, Fatma Bouri, Houcien Boubaker, and Adel M. Alimi (REGIM, University of Sfax, Tunisia). This system is based on the idea to observe visually the writing process on an ordinary paper and to automatically recover the pen trajectory from numerical tablet sequences. On the basis of this work, we developed a handwriting recognition system based on visual coding and fitness evaluation function. The system is applied on Arabic script. The first step of the encoding system is consisting of the smoothing, normalization, base line detection, beta elliptical modeling and visual indices attribution [14, 1, 2]. The second step is based on visual indices similarity in order to calculate the evaluation function [12]. A hierarchical architecture of the ADAB-database is designed as a learning phase. This repartition is depending on stroke number inspection. The authors optimize the times cooling of our system to give the final output (Proposed words).

REGIM-CV-HTK The REGIM-HTK-CV system is submitted by Mahdi Hamdani, Lobna Haddad, Houcine Boubaker, Monji Kherlallah and Adel M. Alimi (REGIM, University of Sfax). This recognition system has as propose to observe visually the writing process on an ordinary paper and to automatically recover the pen trajectory from numerical tablet sequences. On the basis of this work, we developed a handwriting recognition system based on visual coding and HTK system. The system is applied on Arabic script. The first step of the encoding system is consisting of the smoothing, normalization, base line detection, beta elliptical modeling and visual indices attribution [14, 12, 1, 2]. The second step is based on HTK classifier [10].

4. Tests and Results

We evaluated the performance of the 7 different Arabic handwriting recognition systems in three steps. In a first step we used a subset and then the training data sets 1 to 3 of the ADAB-database for a function check of the systems on our competition environment. After the confirmation from

Table 3. Recognition results in % of correct recognized images on reference datasets 1, 2 and 3 and on a subset of the new dataset set 4.

System	set 1			set 2			set 3			set 4*		
	top 1	top 5	top 10	top 1	top 5	top 10	top 1	top 5	top10	top 1	top 5	top10
MDLSTM-1	99.36	99.94	99.96	99.42	99.96	100.00	99.52	99.94	99.94	95.70	98.93	100
MDLSTM-2	98.55	99.60	99.66	98.77	99.88	99.92	98.89	99.64	99.70	95.70	98.93	100
VisionObjects-1	99.46	99.70	99.70	99.82	99.94	99.96	99.58	99.76	99.76	**98.99**	100	100
VisionObjects-2	99.29	99.60	99.60	99.51	99.74	99.74	99.26	99.56	99.56	98.99	100	100
REGIM-HTK	57.87	72.89	77.03	54.26	66.38	71.06	53.75	72.31	76.22	52.67	63.44	64.52
REGIM-CV	100	100	100	94.39	96.06	96.06	96.28	97.14	97.52	13.99	31.18	37.63
REGIM-CV-HTK	28.85	51.92	55.77	35.75	58.30	64.26	30.60	52.80	62.80	38.71	59.07	69.89

the participants that this first tests are conform to the results performed on participant local environment, we have started the competition tests. In this second step we used the new test dataset set 4, unknown to all participants to test the recognition rates of participant systems. In a third step the processing performance of the systems was compared on two data subsets t and t_1.

4.1. General Remarks

To reach an optimal result and due to different processing time, the competition was organized in a closed mode. All participants sent us running versions of their recognition systems trained on all data (sets 1 to 3) of the ADAB-database. The systems can be classified in two classes depending on the operating system: 3 systems are developed under Linux, 3 under Microsoft Windows environment. A first set with 100 input files selected from the training sets was used to test the basic functionality of the systems on our competition environment. All systems passed this test. The second function test was performed on the training sets 1 to 3 of the ADAB-database with all systems.

4.2. Recognition Results

The most important results of our tests are shown in Tables 3 to 4. For each test the best result is marked in bold font. More details will be presented in the special competition session at ICDAR 2009 Conference.

4.3. Tests with known Data (sets 1 to 3)

The comparison of the systems based on the results of sets 1 to 3, which are part of the training set, shows 3 systems with a recognition rate better than 99% on sets 1 to 3. Tow systems have a recognition rate less than 60% on sets 1 to 3. It is interesting to see that the results of the system REGIM-CV on the sets 1 to 3 on top 1.

Table 4. The average recognition time in ms per image on subsets t and t_1.

System	set t	set t_1
MDLSTM-1	1377.22	1574.13
MDLSTM-2	1712.45	2222.04
VisionObjects-1	172.67	179.84
VisionObjects-2	**69.41**	81.93
REGIM-HTK	6402.24	4626.58
REGIM-CV	7521.65	7120.2
REGIM-HTK-CV	3571.25	3158.21

4.4. Main Test (set 4)

The most important test to compare the different systems is of course the test using the new set 4. The features of this set should be similar to sets 1 to 3, as it was collected with writer from the same country. Table 3 shows some interesting results:

- Four systems have more than 95% recognition rate on set 4,

- the variation of best system between top 1, top 5 and top10 is preserved in all sets,

- one system lost about 75% compared to sets 1 to 3 and one system has better results on test dataset 4 as on training datasets 1 to 3.

The best system has a recognition rate of almost 4% higher than the second-best system.

4.5. Speed Tests (sets t and t_1)

The average processing time per name on the two test sets t (450 files) and t_1 (45 files) respectively is shown

in Table 4. A substantial difference in speed can be observed. The slowest system is about 100 times slower than the fastest one. An average processing time of 69ms per name image is a very good result.

5. Conclusions

The competition results show that Online Arabic handwriting recognition systems made a remarkable further progress. Most of the participating systems show a very high accuracy and some also a very high speed. Details and specific features of the systems cannot be presented in this short paper. The systems **VisionObjects** are the winner of this competition. The system VisionObjects-2 is the system with the shortest average processing time.

References

[1] H. Boubaker, M. Kherallah, and A. M. Alimi. New strategy for the on-line handwriting modelling. In *Proc. Ninth International Conference on Document Analysis and Recognition (ICDAR)*, volume 2, pages 1233–1247, 23–26 Sept. 2007.

[2] H. Boubaker, M. Kherallah, and A. M. Alimi. New algorithm of straight or curved baseline detection for short arabic handwritten writing. In *International Conference on Document Analysis and Recognition (ICDAR)*, 2009.

[3] H. El Abed and V. Märgner. Base de données et compétitions - outils de développement et d'évaluation de systèmes de reconnaissance de mots manuscrits arabes. In *10th Colloque International Francophone sur l'Ecrit et le Document , CIFED*, pages 103–108, 2008.

[4] A. Elbaati, M. Kherallah, H. El Abed, A. Ennaji, and A. M. Alimi. Arabic handwriting recognition using restored stroke chronology. In *International Conference on Document Analysis and Recognition (ICDAR)*, 2009.

[5] A. Graves. *Supervised Sequence Labelling with Recurrent Neural Networks*. Ph.D. in Informatics, Fakultat für Informatik – Technische Universität München, Germany, 2008.

[6] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the International Conference on Machine Learning, ICML 2006*, Pittsburgh, USA, 2006.

[7] A. Graves, S. Fernàndez, and J. Schmidhuber. Multidimensional recurrent neural networks. In *Proceedings of the International Conference on Artificial Neural Networks*, 2007.

[8] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, and J. Schmidhuber. A novel connectionist system for unconstrained handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):855–868, 2009.

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

[10] 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 *International Conference on Document Analysis and Recognition (ICDAR)*, 2009.

[11] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[12] M. Kherallah, F. Bouri, and A. M. Alimi. On-line arabic handwriting recognition system based on visual encoding and genetic algorithm. *Engineering Applications of Artificial Intelligence*, 22:153–170, 2009.

[13] 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 *Proceedings of the Artficial Intelligence Applications & Innovations (AIAI), 18th IFIP World Computer Congress*, pages 187–196, 2004.

[14] M. Kherallah, L. Haddad, A. M. Alimi, and A. Mitiche. Online handwritten digit recognition based on trajectory and velocity modeling. *Pattern Recognition Letters*, 29:580–594, 2008.

[15] F. Lacquaniti, C. Terzuolo, and P. Viviani. The law relating the kinematic and figural aspects of drawing movements. *Acta Psychol (Amst)*, 54(1-3):115–130, Oct 1983.

[16] V. Märgner and H. El Abed. ICDAR 2007 Arabic Handwriting Recognition Competition. In *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR)*, volume 2, pages 1274–1278, 2007.

[17] V. Märgner, M. Pechwitz, and H. El Abed. ICDAR 2005 arabic handwriting recognition competition. In *8th International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 70–74, 2005.

[18] J. Sternby, J. Morwing, J. Andersson, and C. Friberg. Online arabic handwriting recognition with templates. *Pattern Recognition*, page in Press, 2009.

[19] P. Viviani and M. Cenzato. Segmentation and coupling in complex movements. *Journal of Experimental Psychology: Human, Perception and Performance*, 11(6):828–845, Dec 1985.