Arabic Word and Text Recognition -- Current Developments

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

Arabic word and text recognition methods for printed or handwritten words are known since many years. Such a recognition technology is a very important tool used in many natural language processing systems. This application makes it important to know the performance of existing recognition systems. We present in the first part of the paper a state of the art of Arabic word and text recognition systems and the used classification techniques. In the second part we discuss how evaluation methods and competitions support a user to select a system, which is best qualified for his application, but also support the development of text recognition systems and methods. Based on the results of the Arabic handwriting recognition competition held at ICDAR 2007 we show a concept to develop efficient recognition systems. On the basis of the actual situation of research, future trends are described in the last part of this paper.

Introduction

The OCR (Optical Character Recognition) research started early in the 1930's and first address reading machines for Latin characters are available since the 1960's. Due to the connectedness of characters within an Arabic printed word it took more than ten further years before the first papers on Arabic word recognition were published. Research work on Arabic Optical Text Recognition increased considerably since the 1980's. First systems for Arabic printed text are available at the market in the 1990's. Today a few systems are available at the market but the only independent system comparison was made 15 years ago. Compared to the high quality and wide spread usage of OCR systems for Latin characters Arabic OCR still has to be developed, especially for the case of handwritten words.

The special style of Arabic printed text where the characters are connected (with some exceptions) within a word is the normal case for cursively handwritten words in many different languages, too. That is why methods developed for words handwritten in languages other than Arabic as for instance methods based on a so called Hidden Markov Model (HMM) or on a Neural Network (NN) is applied to Arabic handwritten word recognition systems successfully. The state of the art in Arabic handwritten word recognition has been presented recently Lorigo & Govindaraju(2006) and Märgner & El Abed(2008) and direct comparison of systems was performed twice on the basis of a competition during the ICDAR 2005 and ICDAR 2007 Märgner et al.(2005) and Märgner & El Abed(2007) conferences. Especially the competition in the year 2007 showed very good results, and a progress compared to the competition in 2005 can be observed.

This paper gives a survey of Arabic text recognition systems, especially for handwritten words, competitions, and evaluation tools necessary to improve recognition systems, in general, and Arabic handwritten text recognition in particular. Evaluation and competition are presented as important tools for the development of recognition systems and a state of the art of Arabic Text Recognition Systems is given. Several aspects important for the further development of Arabic handwriting recognition are presented.

Word and Text Recognition Systems

Arabic OCR for printed characters was a research topic in the 1990th. A comparison of published papers was reported in 2000, in Ahmed & Al-Ohali(2000). Today three Arabic OCR systems (Sakhr, IRIS, ABBY) are available at the market and printed Arabic recognition is not a topic of research any more. Nevertheless the only evaluation of Arabic OCR systems was made in the year 1998, 10 years ago. The performance of OCR systems is more or less unknown, the user has to make tests of his own or trust the advertisement of the companies selling the product. It follows that many people don't know what is available and if the problems they have can be solved by products available at the market. An independent test and comparison of the performance of OCR systems is highly recommended.

The situation for recognition of Arabic handwritten Arabic words or text is quite different. This is an important research topic of today, every year at the conferences ICDAR and ICFHR new results are reported and since 2005 the performance of systems is compared in a competition at the ICDAR conferences. This competition was made possible with the publication of the IFN/ENIT-database consisting of handwritten Arabic

words. Table 1 gives an overview of recently published recognition systems and their accuracy. Especially these systems which are tested with the IfN/ENIT-database can be objectively compared and it becomes possible to learn which features or classifiers are better than others. Questions about segmentation of words into characters and the usage and size of a lexicon are most important and have to be answered for the development of a successful recognition system. From Table 1 it can be seen that tests on the IfN/ENIT-database, using the lexicon of 937 Tunisian city/village names reached today 90% recognition rate.

The application field investigated in recently published papers Oliveira et al.(2006), Bertolami & Bunke(2007) is character/digit recognition. Compared to Latin text where a lot of research work is done, the number of work for the combination of Arabic text recognition systems is quite limited. One of the first works in this field was given by Snoussi Maddouri et al. in Snoussi Maddouri et al.(2002) presenting a combination scheme on feature level being tested on a sub set of pieces of Arabic words and words selected from a local bank checks database. In Essoukri Ben Amara & Bouslama(2003) Essoukri Ben Amara and Bouslema give an overview of multiple sources of information for recognition techniques and analyze different problems related to Arabic script recognition with hybrid architectures. In, Alma'adeed et al.(2004) an approach to combine Hidden Markov Models for Arabic handwritten word recognition is presented, and recently Farah et al.(2004) introduce a system based on the combination of different Neural Networks for the recognition of Arabic literal amount with a recognition rate of about 94% on a small test database containing 4800 words. In El Abed & Märgner(2009) methods to improve recognition rates based on different combination algorithms are presented. Comparing the presented results with the best performing single system at ICDAR2007 competition, an increase of recognition rate without reject from 83:34% to 94:71% is achieved.

Author(s)	Description	Data	Results
Abuhaiba et al.(1994)	Fuzzy Models (FCCGM)	1410 letters	99.4%
Amin et al.(1996)	NN	3000 characters	92%
Alimi(1997)	Neuro-fuzzy	100 words	89%
Dehghani et al.(2001)	Multiple HMM	Farsi-Cities	71.82%
Maddouri et al.(2002)	TD-NN	70 words 2070 images	97%
Khorsheed(2003)	Universal HMM	ancient documents	87%
Alma'adeed et al.(2004)	Multiple HMM's	AHDB	45%
Haraty & Ghaddar(2004)	NN	2132 letters	73%
Souici-Meslati & Sellami(2004)	NN	55 words	92%
Farah et al.(2004)	ANNK-NN, fuzzy K-NN	48 words (100 writers)	96%
Safabakhsh & Adibi(2005)	CD-VD-HMM	50 words	91%
Systems using the IfN/ENIT-database (* participant at ICDAR'05 competition, ** participant at ICDAR'07 competition))			
Pechwitz & Märgner(2003) (ARAB-IfN)*	SC-1D-HMM	training: a-c, test: d	2003: 89%
		training: a-d, test: e	2005: 74.69%
Jin et al.(2005) (TH-OCR)*	statistical methods	training: a-d, test: e	29.62%
Touj et al.(2005) (REAM)*	Planar HMMs	training: a-d, test: e	
Kundu et al.(2007) (MITRE)**	VD-HMM	training: a-e, test: f	61.70%
Ball (2007) (CEDAR)**	HMM	training: a-e, test: f	59.01%
Pal et al.(2006) (MIE)**		training: a-e, test: f	83.34%
Schambach(2003) (SIEMENS)**	HMM	training: a-e, test: f	87.22%
Al-Hajj et al.(2006) (UOB-ENST)*,**	HMM	training: a-d, test: e	2005: 75.93%
		training: a-e, test: f	2007: 81.93%
Abdulkadr(2006) (ICRA)*,**	NN (Two-Tier approach)	training: a-d, test: e	2005: 65.74%
		training: a-e, test: f	2007: 81.47%
Menasri et al.(2007) (Paris V)**	hybrid HMM/NN	training: a-e, test: f	80.18%
Benouareth et al.(2008)	HMM	training: a-c, test: d	89.08%
Zavorin et al.(2008) (CACI)**	HMM	training: abce, test: d	52%
Dreuw et al.(2008)	HMM	training: a-d, test: e	80.95%
Graves & Schmidhuber (2008)	MDR-NN	training: a-e, test: f	91.43%
Kessentini et al.(2008)	HMM multi-stream	lexicon of 500 words	86.2%

Table 1: Overview of Arabic text recognition systems

Arabic Handwriting Databases

Datasets of typical printed or handwritten words yield the most important part in a recognition system design. Papers on Arabic printed or handwritten word recognition have been published for many years. The first paper on Arabic OCR dates to 1975, the first Arabic OCR system was made available in the 1990s. However, only three papers comparing OCR systems have been published to dates, and the newest one is ten years old Kanungo et al.(1998). To overcome the problem of the lack of large datasets for developing Arabic OCR systems, and to motivate the research on statistical methods, a system for synthetic generation of Arabic datasets was developed according to the approach of English OCR Märgner & Pechwitz(2001). This allows for fast and simple generation of large datasets.

The situation is no better for Arabic handwritten word recognition. For many years the work published about Arabic handwriting recognition has used small, private datasets, which makes a comparison of methods essentially impossible. Another disadvantage lies in the fact these private datasets are often too small for reasonable statistical methods. Only a few datasets have been published. Recently, an overview about the state of the art of Arabic offline handwriting recognition was given Lorigo & Govindaraju(2006).

In the following, datasets of Arabic words or text are discussed. It must be mentioned that databases presented in papers and used for experiments are often not publicly available.

IfN/ENIT-Database

At the CIFED Conference in 2002, the Institute for Communications Technology (IfN) at Technical University Braunschweig, Germany, and the Ecole Nationale d'Ingénieurs de Tunis (ENIT), Tunisia, presented a database with handwritten Tunisian town names Pechwitz et al.(2002). This dataset was collected on specially designed forms to make the labeling procedure as simple as possible.

The aim was to collect images of handwritten town names written in a similar quality to town names in an address on a letter. The form was designed to:

- encourage writing without strong constraints,
- collect writing similar to writing on a letter,
- be easy to process automatically, and
- provide additional information about the person who completed it.

The form consists of three columns and a text block at the bottom. A column on the right hand side of the form lists 12 lines with printed Tunisian town/village names and their respective postcodes, which are automatically selected from the possible 937 names. The sample writers were expected to write the postcode in the left column and the town/village name in the middle column in their individual writing style. Writers had neither a writing line nor a box in which to write, because the processing of the scanned data should be as simple as possible. To provide a light writing guidance, black rectangles were printed on the backside of each page, which mark where to write. In the scanning process, these rectangles can be removed using a simple threshold. Further segmentation operations are unnecessary. The names printed on each form were selected randomly with the condition that each character shape should occur at minimum of 200 times. Therefore, those names with rare character shapes occur more often than names with frequent ones.
Each word appears at least three times in the database. A page number provides a form identifier for the subsequent processing. The block at the bottom gives additional information about the age, profession, and identity of the writer. Each writer was asked to complete five forms, with each writing 60 names.

The database (www.ifnenit.com), in version version 2.0 with patch level 1e (v2.0p1e), consists of 32492 handwritten Arabic names by more than 1000 writers. 937 Tunisian town/ village names are written. Each writer filled some forms with pre-selected town/village names (referred to as "names" in the following) and the corresponding post code. Ground truth was added automatically to the image data and verified manually. Table 2 shows a dataset entry of the IfN/ENIT-database.

This database has an interesting feature in the detailed labeling of the postcode as a label on word level. The character shape sequence, where each character shape depends on its position in the word, receives different labels for each position. Additionally, the position of the baseline is given as straight line. All this information, together with some additional quality measures, has been verified manually several times. The label information allows to train and test recognition systems, either on word or on character level. The given baseline position can be used for the testing of baseline estimation algorithms and the dependency on the baseline accuracy.
Table 2 shows some of the most important statistics of the different data sets of the IfN/ENIT database. Only the sets a-e are available, the sets f and s are used for the ICDAR'07 competition only.

set	names	characters	PAWs
a	6537	51984	28298
b	6710	53862	29220
c	6477	52155	28391
d	6735	54166	29511
e	6033	45169	22640
Total(v2.0p1e)	32492	257336	138060
f	8671	64781	32918
s	1573	11922	6109

Table 2: Statistic of IfN/ENIT-database v2.0p1e and the test datasets f and s.

Different databases for Arabic word and text recognition related topics are published during the last 10 years. The most of these databases or datasets are used for a limited period only. An overview of available databases for Arabic word and text recognition is presented 2008 in El Abed & Märgner(2008) .

Evaluation and Competitions of Arabic Text Recognition Systems

Another crucial aspect for recognition system development is the discussion and competition of different approaches. Only the testing of different methods on identical datasets allows for informative comparison. Furthermore, objective quality measuring methods are necessary for ranking the systems. This also constitutes a high motivation for developing advanced methods.

Evaluations Tools

Testing recognition systems with large identical datasets is crucial for performance evaluation. Another challenge comes from their complexity, because they consist of many specialized parts solving very diverse tasks. The recognition rate is a convenient measure for comparing different systems, but it is a global parameter hardly significant for system component development. To improve the overall system quality, it is essential to know the effectiveness of its modules.

The development of meaningful aspects of system evaluation methods was an important part of the aforementioned annual OCR tests at ISRI. The goal of these tests not only publicized the state-of-the-art of page reading systems, but also provided information for improvement through competition and objective assessment. While much has been achieved concerning the evaluation problem (e.g. Rice(1996)), the availability of tools and data remains an issue for research, as discussed in the paper Märgner et al.(2005), published in 2005. For example, it is not enough to measure the quality, based on the symbol output of the recognizer, only by considering the word accuracy. The quality of zoning and the segmentation into words or characters represent an important feature of a recognition system, and should be evaluated too Thulke et al.(1999). A more general concept for evaluating system modules separately is presented in Märgner et al.(1997).

ICDAR 2005

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 Märgner et al.(2005). To reach an optimal result, the competition was organized in a closed mode, i.e. the participants developed their systems using the IfN/ENIT-database for training and sent it to the IfN where the tests were carried out using new test data, unknown to the participants. Five groups submitted systems to this competition (Table 3).

System name	ID	Top 1	Top 5	Top 10
ICRA	1	65.74	83.95	87.75
SHOCRAN	2	35.70	51.62	51.62
TH-OCR	3	29.62	43.96	50.14
UOB	4	75.93	87.99	90.88
REAM*	5	15.36	18.52	19.86
ARAB-IFN	6	74.69	87.07	89.77

Table 3: Recognition results in % with the new dataset e

ICDAR 2007

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 International Conference on Document Analysis and Recognition (ICDAR) 2007 Märgner & El Abed(2007). The main results are shown in Table 4. This competition compared 14 systems submitted from 9 groups (some groups delivered more than one system). The table shows in the first column the results with datasets which are part of the training data (set d and e) and with the new data from set f. We show here only the recognition rate. A comparison with the 2005 tests shows an improvement of about 5% and more for many systems. But of course there are still errors of about 20% of the best systems.

ID	set d	set e	set f	set s
1	66.34	64.89	61.70	49.91
2	40.45	37.73	11.95	8.01
3	70.62	68.62	15.79	14.24
4	48.68	44.04	14.28	10.68
5	68.07	57.37	59.01	41.32
6	93.63	86.67	83.34	68.40
7	91.23	84.27	82.77	68.09
8	94.58	87.77	87.22	73.94
9	90.02	81.80	79.10	64.97
10	92.12	83.52	81.65	69.61
11	92.38	83.92	81.93	69.93
12	93.32	85.13	81.81	70.57
13	88.33	83.87	81.47	72.22
14	89.80	80.24	80.18	64.38

Table 4: Recognition results in % of correct recognized images on reference datasets d and e, new datasets f and s.

ICDAR2009

This year the third competition for Arabic handwritten word recognition using IfN/ENIT-database is planned and it will be again a step forward to better recognition systems. In this competition 8 groups are participating with a total of 20 systems.

Conclusions

Considering all the aspects discussed in the previous sections, the next steps to provide better Arabic handwriting recognition systems are obvious. In the following, the necessary steps are listed:

- Selection of interesting fields of application for Arabic handwriting recognition.
- Collection of real world data, perform scanning and labeling of the data to construct a database.
- If real world data are not available: development of a concept to generate an artificial database by selecting people to fill forms. Scanning and labelling may be easier, as the form can be specially designed.
- Organization of a workshop to discuss newest research results and to present a performance evaluation of different systems or modules based on a common dataset, performed by an independent group.

- Formation of new datasets available to research teams immediately. The labels should have the same format for different datasets.
- Development of performance measurements of processing and recognition modules should be considered in the workshops.
- Workshops should be conducted annually, and interdisciplinary contributions encouraged.
- Exchange of methods between handwriting recognition of different languages.

From experience gathered on other languages, these steps should help us reach the goal of developing better Arabic handwriting reading machines. In addition to this short-term objective, we hope to approach a better understanding of the nature of the reading process of humans in general, to reduce the work needed to adapt a system to a new application.

References

Abdulkadr, A. (2006). Two-tier approach for Arabic Offline handwriting recognition. In 10th International Workshop on Frontiers in Handwriting Recognition (IWFHR), pp. 161--166.

Abuhaiba, I., S. Mahmoud, and R. Green (1994). Recognition of handwritten cursive Arabic characters. IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (6), 664--672.

Ahmed, P. and Y. Al-Ohali (2000). Arabic character recognition: Progress and challenges. Journal of King Saud University. Computer & Information Sciences 12 (1), 85--116.

Al-Hajj, R. (2007). , Reconnaissance hors ligne de mots manuscrits cursifs par l'utilisation de systémes hybrides et de techniques d'apprentissage automatique,. Ph. D. thesis, l'Ecole Nationale Supérieure des Télécommunications (ENST).

Al-Hajj, R., C. Mokbel, and L. Likforman-Sulem (2006). Reconnaissance de l'écriture arabe cursive: Combinaison de classifieurs MMCs à fenêtres orientées. In Colloque International Francophone sur l'Ecrit et le Document (CIFED), pp. 271--276.

Alimi, A. M. (1997). An evolutionary neuro-fuzzy approach to recognize on-line Arabic handwriting. In 4th International Conference on Document Analysis and Recognition, pp. 382--386.

Alma'adeed, S., C. Higgins, and D. Elliman (2004). O_-line recognition of handwritten Arabic words using multiple hidden Markov models. In the Twenty-third SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Volume 17, pp. 75--79.

Amin, A., H. Al-Sadoun, and S. Fischer (1996). Hand-printed Arabic character recognition system using an artificial network. Pattern Recognition 29 (4), 663--675.

Atukorale, A. and P. Suganthan (1999). Combining classifiers based on confidence values. In 5th International Conference on Document Analysis and Recognition (ICDAR), pp. 37--40.

Ball, G. R. (2007). Arabic Handwriting Recognition using Machine Lerning Approaches. Ph.D. thesis, The Faculty of Graduate School of State University of New York at Buffalo.

Benouareth, A., A. Ennaji, and M. Sellami (2008). Arabic handwritten word recognition using HMMs with explicit state duration. Journal on Advances in Signal Processing.

Bertolami, R. and H. Bunke (2007). Multiple classifier methods for offine handwritten text line recognition. In 7th International Workshop on Multiple Classifier Systems, pp. 72--81.

Cordella, L. P., P. Foggia, C. Sansone, F. Tortorella, and M. Vento (1999). Reliability parameters to improve combination strategies in multi-expert systems. Pattern Analysis & Applications 2 (3), 205--214.

Dehghani, A., F. Shabani, and P. Nava (2001). Off-line recognition of isolated Persian handwritten characters using multiple hidden Markov models. In International Conference on Information Technology: Coding and Computing (ITCC).

Dreuw, P., S. Jonas, and H. Ney (2008). White-space models for offline Arabic handwriting recognition. In 19th International Conference on Pattern Recognition (ICPR).

El Abed, H. and V. Märgner (2007). Comparison of different pre-processing and feature extraction methods for offline recognition of handwritten Arabic words. In 9th International Conference on Document Analysis and Recognition (ICDAR), Volume 2, pp. 974--978.

El Abed, H. and V. Märgner (2008). Base de Donnes et Comptitions – Outils de Dveloppement et d_Evaluation de Systmes de Reconnaissance de Mots Manuscrits Arabes. In Colloque International Francophone sur l'Ecrit et le Document (CIFED), pp. 103--108.

El Abed, H. and V. Märgner (2009). Improvement of Arabic handwriting recognition systems - combination and/or reject? In Document Recognition and Retrieval XVI, IS&T/SPIE International Symposium on Electronic Imaging, Number 7247--10.

Essoukri Ben Amara, N. and F. Bouslama (2003). Classification of Arabic script using multiple sources of information: State of the art and perspectives. International Journal on Document Analysis and Recognition 5 (4), 195--212.

Farah, N., L. Souici, L. Farah, and M. Sellami (2004). Arabic words recognition with classifiers combination: An application to literal amounts. In Artificial Intelligence: Methodology, Systems, and Applications, pp. 420--429.

Farah, N., L. Souici, and M. Sellami (2006). Classifiers combination and syntax analysis for Arabic literal amount recognition. Engineering Applications of Artificial Intelligence 19 (1), 29--39.

Graves, A. and J. Schmidhuber (2008). Offline handwriting recognition with multidimensional recurrent neural networks. In 22sd Conference on Neural Information Processing Systems (NIPS).

Günter, S. (2004). , Multiple classifier systems in offline cursive handwriting recognition,. Ph.D. thesis, Institute of Computer Science and Applied Mathematics (IAM).

Haraty, R. and C. Ghaddar (2004). Arabic text recognition. International Arab Journal of Information Technology (IAJIT) 1, 156--163. [Ho et al.(1994)] Ho, T. K., J. Hull, and S. Srihari (1994). Decision combination in multiple classifier systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (1), 66--75.

Jin, J., H. Wang, X. Ding, and L. Peng (2005). Printed Arabic document recognition system. In Proc. SPIE-IS&T Electronic Imaging, Volume 5676, pp. 48--55.

Kanungo, T., G. A. Marton, and O. Bulbul (1998). Performance evaluation of two Arabic OCR products. In SPIE (Ed.), Proceedings of AIPR Workshop on Advances in Computer Assisted Recognition, Volume 3584.

Kessentini, Y., T. Paquet, and A. M. Benhamadou (2008). Multi-script handwriting recognition with n-streams low level features. In Proc. 19th International Conference on Pattern Recognition (ICPR), pp. 1--4.

Khorsheed, M. S. (2003). Recognising handwritten Arabic manuscripts using a single hidden Markov model. Pattern Recognition Letters 24 (14), 2235--2242.

Kittler, J. (2000). A framework for classifier fusion: Is it still needed? In Advances in Pattern Recognition, pp. 45--56.

Kundu, A., T. Hines, J. Phillips, B. Huyck, and L. Van Guilder (2007). Arabic handwriting recognition using variable duration HMM. In 9th International Conference on Document Analysis and Recognition (ICDAR), Volume 2, pp. 644 --648.

Lin, X., X. Ding, M. Chen, R. Zhang, and Y. Wu (1998). Adaptive confidence transform based classifier combination for Chinese character recognition. Pattern Recognition Letters 19 (10), 975--988.

Lorigo, L. and V. Govindaraju (2006). Offline Arabic handwriting recognition: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (5), 712--724.

Maddouri, S. S., A. Belaid, C. Choisy, and H. Amiri (2002). Modéle perceptif neuronal á vision globale-locale pour la reconnaissance de mots manuscrits arabes. In Colloque International Francophone sur l'Ecrit et le Document (CIFED), pp. 11--20.

Märgner, V. and H. El Abed (, 2008). , Arabic and Chinese Handwriting Recognition, , Volume 4768, Chapter Databases and Competitions: Strategies to Improve Arabic Recognition Systems. Springer, LNCS.

Märgner, V. and H. El Abed (2007). ICDAR 2007 Arabic handwriting recognition competition. In 9th International Conference on Document Analysis and Recognition (ICDAR), Volume 2, pp. 1274--1278.

Märgner, V., P. Karcher, and A.-K. Pawlowski (1997). On benchmarking of document analysis systems. In 4th International Conference on Document Analysis and Recognition (ICDAR), Volume 1, pp. 331--336.

Märgner, V. and M. Pechwitz (2001). Synthetic data for Arabic OCR system development. In 6th International Conference on Document Analysis and Recognition (ICDAR), pp. 1159--1163.

Märgner, V., M. Pechwitz, and H. El Abed (2005). ICDAR 2005 Arabic handwriting recognition competition. In 8th International Conference on Document Analysis and Recognition (ICDAR), Volume 1, pp. 70--74.

Menasri, F., N. Vincent, E. Augustin, and M. Cheriet (2007). Shape-based alphabet for off-line Arabic handwriting recognition. In 9th International Conference on Document Analysis and Recognition (ICDAR), Volume 2, pp. 969--973.

Oliveira, L. S., M. Morita, and R. Sabourin (2006). Feature selection for ensembles applied to handwriting recognition. International Journal on Document Analysis and Recognition 8 (4), 262--279.

Pal, U., K. Roy, and F. Kimura (2006). A lexicon driven method for unconstrained Bangla handwritten word recognition. In 10th International Workshop on Frontiers in Handwriting Recognition (IWFHR), pp. 601--606.

Pechwitz, M., S. S. Maddouri, V. Märgner, N. Ellouze, and H. Amiri (2002). IFN/ENITdatabase of handwritten Arabic words. In Colloque International Francophone sur l'Ecrit et le Document (CIFED), pp. 127--136.

Pechwitz, M. and V. Märgner (2003). HMM based approach for handwritten Arabic word recognition using the IFN/ENIT-database. In 7th International Conference on Document Analysis and Recognition (ICDAR), pp. 890--894.

Rice, S. V. (1996). Measuring the Accuracy of Page-Reading Systems. Ph. D. thesis, Department of Computer Science, University of Nevada, Las Vegas.

Safabakhsh, R. and P. Adibi (2005). Nastaaligh handwritten word recognition using a continuous-density variable-duration hmm. Arabian Journal for Science & Engineering 30 (1B), 95-- 118.

Schambach, M.-P. (2003). Model length adaptation of an HMM based cursive word recognition system. In 7th International Conference on Document Analysis and Recognition (ICDAR), Volume 1, pp. 109-- 113.

Snoussi Maddouri, S., H. Amiri, A. Belaid, and C. Choisy (2002). Combination of local and global vision modelling for Arabic handwritten words recognition. In 8th International Workshop on Frontiers in Handwriting Recognition (IWFHR), pp. 128--135.

Souici-Meslati, L. and M. Sellami (2004). A hybrid approach for Arabic literal amounts recognition. Arabian Journal for Science and Engineering (AJSE) 29 (2B), 174--194.

Suen, C. Y. and J. Tan (2005). Analysis of errors of handwritten digits made by a multitude of classiérs. Pattern Recognition Letters 26 (3), 369--379.

Thulke, M., V. Märgner, and A. Dengel (1999). A general approach to quality evaluation of document segmentation results. In 3ird IAPR International Workshop on Document Analysis Systems (DAS), Volume 1655, pp. 43--57.

Touj, S., N. Ben Amara, and H. Amiri (2005). Arabic handwritten words recognition based on a planar hidden markov model. International Arab Journal of Information Technology 2 (4).

Van Erp, M. and L. Schomaker (2000). Variants of the borda count method for combining ranked classifier hypotheses. In 7th International Workshop on Frontiers in Handwriting Recognition (IWFHR), pp. 443--452.

Xu, L., A. Krzyzak, and C. Suen (1992). Methods of combining multiple classiérs and their applications to handwriting recognition. IEEE Transactions on Systems, Man, and Cybernetics 22 (3), 418-- 435.

Zavorin, I., E. Borovikov, E. Davis, A. Borovikov, and K. Summers (2008). Combining different classification approaches to improve o_-line Arabic handwritten word recognition. In Proceedings of SPIE - The International Society for Optical Engineering, Volume 6815.

Zouari, H. K. (2004). , Contribution _a l'évaluation des méthodes de combinaison parallèle de classifieurs par simulation,. Ph.D. thesis, Université de Rouen.