Multi-level Arabic Handwritten Words Recognition

H. Miled[♦]*, M. Cheriet* and C. Olivier[♦♣]

[♦] PSI-La3i, UFR des Sciences, Université de Rouen, France
* LIVIA, École de Technologie Supérieure, Montréal, Canada
[♣] SIC-IRCOM, UMR CNRS 6615 Université de Poitiers, France
Tel : (33) 02 35 14 67 86   Fax : (33) 02 35 14 66 18
E-mail: miled@la3i.univ-rouen.fr

ABSTRACT

In this paper, we present a strategy of Arabic words recognition by combining two levels which are based on global and analytical approaches according to the topological properties of Arabic handwriting. In the first level (global), we consider the visual indices which can be generated by: diacritics and strokes (denoted tracing) that form the main shapes of the word. Each word is described as a sequence of visual indices which is treated by a "global" classifier based on Hidden Markov Model (HMM). In the second level, the word is segmented into graphemes, then each grapheme is transformed into a HMM observation by a vector quantization phase. An analytical HMM is developed in order to manage the observation sequences. At this level the diacritics are not taken in consideration which allows to reduce the number of estimated character models. Finally we combine the two approaches to decide on the class of an unknown word. In fact, the global model serves as a filter. It produces a set of hypotheses to the analytical model, which in turns, defines and outputs the final decision.

Key words: Arabic handwriting recognition, character and cursive scripts recognition, visual indices, HMM modeling, classifier combination, hybrid approach.

1. Introduction

Since the appearance of the machine, man has been trying to mimic his own *behaviour*. The need to understand how he functions, pushed him to model mechanically not only his motion, but also the way he thinks. This curiosity reached the field of pattern recognition, and in particular automatic reading, and gave birth to hundreds of research studies [1]. Our main interest is the optical reading of Arabic cursive handwriting. Arabic is the official script of Arabic and Persian countries which places it in a fairly high position on the ranking of the world mostly used scripts. Still, the number of studies done on the field of Arabic writing recognition is relatively low [2-3]. Because the Arabic writing is based on an alphabet and rules different from those of Latin, it makes it a foreign writing for the majority of our scientific community.

Handwriting recognition is based mainly on two types of approaches: global and analytical. However, each approach has its advantages and its disadvantages. In the literature, we find a thorough research works treating combination of classifiers in case of character recognition [4-5]. However research works treating the strategy of

combining classifiers for word recognition are less frequent [6-7]. In our application, we find in the combination of different classifiers a solution to extract complementary information. Consequently this combination might improve the performances of the developed handwritten word recognition system in the case of a complex problem with a large number of word classes (232).
This paper is divided into two main parts. we introduce in the first part the different visual indices used for the word coding and the global modeling of handwriting. The second part deals with the analytical approach which is the core of our final system. For both modeling levels (global and analytical), we adopted Hidden Markov Models (HMMs). We terminate this paper by discussing results of the decision level (strategy of classifiers combination) and we give a general conclusion.

2. Visual Aspect of the Arabic Handwriting: The Global Approach

This approach has as goal to describe the word (as a global entity) by a sequence of visual indices. So, first we detect the information zones in the word. This phase is achieved by the extraction of the image external contours. These components represent two types of image information: tracings (strokes, see Fig. 1-b) and diacritics [8], (Fig. 1-c).

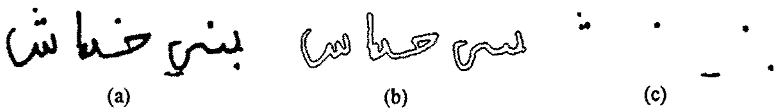


Fig. 1. Detection and separation of information zones of the image: (a) word image; (b) external contours of tracings; (c) diacritics (dots).

2.1 Visual indices extraction

We define the set of visual indices extracted from the tracing zone and the diacritics.

2.1.1 Zones of tracings

It contains the majority of the image information (word). The visual indices from this zone are of two types, defined as regularities and singularities [9]. The first type groups indices extracted from the middle zone: (Fig. 2.) loops, valleys and inter-tracing spacing (noted by "#"); the second type includes the prominent features: alifs, ascenders, descendents and tanks [10].

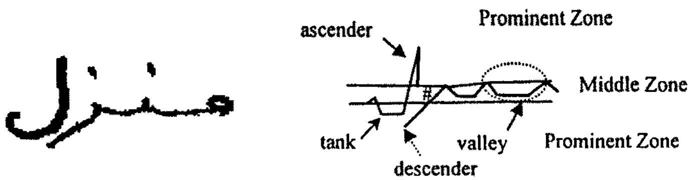


Fig. 2. Different Visual Indices extracted from the tracing zone.

2.1.2 Diacritics

Diacritics are frequently used in Arabic script. The most pertinent ones are dots (single or multiple) because distinguish characters having the same main body. Multiple dots come in two types: double (Fig. 3-b) and triple (Fig. 3-c). In the case of handwriting, the diacritics are complex that we decompose each of these two types into their number of single dots. To refine the word description and to increase our information about the word, we separate detected diacritics (dots) into two distinct visual indices according to their relative position to the baseline.

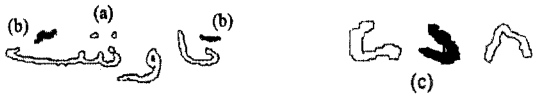
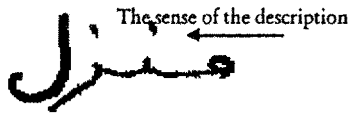


Fig. 3. Detected diacritics: (a) simple dots (b) double dots (c) triple dot shapes.

2.1.3 Global Markovian Modeling

At this level, the visual indices described above represent the set of features used in word description. The words are represented by a chronological observation sequence y_1^T (visual indices, see Fig. 4). The description direction follows the classical Arabic reading/writing direction, which is from right to left. The management of these observation sequence is based on Hidden Markov Models (HMMs).

Ta/As/#/Ds/Ud/V/Ud/V/L ←



$$y_1^T = (y_1, y_2, \dots, y_{T-1}, y_T) \text{ with } T = 9 \text{ in this particular case}$$

Fig. 4. Examples of word presentation as a sequence of visual indices: *As*: ascender; *Ds*: descender; *Ud:* upper-dot ; *L* : loop ; *V* : valley ; *Ta* : tank ; #: inter-trace spacing.

HMMs are soft elastic models widely used in speech recognition [11], and also in writing recognition [12-14]. This modeling tolerates variability in the writing, and adapts perfectly to our type of data. The only problem is that we have a relatively large lexicon (232 classes of words), and few samples per word classes.

In this paper, the "global" HMM of the word class ω_i is noted $\lambda_i^G = (A_i, B_i, \Pi_i)$. The used models have classical right to left topology and they are trained by *Baum-Welch* algorithm. At this level, our goal is to build a pre-classification module.

The classifier appropriated is Maximum Likelihood (ML) module based on this HMM global word modeling. This classifier will be called global classifier. It computes for each model λ_i^G of word class ω_i the associated probability $P(y_1^T / \lambda_i^G)$.

Table 1: Performance of the global classifier.

	top1	top2	top3	top4	top5	top10	top20	top50	top80	top150
τ_{reco}	58,9%	68,3%	72,5%	76%	78,2%	86,8%	93,3%	98%	99,1	99,9%

Table 1 presents the performance of the global classifier trained on 4720 words and tested on 5900 other words. We point out the weakness of the recognition rate (τ_{reco}) in the first choice (less than 60% in top 1). However these recognition rates is equal to 98% in the 50th choice (top 50) and greater than 99% in the top 80. We note that from the top 80 to the top 150 the increase of recognition rates is less than 1%.
This classifier is used as a filter to reduce the lexicon size. It selects a set of the most probable word classes noted by Ω_G ($\Omega_G \subset \Omega_{word}$, with Ω_{word} the set of all word classes):

$$\forall \omega_l \in \Omega_G \; ; \; \omega_j \in \Omega_{word} - \{\omega_l\} \; if \; P(y_1^T / \lambda_j^G) \geq P(y_1^T / \lambda_l^G) \Rightarrow \omega_j \in \Omega_G$$

Then, the second level consider only the reduced lexicon. In order to optimize the combination results, we will study (in §4) the influence of the size (noted by N_G) of Ω_G on the recognition system performances.

3. The Analytical Approach

In cursive Arabic handwriting, a tracing is a set of characters or portions of a character (graphemes) attached with links. In order to describe a word as a sequence of features, we developed a segmentation module, described in [15], to cut tracing into graphemes. To stabilize the grapheme's morphologic representation, we filter out the links. In this way, the grapheme is restricted to the component representing its main body (Fig. 5).

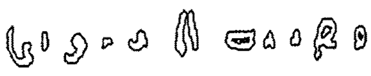


Fig. 5. Filtering of links and detection of grapheme main bodies:

3.1 Observations vs. Graphemes

The transfer of a grapheme (a shape) into an Markov Models observation (logical entity) is primordial phase for HMM word modeling. At this level, each grapheme is represented by a vector of the different measurements used by the classifier to compare an unknown grapheme to known ones. The decision of the observation given to each grapheme is made using a k nearest neighbors classifier (k-NN) [16].

3.2 Analytical Markovian Modeling

The main subject of modeling is the character component. Let us remember that the word is a sequence of pseudo-words. We are taking into account in our model to get a specific Arabic handwriting.

The modeling process is directly linked to the segmentation phase. This segmentation does not depend on the character as a logical entity but rather depends on its shape. Based on this, we go from a model per character to a model per family of characters (which have the same main body); which reduces the number of models from approximately 30 to only 18. The character model is a right-left 3 states model (Fig. 6-

a), and has to take into consideration all the possible segments of a character. The states of this model are called μ-states. The transition probabilities between the different states are estimated on the entire database.

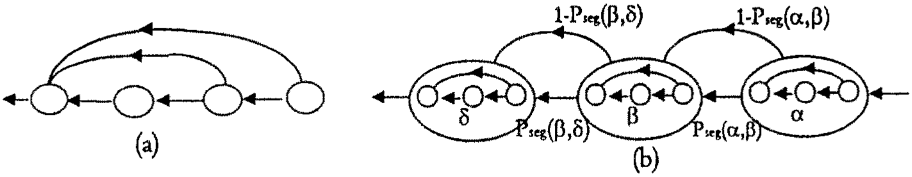


Fig. 6. (a) character model (b) Example of model for a 3-character pseudo-word (αβδ).

A pseudo-word is a sequence of characters. The pseudo-word model can be viewed as a concatenation of the character models composing the word (Fig. 6-b). The segmentation phase depends essentially on the set of bi-grams. The segmentation engine behaves in the same way each time it encounters the same bi-gram in a pseudo-word. The parameter of interest at this level is the probability of segmentation by bi-grams. This probabilities can be computed over the entire database.

The final model is a fusion of the models defined earlier. Its states are all the combinations of the μ-states (from character models), encountered while tagging the training set.

The classifier associated, for this second level of our system, is also a ML based on these HMMs analytical word modeling. It is called analytical classifier, As previous, this classifier computes, for each candidate model λ_i^A of the class ω_i (with $\omega_i \in \Omega_G$), the associated probability $P(y_1^T / \lambda_i^A)$.

Table 2 presents the performance of this classifier tested on 5900 words. The models are trained on a tagged database of 4720 words.

Table 2: Performances of the analytical level classifier.

	top 1	top 2	top 3	top 4	top 5	top 10
τ_{reco}	81,6%	88,0%	90,4%	92,2%	93,0%	94.9%

The analysis of the main recognition errors (confusions) shows that the principal cause of confusions are word classes which have the same handwritten main shapes and which can be distinguished by diacritics (Fig. 7).

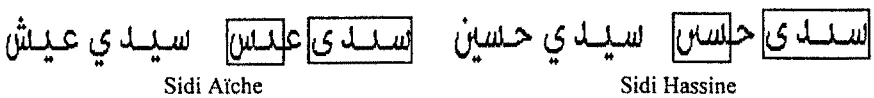


Fig. 7. Similar main shapes of the different word classes.

4. Decision Level and Results

The strategy of recognition of the unknown word is composed by two levels (see Fig. 8). Each level is a maximum likelihood (ML) classifier. Which takes the unknown pattern (noted by x in the Fig. 8) as being a sequence of observation $y_1^T = (y_1, y_2, \ldots, y_T)$ presenting visual indices in one level and graphemes in the other level.
We remember that the first level is based on the global approach which select a set Ω_G of its outputs from the most probable word classes (models).

$$S_G = (s_G^1, s_G^2, \ldots, s_G^{N_\omega});$$

$$s_G^i = \begin{cases} P_G(y_1^T / \lambda_i^G) & \text{if } \omega_i \in \Omega_G \\ 0 & \text{otherwise} \end{cases}$$

with N_ω *number of elements of the word set* Ω_{word}

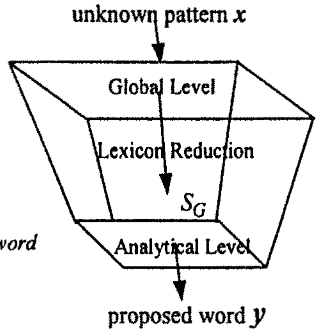


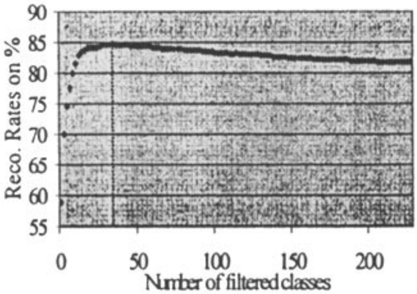
Fig. 8. Combination of the two level classifiers: **x** is the unknown word and **y** is the decision

Only the word classes ω_i belonging to the set Ω_G $(s_G^i \neq 0)$ are treated by the second level. Each modeling level, of this system, exploits an appropriate information about the word being processed (different signal representation). This information complementarily increases the performance of the final recognition module. We use two methods of combination:
In the first method, we consider that the two levels of the system are independent. The first level filters the word classes to give to the second level a set of candidates without any other information. The word is attached to the class ω_k $(y = \omega_k)$ for which the model λ_k^A maximizes the emission probability of y_1^T by the second level (the analytical classifier):

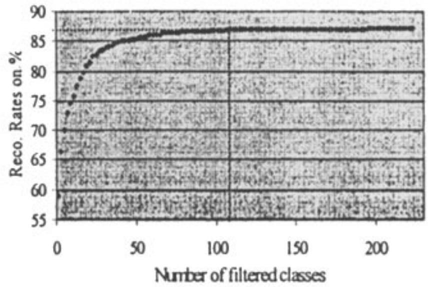
$$\lambda_k^A = \arg\max_{\lambda_i^A} P(y_1^T / \lambda_i^A) \text{ with } \lambda_i^A \text{ model of the class } \omega_i \in \Omega_G \quad \text{(eq. 1)}$$

The second method is based on the stochastically independence assumption of the classifiers. The global classifier gives a set of candidates with their a posteriori probabilities. A score S_i^{comb} is given to each candidate $\omega_i \in \Omega_G$ and which is equal to $P(y_1^T / \lambda_i^G, \lambda_i^A)$. Using the classifiers independence assumption, we can simplify the computation of the score of each candidate: $S_i^{comb} \approx P(y_1^T / \lambda_i^G).P(y_1^T / \lambda_i^A)$
Finally, the word is attached to the class ω_k $(y = \omega_k)$ for which maximizes the score after the combination of the two classifiers:

$$k = \arg\max_i S_i^{comb} \text{ with } S_i^{comb} \text{ the score of the class } \omega_i \in \Omega_G \text{ after the combination} \quad \text{(eq. 2)}$$



(a) : with the rule (eq. 1)



(b) : with the rule (eq. 2)

Fig. 9. The recognition rates (top 1) variation regarding to the values of N_G.

Our lexicon counts 232 different word classes $(N_\omega = 232)$. Let us remember that the models are trained on 4720 tagged words and the system performances are evaluated on a test set of 5900 words. The recognition strategy by combining the two classifiers improves the system performance in both cases (Fig 9.). The performances of the first combining method (eq. 1) are optimum (about 85% in top 1) for a size $N_G = 40$ of Ω_G, and they decrease for high values of N_G (Fig. 9-a). However, the performances of the second combining method are optimum (about 87% in top 1) for a value of N_G superior to 100 filtered candidates and they remain slightly constant for highest values of N_G, see Fig 9-b.

These results are encouraging to undertake further research and continue such paradigm by refining both classifiers and adding other characteristics that are intrinsic to Arabic handwriting.

5. Conclusion

This paper present two types of classifier based respectively on the global approach and the analytical one. These two classifiers treat about different representation of the word which assume a complementarily between them. The global modeling treats about the visual aspect of the Arabic handwritten. However, the analytic approach models the word as a fusion of different levels of abstraction: the character, the pseudo-word and the word. A strategy of combining the two classifier is developed to use a maximum of information about the unknown word. This strategy improves the good recognition rate about 3.5% in the first choice for one method and 5.5 % for the other method which is based on the classifier independence assumption. These results are encouraging to continue the development of the recognition system in same way.

6. References

[1] M. Fayol, J. E. Gombert, P. Lecocq, L. Sprenger-Charolles, D. Zager, "Psychologie Cognitive de la Lecture", *Presses Universitaires de France*, 1992.

[2] A. Amin, "Off Line Arabic Character Recognition- A survey", *In Proc. of ICDAR'97*, Ulm, Germany, pp 596-599, Aug. 1997.

[3] B. El-Badr, S. A. Mahmoud, "A Survey and Bibliographies of Arabic optical text recognition", *Signal Processing*, vol. 41, pp. 49-76, 1995.

[4] T. K. Ho, J. J. Hull, S. N. Srihari, "Decision Combination in Multiple Classifier Systems", *IEEE PAMI*, vol. 16, no. 1, Jan. 1994.

[5] L. Xu, A. Krzyzak, C. Y. Suen, "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 3, pp 418-435, 1992.

[6] R. K. Pawalka, N. Sherkat, R. J. Whitsow , "Recognizer characterisation for combining handwriting recognition results at word level", *In Proc. of ICDAR'95*, Montréal, Canada, pp 68-73, Aug. 1995.

[7] B. Plessis, A. Sicsu, L. Heutte, E. Menu, E. Lecolinet, O. Debon, J. V. Moreau, "A Multi-classifier combination strategy for recognition of handwritten cursive words", *In Proc. of ICDAR'93*, Tsukuba Science City, Japan, pp 642-645, Aug. 1997.

[8] A. Ameur, K. Romeo, H. Miled, M. Cheriet, "Approche Globale pour la Reconnaissance des Mots Manuscrits Arabes", *Proc. of CNED'94,* Rouen, France, pp 151-157, Juillet 1994.

[9] J. C. Simon, O. Baret, "Handwriting recognition as an application of regularities and singularities in line pictures", *Proc. of IWFHR*, Montréal, Canada, pp 23-36, 1990.

[10] M. Cheriet, H. Miled, C. Olivier: ''Visual Aspect of Cursive Arabic Handwriting Recognition'', *Visual Interface*, Vancouver , Canada, Aug. 1998.

[11] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. of IEEE*, vol. 77, no. 2, pp 257 285, 1989.

[12] H. Bunke, M. Roth and E. G. Schukat-Talamazzini, "Off-line cursive handwriting recognition using Hidden Markov Models", *Pattern Recognition*, vol. 28, no. 9, pp 1399-1413, 1995.

[13] N. Ben Amara, A. Belaid, "Printed PAW Recognition Based on Planar Hidden Markov Models", *Proc. of ICPR'96*, Vienna, Autria, vol. 2, pp 220-224, Aug. 1996.

[14] C. Olivier, T. Paquet, M. Avila , Y. Lecourtier, "Optimal Order of Markov Models applied to Bankchecks", *Inter. Journal of Pattern Recognition and Artificial Intelligence*, vol. 11, no. 5, pp 789-800, 1997.

[15] C. Olivier, H. Miled, K. Romeo, Y. Lecourtier, "Segmentation and Coding of Arabic Handwritten Words", *Proc. of ICPR'96*, Vienna, Austria, vol. 3, pp 264-268, Aug. 1996.

[16] H. Miled, C. Olivier, M. Cheriet, Y. Lecourtier, "Coupling Observation/Letter for a Markovian Modelisation Applied to the Recognition of the Arabic Handwriting", *Proc. of ICDAR'97, Ulm*, Germany, pp. 580-583, Aug. 1997.