Human Reading Based Strategies for off-line Arabic Word Recognition

Abdel Belaïd, Christophe Choisy

▶ To cite this version:

HAL Id: inria-00110984

https://hal.inria.fr/inria-00110984

Submitted on 2 Nov 2006

Human Reading Based Strategies for off-line Arabic Word Recognition

A. Belaïd[1] and Ch. Choisy[2]
[1]University Nancy 2-LORIA
Campus Scientifique, 615, rue du jardin botanique
54600 Villers-Lès-Nancy, France,
Email : abelaid@loria.fr

[2]ITESOFT AIMARGUES, Parc d'Andron - Immeuble Le
Séquoia, 30470 Aimargues – France
Email : Christophe.Choisy@itesoft.com

Abstract

This paper summarizes some techniques proposed for off-line Arabic word recognition. The point of view developed here concerns the human reading favoring an interactive mechanism between global memorization and local checking making easier the recognition of complex scripts as Arabic. According to this consideration, some specific papers are analyzed and their strategies commented.

1 Type of survey

Concerning Arabic recognition, several surveys have been proposed in the literature considering different points of view:

- By stressing the multiplication of the source of information, from simple classifier to a combination of them, with simple or hybrid choice of the primitives, as described by Essoukri Ben Amara and Bouslama [1].

- By considering the nature of the script: printed or handwritten, its recognition engines and its applications, like in Lorigo and Govindaraju [2].

- By pointing out the nature: symbolic or numeric of the methods as made by Amin [3].

We propose another survey based on the functioning of the human perception going from coarse to detail (i.e. local, analytical or precise). It makes it possible to better justify the choice of the observations, to order them in the classifier cascades, and to propose solutions of correction in the case of conflict or problem and gives finally a smell to all the chain of recognition.

2 Human Perception of Arabic Writing

Arabic is a calligraphic language privileging the global rendering of the whole word to the detail of the letter which is often thinned, crushed, sketched provided that it contributes to the embellishment of the unit (see Figure 1).

Figure 1: Same word written with different possible elongations as described in [1]

Thus the letter can take one to four different forms according to its position in the word. The global becomes the form to be recognized and the letter passes in the second plan favoring the total appearance (see Figure 2). The consequence is a bigger alphabet containing around 100 possible forms [1].

Figure 2: Examples of style fonts of Arabic as described in [4]

However, to facilitate this calligraphic reading, diacritics and accents come first to contribute to the deciphering of the letters which have very similar base shapes. Second, in order to not force too much the writer to maintain the pen lowered to calligraphy only one cursive shape, Arabic offers a decomposition in PAW (Part of Arabic Word) which introduces pauses in the writing having an influence on the recognition process. The PAWs simplify the script apprehension and make easier the linear recognition. Figure 3 gives an example of the Arabic writing complexity, with sub-words and diacritic information.
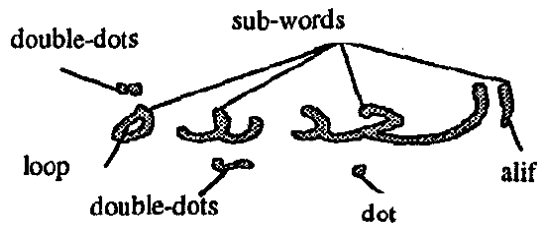
Figure 3: Arabic writing complexity: example of an handwritten word as shown in [28]

Considering the reading process itself and the perception of the writing, Arabic reading seems to be more global than syllabic. It is facilitated by chopping the word in PAWs which makes it finally semi-global.

Some psycho-cognitive experiments proved that the procedure of reading with human starts with a first global vision of the relevant characteristics. The basic experience of reading letter in and outside the word showed the "Word Superiority Effect". In order to illustrate this phenomena, Mc Clelland and Rumelhart proposed a reading model [5]. As illustrated in Figure 4, the model is based on three fundamental hypotheses: 1) the perception is operated in three different processing levels, each one of them is representing a different abstraction level, 2) the perception implies parallel processing on the visual information, 3) the related processes are interactive, i.e. bottom-up and top-down.
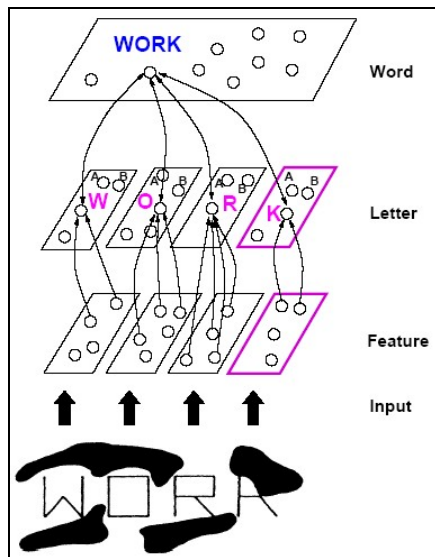

Figure 4: McClelland and Rumelhart Model

-   Human builds a complete image of his environment by accumulating different sources of sensory data. In these various stages of decision-making, he proceeds by a general study of the problem. If this global vision is not sufficient, he seeks to go into details [6].

-   In the case of Arabic writing, the natural "global" pattern is the PAW: words are combinations of PAWs. Furthermore, there are no clear physical limits between words: words are mainly recognized through the sense of the different PAW aggregation possibilities, where only one should give a sense for an entire sentence. PAWs could be compared to Latin syllables written separately, witch should be correctly gathered to have a meaning.

We can conclude from these standpoints that Arabic writing fits very well the reading principle of Mc Clelland and Rumelhart as it clearly privileges the "Word Superiority Effect", while adding some local perceptual information to help word understanding.

But the corresponding model has to be adapted to consider the PAW intermediate reading level and the letter distortions: PAWs introduce an intermediate global level of information, while letter shape variations make more complex their localization and their modeling.

3   Computing perception levels

All computing methodologies try to simulate a human perception level. Considering human perception of Arabic writing with the particularity of PAW, this leads to divide the classification methods in four classes:
-   Global-based vision classifiers.
-   Semi-global-based vision classifiers.
-   Local-based vision classifiers.
-   Hybrid-level classifiers.

3.1   Global-based Vision Classifiers

In this holistic approach, the word is regarded as a whole, allowing correlations on the totality of the pattern. This approach avoids a heavy task of letter localization and recognition and remains very used. However, its interest remains limited to small vocabularies or as a pre-classification step, because its complexity grows linearly with the number of word models.

This category is mainly assimilated to what is called segmentation-free approach. In fact it means that even if a segmentation is used, no local interpretation is made but information is gathered at the word level. In such an approach, one should find the best interpretation possible of a word based on an observation sequence derived from the word image without performing a meaningful segmentation first [7].

Several works on Arabic writing are almost directly derived from Latin studies. Thus the global approach leads to two questions:
-   the first one is natural: "it is possible to correctly adapt classical Latin script approaches to Arabic script ?"

-    the second one is: "it is possible to extract Arabic words as 'simply' as in Latin script ?"

Srihari et al. propose in [8] a handwritten Arabic word spotting system based on a feature vector similarity measure. The GSC (Gradient, Structural and Concavity) binary features previously used for Latin work in [9] give the best performances. The similarity measure is common to the two languages [9,41]. A precision of 70% is achieved at a recall of 50% when 8 writers were used for training.

The specificity of Arabic writing appears only in a particular part of the work: the word segmentation. Due to the Arabic writing nature, the authors cannot directly evaluate the gap between two consecutive PAWs to decide concerning the word limits (see Figure 5). They use then an NN on a set of 9 features. The authors indicate that one of the most relevant feature is the presence of the "Alef" letter as first letter of many Arabic words [8]. As this hypothesis is not always verified, and due to the natural homogeneous gaps betweens PAWs, the authors achieve only 60% of correct word segmentation.
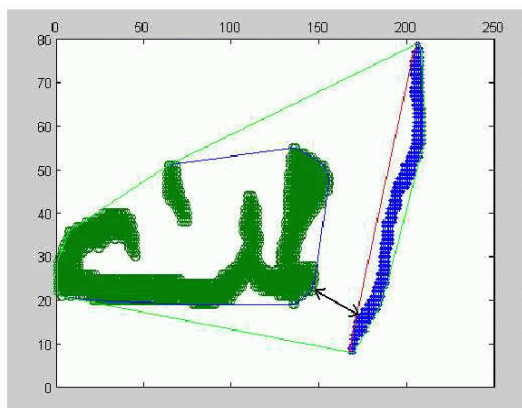

Figure 5: Gap calculation by Srihari et al. in [8]

Al-Badr et al. consider in [28] that segmenting Arabic words into letters is a too much difficult task considering the particular nature of the Arabic writing style, even if it is type printed text. It is why they propose a segmentation-free approach to recognize words. The key idea is to detect a set of shape primitives on the analyzed word, and to arrange them at best in the word space. The interpretation of each primitive depends of its context and position and the posterior probability maximization, allowing to tolerate local misrecognition. Word recognition scores varies according to if the words are clean (99.39%), degraded (95.60%) or scanned (73.13%).

This approach is not specifically dedicated to Arabic Script. Indeed the primitive shapes are very classical: lines of different lengths and orientations, corners, arcs,

curves, etc. The independence with the language is so important that the authors assume that they recognize isolated word. Hence, they elude the important problem of Arabic word segmentation even though the event was underlined in the introduction of their paper.

Amin and Mansoor [29] proposed an MLP-based holistic word recognition method for handwritten Arabic words. The MLP input is a global vector composed of 6 kinds of feature vectors carefully chosen to represent globally the word, such as: number of sub-words (up to five), number of peaks of each sub-word (up to seven), number and position of complementary characters and curves within each peak with height and width of the peak. Features are dedicated to Arabic word representation, making the system very specific to the language, even though the models are classical. The recognition rate of 98% on different fonts accredits the interest of adapted language specific features.

But here also a question remains: how the words are located in the text? Even though the authors discusses the problem of PAW extraction and their interest in the word recognition, never they explain how they gather several sub-words into a whole word.

Farah et al. [10] use a battery of three NNs for word recognition, each one is fed by some specific features: statistical, structural or a mixture. Then several combination procedures are tested. The NNs used are classical MLPs with back-propagation algorithm for training. Statistical features are language-independent as they are pixel-based information: the features are the pixel density in various homogeneous zones of the image. Some of the structural features are identical to what is used in Latin systems: ascenders, descenders, loops, writing baseline. Some others are specific to Arabic writing: presence and number of diacritic dots and their position according to the baseline. Words are already isolated in the database, leading the authors to not address in this work the problem of their location. The tests on 2400 word images from 100 different writers achieve 94.93% recognition rate, for 0.97% errors and 4.10% of rejection.

Pechwitz and Märgner [19] used semi-continuous HMMs (SCHMM) representing characters or shapes, as developed by Huang [20]. For each binary image of a word, parameters are estimated after a pre-processing phase normalizing the size and the skew of the word. Then, features are collected using a sliding window approach, leading to a language-independent features (see Figure 6). As in Latin script the middle band of the writing contains the word complexity. For Arabic writing it seems better to look at 3 lines parallel to the baseline at fixed position. The Viterbi algorithm is then used for training and recognition. In training phase a

segmental k-means algorithm is performed. The recognition is done by applying a frame synchronous network Viterbi search algorithm together with a tree structured lexicon representing valid words. The models were combined into a word model for each of 946 valid city names. The system obtained 89% word-level recognition rate using the IFN/ENIT database (26,459 images of Tunisian city-names). The words are yet isolated in the database, thus this work does not have to deal with the word segmentation problem.
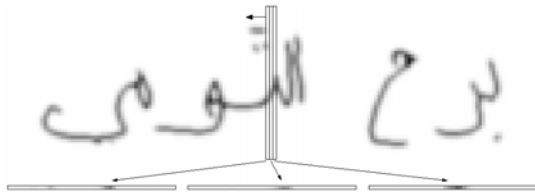

Figure 6: Feature extraction considered by Pechwitz and Märgner in [19]

Khorsheed and Clocksin propose the use of spectral features for printed Arabic word recognition [33]. As mentioned in several works the problem of word segmentation is discarded, assuming to have word images at the input. The originality of this work is in the use of a polar transformation coupled with a Fourier transform that allows to deal with rotation problems (see Figure 7). In a multi-fonts approach the system obtains 95.4% of good word classification by a simple matching with prototypes using the Euclidian distance on 1700 samples of different size, angle and translation.
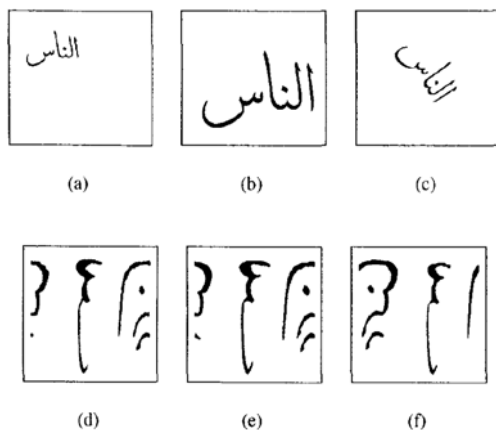

Figure 7: Polar transform of three images of the same word with different scales and rotation [33]

These works clearly accredit the word superiority principle of Mc Clelland and Rumelhart. Indeed many feature combinations and models perform very well.

But the Mc Clelland and Rumelhart model is not respected as only two levels are considered: input features level and word level. This allows to need less precision in the feature research, but the vocabulary should be limited and each word needs its own modeling.

We note that when the input uses low-level features, it is not necessary to introduce specific information, contrarily to the high-level features which need to be adapted to the writing nature. Perhaps should we extend the Mc Clelland and Rumelhart vision model with another layer, linking the pixel information with high-level features, as works the brain with the input images?

Once having a look on these different systems, we can now answer questions asked in the beginning of this section.

The first question is about the direct reuse of Latin systems on Arabic writing. This question has two answers, considering either the models, or the nature of the basic information extracted.

Considering the models, this is clear that almost Arabic systems use the same models and measures than Latin ones. We can deduce that the adaptation of the models is not necessary. Its logical because all the classical models perform information without a priori on its nature [8,10,19,28,29].

Considering the information extraction, all the approaches maintain some classical features used on Latin script. Low-level information based approaches seem able to avoid the add of specific features as they learn them directly [8,29,33]. When high level features are considered, the particularities of Arabic writing lead some authors to search for more specific features, like diacritic points, elongations and curves in the beginning or the end of words [10,29].

The second question asks about Arabic word segmentation (i.e. location in the text) possibilities. The answer is clearly no as the only work that proposes a segmentation method obtains really low segmentation results [8]: all others authors assume to deal with segmented words without approaching this problem.

Arabic word segmentation is much more difficult than for Latin writing for different reasons. The major one becomes from the PAW level which introduces a natural segmentation of the writing, with similar intra-word and inter-words gaps. This problem is underlined by Al-Badr and Haralick in [28] that indicate that the justification of Arabic text is not based on inter-word space adjustment but on elongation of some parts of words. Khedher and Abandah confirm this fact by a statistical study of 262647 Arabic words in [34], showing an average of 4.3 letters and 2.2 PAWs per word: they conclude that the PAW level is the real basic block to be processed rather than word level.

But curiously several works assume a prior word separation, without considering the difficulty of the task. Perhaps is it a reminiscence of Latin works where word segmentation is much easier? But we can think that such an hypothesis can explain why it is so few practical, industrial applications on Arabic language despite of all these good results.

3.2 Semi-global-based Vision Classifiers

The particular nature of Arabic writing allows us to describe the language in a fewer natural level: the PAW level. Indeed Arabic words are built by concatenation of several independent written parts that give another natural segmentation level. This natural segmentation allows us to refine the analysis by reducing the basic vocabulary. It is why some approaches have based their work on this level.

The principal effect of reducing the base vocabulary is the possibility to extend the dictionary. Ben Amara et al. illustrate this fact in [15] where the PAW level allows to deal with a medium vocabulary of city names, that is usually not treatable with a word global approach. The proposed system used a PHMM (Planar Hidden Markov Model) adapted to the PAW morphology. Hence, a shape is vertically decomposed into five horizontal bands, corresponding respectively from top to bottom to the ascenders, the upper diacritic dots, the writing central band, the lower diacritic dots and to the descenders. At each band is associated a super state for which corresponds an horizontal HMM modeling the concerned zone (see Figure 8).

Although the PHMM is a classical model yet used for Latin script [22], its use is clearly dedicated to Arabic writing by the integration of Arabic specificities inside the model. Arabic features are directly integrated in the PHMM by adapting the principal HMM to locate the 5 specific bands of Arabic shapes, while secondary HMMs model the stroke length variations and the diacritic information, aspects that are very specific to Arabic writing. Results are very good as the authors achieve up to 99.84% good recognition for 33168 samples from a vocabulary of 100 PAWs [31].
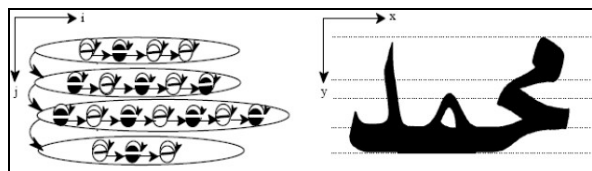

Figure 8: the PHMM architecture as defined by Ben Amara in [15,31]

Burrow confirms in [17] that Arabic word segmentation is harder than for Latin script and proposes to tackle the problem by recognising the PAWs separately. He hopped that this method will be able to cope with the large lexicon of the full database (i.e. IFN/ENIT with 946 different town names from 411 different writers). Interested only by the word shape, he considered the tracing approach as detailed in [10,30] for Latin script. He will in effect be converting an off-line representation into pseudo-on-line representation. Because of diacritics are points, their tracing does not have sense; so, they are discarded from the PAW images. Then, a PAW is transformed in an ordered series of points describing the trace.

Once K-NN classification approach is applied on each PAW, a majority vote is taken on its overall class and repeated for each PAW sample.

First results give 47% accuracy on PAWs. Refining the scoring system and adding some features including the number of dots, this allow the author to score at 74% for PAWs on correctly represented classes. Global word recognition with the add of word-global features is studied and improve greatly results, but one more time the input images are supposed to be entire word in this case: no study is made on the possibility to gather PAW information to find words in a text line.

Concerning the dependence or not of the features on the language specificity we can observe one more time that on one side low-level features are able to simply deal with feature language specificities as Amara showed (i.e. horizontal run length) by adapting the PHMM model [15,31]. On the other side, structural features proposed by Burrow need to be correctly chosen in order to reflect language specificities, and the use or not of them highly influence the results [17].

Curiously, very few work have been made on the PAW level. As assessed in the previous section, this level is however the natural global level of Arabic writing. Thus this section leads to a similar question: is it the influence of Latin works that tend to recognize whole words rather than PAWs?

This is right that for Latin human reading, the PAW level does not exist. The McClelland and Rumelhart vision model confirms this fact, as no intermediate level is given between letters and words. Two solutions exist to adapt this model: the first one is to extend it by adding a PAW layer between letters and words; the second one is to decrease the word level to PAW level, assuming that it is the real global level. The second solution has some advantages and global approaches would benefit to be applied to PAW level for several reasons:

- firstly, the PAW vocabulary is reduced according to the word vocabulary. Thus it is easier to deal with larger vocabularies.
- secondly, as PAW give a natural segmentation of word, the word representation will integrate

it in a way or another. It is thus more logical to divide the representation according to these limits.

- thirdly, it transforms the word segmentation problem into a PAW gathering problem. Now the segmentation problem has only empirical solutions, whereas the PAW gathering can use theoretical frameworks as HMM that can guarantee the optimality of the solution.

We remark that the McClelland and Rumelhart vision model could be extended to a more general approach, where the information gathering could be made recursively through as many levels as necessary. This idea is reinforced by the fact that "good readers" are able to recognize word groups rather than isolated words [44]. A level-recursive approach can then simulate this fact by gathering information through several higher-abstract structures.

3.3 Local-based Vision Classifiers

In this vision level the objective is to focus on letters or smaller entities for their interpretation. The process is thus to gather, bind, confront these entities to identify the word. But such an analysis level leads to the Sayre dilemma: to find letter limits human has to recognize them, and to recognize them human have to localize them. This problem is usually eluded by the use of implicit or explicit segmentation methods.

Fahmy and Al Ali proposed a system with structural features [11]. During a pre-processing phase, slopes and slants are corrected, then some measurements are achieved like stroke width and height of letters. Word is then normalized and encoded in a canonic form, using a skeleton coding approach used on Latin script [21] but adapted to the Arabic writing style. The word image is divided in several frames focusing on character parts, and each frame is divided into three segments. Then, classical features like turnings, junctions and loops were detected from skeletons and used as the input of an ANN. The number of inputs is 35, representing 11 features for each of the three segments of a frame, plus two inputs representing dots. One of these two inputs represents dots if they are above the baseline, while the other input represents dots if they are below the baseline. A recognition rate of 69.7% was obtained on 300 different words written by one writer, with a second writing of the 300 words for the training stage.

Let note that the system try to class the frames: there is no try to gather frames to form a complete character. Another point concerns the word segmentation: here words are separated at the writing time, eluding the difficult problem of word segmentation.

Trenkle et al. propose in [32,39] a printed text recognition system based on an over-segmentation approach (see Figure 9). A full page of text is divided in blocks and lines, and each line is segmented into atomic segments, that are part of character. During the recognition atomic segments are grouped in order to retrieve the whole character according to a Viterbi analysis. Each segment group gives 424 features obtained from horizontal and vertical projections and a edge-based chain code.

An NN is used to classify the segment groups: it has 229 outputs according to the 117 regular Arabic character forms, 80 ligature forms, 10 Arabic digits, 20 punctuation characters, and 2 rejection classes. Classification is also done by a set of decision trees. A Viterbi beam search allows to find the best decoding path on the entire line given an Arabic text model in which one are encoded the rules of Arabic typography. The model combines lexicon-free and lexicon-based approaches, with a vocabulary of 50000 common Arabic words. A dataset of 722 text images of different qualities is used for the realistic tests: the neural network achieves 89.1% of good recognition as the set of decision trees obtains 90.7% recognition rate.

This work is very complete as it addresses all of the problems of printed Arabic text recognition, from page processing and segmentation to text recognition with an ASCII output. The word segmentation problem is elegantly solved by the use of a language model to gather information. We note that the features used are low-level based, allowing to not need integration of Arabic specificities at the character level.
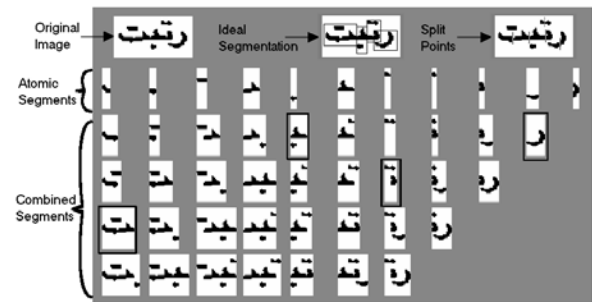

Figure 9: Over-segmentation applied by Trenkle et al. in [46,53]

In Abuhaiba et al. [14], a method for the recognition of free handwritten text is proposed. Based on the skeleton representation, the sub-words are segmented into strokes that were further segmented into "tokens". Tokens are single vertices representing dots or loops or sequences of vertices. A "fuzzy sequential machine" is employed to identify the classes. This machine is composed of sets of initial and terminal states. Stroke directions are used for entering states, and a function for transitioning between states. Tokens are either recognized directly or used to augment the recognizer.

This system achieves 55.4% of good recognition for PAWs with 17.6% of rejection, characters having 51.1% of correct answers with 29.3% of rejection. No lexicon was used, but the PAW vocabulary remains naturally limited. Although the computer used for tests is old and very slow, the approach needs very huge calculation time. As assessed by the authors commenting the relative results, the objective of this work is mainly to propose new theoretical basis and concepts.

Clocksin and Fernando propose in [12] an analytic system for Syriac manuscripts, a West Semitic language which is less grammatically complex than Arabic. The word segmentation is much simpler than for Arabic writing as the intra-word gaps seem to be clearly smaller than inter-word gaps (see Figure 10), contrarily to Arabic writing. But as the grammatical function are almost written as word prefixes or suffixes instead of as separate words, it is not possible to have a global word approach without a very huge dictionary: this language construction is close to the Arabic one. Thus the authors focus on character recognition.

Figure 10: Syriac writing: inter-word gaps are larger than intra-word gaps [12]

A text page is then segmented into words using horizontal and vertical projections. Words are segmented into letters by over-segmenting and removing bad segmentation points, with a segmentation approach specially adapted to Syriac writing: approximately 70% of the characters are correctly segmented. Some features are extracted from character images: moments of different parts of the image and polar transformation. Classification is based on Support Vector Machine (SVM) considering a 'one against one' scheme. As in Arabic writing, a letter can have different shapes and thus can belong to different classes. The best feature combination gives 91% recognition rate for manuscript word letters, and 97% for typeset word letters.

As other previous works, this one uses low-level features, and its adaptation to the language specificities is made through the model learning. The low rate of 70% good character segmentation accredit the fact that, as for Latin scripts, a letter segmentation cannot be done correctly.

Miled et al. [18] proposed an analytical approach based on HMMs for the recognition of Tunisian state names. They integrate the notion of PAW in their system. They group letters having the same body but different diacritics in order to "solve" the problem of diacritic detection and classification. A text line is segmented into PAWs and isolated letters, then PAWs are segmented into graphemes using their upper contour and some heuristic rules, in a way that looks very similar to approaches for Latin script. Each grapheme is encoded into two vectors: the first contains topological features corresponding to human perception, like loops, openings, relative size, relative position, etc., the second one containing moment-like descriptors (here Fourier descriptors were kept). The grapheme identification is based on a K-NN classifier which obtains 84.90% of good recognition.

HMM are used to describe the word composition. Word models are built by concatenating grapheme states, considering several ways to deal with over- and sub-segmentation and a space state to deal with intra-word gaps (see Figure 11). The approach is 'flat': each word has its own HMM representation, and the input sequence is analyzed by all these models. Tests are made on isolated word images, and reach 82.52% of good recognition with a lexicon of 232 words.

This work does not focus the word separation problem, but the HMM approach can solve it at an higher representation level. An interesting aspect of this work is to consider over- and under- segmentation and to integrate the intra-word separation as a blank character: all the segmentation problems of Arabic script are then covered.
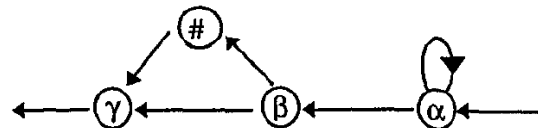

Figure 11: Word model proposed by Miled et al. [18]. α, β, and γ stand for characters, and # represent a space

Fakir et al. propose to use the Hough transform for the recognition of printed Arabic characters [38]. A full text page is digitized, noise cleaned, deskewed and segmented into lines. Then lines are segmented into words assuming that it is a bigger gap between words than between PAW. Character segmentation is based on the projection profile of the middle zone of the word with a fixed threshold that determines the breaks in the projection profile. A second segmentation is applied in order to extract diacritics.

Features are then extracted using the Hough transform, which is applied on the character skeleton to detect strokes (see Figure 12). Thus a character is

represented as a set of strokes. At recognition step the set is compared with the one of reference patterns. A second stage ends the recognition by refining the classification according to the diacritic information. 95% of correct recognition is achieved on 300 characters obtained from the segmentation process. The most common confusions are due to the thinning that brings closer some different patterns.

No information is given about the effectiveness of the segmentation itself, but 300 words were collected and only 300 letters were used for the test: this let suppose that not all the characters were correctly segmented.
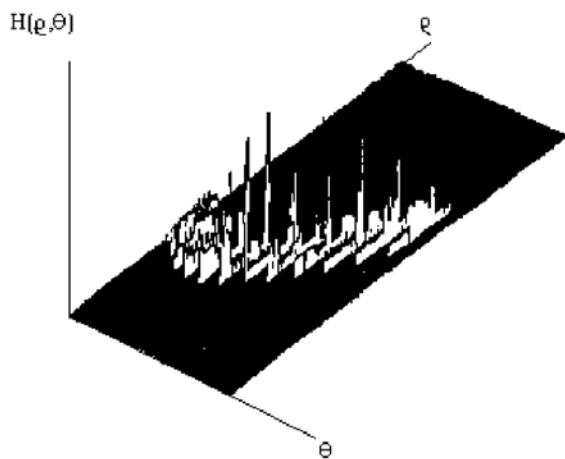

Figure 12: Array of accumulators content given by the Hough transform of a character [38]

Some researches are dedicated to character recognition. They assume that text can be segmented purely in letters. This assumption is not realistic: the segmentation problem was raised long time ago for Latin script, and many studies showed that it is globally not possible. As it seems harder to segment Arabic script, logically this segmentation will not be possible. This point is confirmed by Sari et al. [40] who propose a segmentation system dedicated to Arabic writing: they obtain only 86% of correct letter segmentation.

The interest of these works based on character recognition is reduced to the models and features used that could be integrated into others more complete works.

Alnsour and Alzoubady proposed in [35] a Neocognitron to classify handwritten characters. The input of this particular NN is composed of structural features: Freeman code chain, coordinates of starting and ending points, loops, and primitives like segments with different orientations, corners and dots. These features allow them to achieve 90% character recognition with 3.57% rejection. The system assumes a context of a handwritten Arabic document recognition

software that should be able to segment words directly in letters.

Asiri and Khorsheed propose to use two different NN architectures for handwritten Arabic character recognition [36]. The first architecture has 6 output nodes, and is designed to classify the character into one of the 6 groups of similar shapes. According to this first answer, a second NN that correspond to the group will take the final decision. For all NNs the input correspond to a certain number of Haar wavelet transform coefficients. The best results are achieved for 1024 coefficients and give 88% good recognition.

For this work, character samples were collected individually: writers are asked to write isolated letters into small rectangles.

Cowell and Hussain worked on isolated Arabic printed characters [37]. A character image is normalized in 100x100 pixels, then a signature is extracted by counting the black pixels in each row and column. This signature is compared to those of a template set, and the modulus of the difference for each row and each column is summed: lower is the value, closer should be the forms.

The objective is to have a very quick matching method: here the signature matching carries 200 comparisons per template, against 10000 for a direct image matching. No clear result is given but the confusion matrix let suppose a 100% recognition rate.

As explained previously, several approaches give good results, showing that as for Latin script the analytic approach can perform very well. But such an approach presents some drawbacks. First one is the classical problem of bad segmentation that can lead to over- or under- segmentation; such errors generally lead to misclassification. More, the segmentation process has to be adapted to Arabic script, in order to take into account its specificities like vertical ligatures: thus Latin segmentation methods cannot be used efficiently without adaptation.

Some others problems lie in the approach itself, possibly accented by the language.

Thus one problem lies in the observation independence hypothesis. As letters or segments are recognized independently, any error perturbs the whole recognition process. In fact, the McClelland and Rumelhart word superiority effect is not taken into account because the word is not considered as a whole but as a sum of small parts.

Another problem is the inadequacy between segmentation and models. Indeed, the segmentation is based on structural information that is totally independent of the model nature. Thus the modeling is biased at the source by forcing the model to align on non-optimal limits. This is usually solved in two ways:

- the first way deals with the use of higher level features: as the image is interpreted, the distortions are implicitly removed. The drawback is that any bad interpretation of the image makes lose a huge quantity of information, often leading to a misclassification.
- The second way deals with the use of an implicit segmentation: thus the models are able to "choose" their best limits. Unfortunately all the models cannot be used in such an approach without exploding the calculation time.

We note that in Arabic script the notion of letter limits is very variable as horizontal stroke elongations frequently occur in letters and letter ligatures. This accredits the point of view of Choisy [42] that proposed to not search any letter limits: thus the model can focus on pertinent letter information without taking hard separation decision on the fuzzy ligature parts. This proposal fits the McClelland and Rumelhart model, where precise information position is not important but only its presence in an approximate location.

The recognition process is basically based on the use of a combination of a random field (Non Symmetric Half Plane) drawing its observation directly in the image and a HMM taking into account the column observations in the image, hence tackling the problem of length word variations. Figure 13 shows the aspect of this system applied in our Laboratory on Bangla script presenting some calligraphic similarities with Arabic.
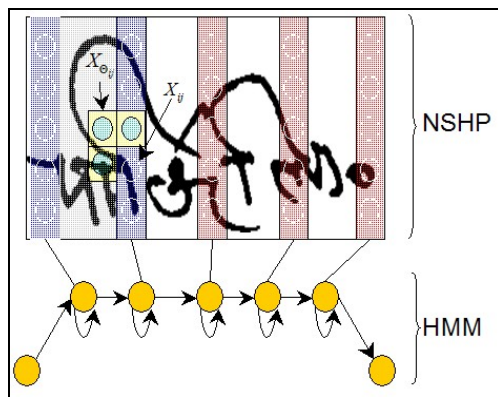

Figure 13: The NSHP-HMM system applied on Bangla script.

3.4 Hybrid-level classifiers

It is possible to combine different strategies so as to approach more the principle of human reading: the analysis must be global for a good synthesis of the information, while being based on local information suitable to make emerge this information [16,23]. Such a combination better fits the human reading, that is proved to firstly analyze global word shapes and searches for local information only to discriminate ambiguous cases.

As local-based approaches gather local information up to words, they could seem to be hybrid ones. But there is an important difference between the two approaches: hybrid approaches try to have a multi-level analysis of the writing, when local-based approaches are only based on the gathering of local information. The aim of hybrid approaches is to combine different levels of features and interpretation, leading to systems closer to the McClelland and Rumelhart proposal.

Souici et al. [24] propose a neuro–symbolic hybridization considering that people rarely, if ever, learn purely from theory or examples. A hybrid system that effectively combines symbolic knowledge with an empirical learning algorithm might be like a student who is taught using a combination of theoretical information and examples. The neural and symbolic approaches are complementary, so their integration is an interesting issue.

For that purpose, they defined a neuro-symbolic classifier for the recognition of handwritten Arabic words. First structural features are extracted from the words contained in the amounts vocabulary. Then, a symbolic knowledge base that reflects a classification of words according to their features is built. Finally, a translation algorithm (from rules to NN) is used to determine the NN architecture and to initialize its connections with specific values rather than random values, as is the case in classical NNs. This construction approach provides the network with theoretical knowledge and reduces the training stage, which remains necessary because of styles and writing conditions variability. The recognition rate varies from 83.55% (4.75% substitution) given by the rule-based approach, 85.5% (14.5% substitution) given by the NN, to 93% (7% substitution) given by the combination.

A similar approach has been applied to handwritten Arabic city-names recognition [43]. The Knowledge Based Artificial NN (KBANN) generated using translation rules is compared to a classical MLP. As the MLP obtains 80% on a 55 vocabulary words, the KBANN performs 92%. The MLP has a less complex architecture than the KBANN, but has a little more neurons.

The hybrid aspect of these work resides in the NN creation: it is based on a multi-level word description, that considers different levels of rules to classify the word according to its number of PAWs, its features and its diacritic information. Thus the network implicitly looks at different perception levels.

Maddouri et al. proposed a combination of global and local models based on a Transparent NN (TNN) [23]. This model is stemmed from the model proposed by

McClelland and Rumelhart for global reading and adapted by Côté [25] for Latin recognition. The TNN is composed of several layers where each one of them is associated to a decomposition level of the word. As Coté's TNN had three layers corresponding to features, letters and words, Maddouri extended it to take into account the Arabic PAW level. Hence the first level corresponds to features, the second to letters, the third to PAWs and the fourth to words. In each level the NN cells represents a conceptual value: primitive, letter, PAW or word (see Figure 14). Training was operated manually by fixing the weights for the cell connections. These weights are determined statistically knowing for each word the various decompositions in the three conceptual levels.

The recognition process is operated during several perceptive cycles, propagating hypotheses from the feature level to word level, looking for their association to the composition levels of the word, and retro-propagating information from the word level to refine the previous features or to extract others. More precisely, in propagation movement the global model proposes a list of structural features characterizing the presence of some letters in the word. Then it proposes a list of possible letters, PAWs and words containing these characteristics. In the back-propagation movement, the activated words and PAWs emit some hypotheses on the letters that could be present. These hypotheses conduce to research the corresponding features, or directly the letter if it has no robust feature. In this last case, a correspondence between the letter image and the corresponding printed one is performed by a local-based model using the correspondence of their Fourier descriptors, playing the role of a letter shape normalizer.
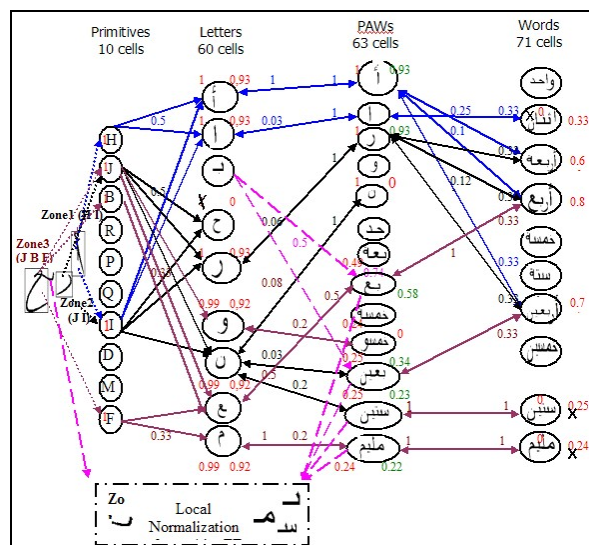
This principle was applied to PAW and word recognition, PAW recognition being made by removing the word layer. The handwritten database contains 2100 images of the 70 word vocabulary of Arabic literal amounts, containing 63 different PAWs. Using only global features, which is like to a simple propagation, PAW recognition rate is 68.42% and word recognition rate is 90%. The add of local features in next perceptive cycles permits to reach a score of 95% for PAWs and 97% for words.

The interest of this approach is to progressively refine the analysis according to the discriminative need of the word dictionary: thus for very distinct shapes a simple propagation can be sufficient, while for words having close shapes more precise information is needed to discriminate them. The "drawback" leads in the information localization, that becomes more and more difficult with the information precision, but this "problem" is inherent to the spirit of such an approach.

This work is very close to the McClelland and Rumelhart approach. The word superiority effect is raised as word recognition performs better than PAW recognition: word shape features are thus sufficient to achieve correct results. The analysis refining principle is clearly efficient, as shows the grow of the scores.

As it is very few hybrid approaches for Arabic writing, we just have a look on some other interesting works on another languages.

Pinales and Lecolinet in [26] proposed a system which is both analytical and global and emphasizes the role of high-level contextual information (see Figure 15). This model is based both on a top-down recognition scheme called backward matching and a bottom-up feature extraction process which is working in a competitive way. This approach has some similarities with the TNN proposed by Côté and Maddouri, as words "retro-propagate" their information to resolve ambiguities and complete missing letters. First results are very encouraging and show that such an approach is very pertinent.
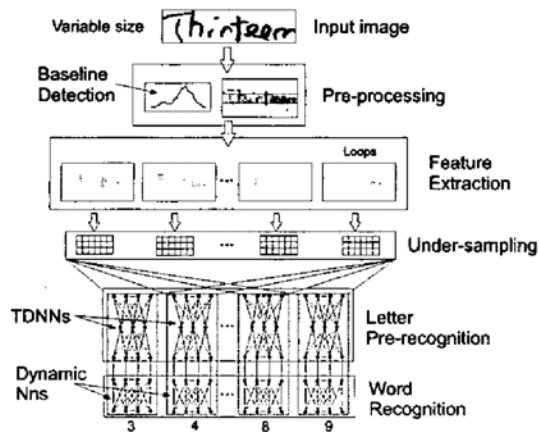

Figure 14: The TNN approach as defined by Maddouri in [23] on the word "Arbaa"

Figure 15: Pinales and Lecolinet Neural Networks combination architecture [26]

Another work has been proposed by Choisy on Latin script [16]. The approach proposes to use the elasticity property of the NSHP to normalize word images in a non-linear way. NSHP states focus pixel features according to the learned ones, giving an implicit state-based normalization. The normalized images are then analyzed by a classical Neural Network. Results show the efficiency of this approach. The drawback of this approach is the compression of the information: even though, contrary to the SDNN, there is a real adaptation of the NN input to the image, normalization is a source of information loss.

Results show that hybrid approaches are very efficient. Corroborating the McClelland and Rumelhart approach, the multi-level analysis allows to refine the analysis with more flexibility than other approaches. In particular, the principle of information focalization rather than segmenting is raised, and seems to be an important point in hybrid approaches.

We can conclude that even if it is not proved that the McClelland and Rumelhart model is the right one to simulate human reading, their approach leads to efficient reading systems. This is very interesting because it links the psychology works and the forma representations by computers. We still remark that the proposed model is clearly oriented towards Latin script reading. For Arabic script it seems necessary to extend the model with a PAW-level. Thus the McClelland and Rumelhart principle is validated but the corresponding model should be adapted to the language considered.

4 Conclusion

Several conclusions can be drawn considering all the research reported in this paper about the recognition of Arabic writing regarding the reading aspect theoretically speaking without neglecting the language characteristics.

First one is that low-level features are language independent. Once extracted (similarly for all the scripts), the training process can arrange their proximity to the language studied. At the opposite, high-level features are language-dependent, and thus need to develop specific extraction methods to retrieve all the information. Obviously, a combination of these two kinds of features should perform better, each feature level complementing the drawback of the other.

Another point is that the PAW level is very important for Arabic script modeling: contrarily to Latin script, the basic entity is not the word. Global approaches should be based on PAW. Analytical ones gain to integrate this information level. A first effect is to reduce the vocabulary complexity by gathering the information on an intermediate level.

Considering the reading approaches, hybrid ones seem to be very promising. They efficiently combine different perceptive levels, allowing to discriminate words without a complete description. In comparison with global approaches, the add of local information allows to extend the vocabulary with less confusions. Compared to local approaches, hybrid ones avoid the full-segmentation problems, and are less disturbed in case of information loss.

Another conclusion stemmed from the works themselves. In particular, two points were raised as very problematical: the segmentation in words and in letters.

The letter segmentation problem was raised long time ago for Latin scripts. Several works were made on this case, and nowadays it is commonly accepted that this problem has no solution. As Arabic writing is mainly described as more complex than Latin one, it seems obvious that a letter segmentation cannot be effective. This leads to the question: why many works are based on such an hypothesis? It seems that the experience gained on Latin languages was not totally transposed to Arabic writing.

Concerning the word segmentation (i.e. location in the text), the problem was probably hidden by the works on Latin script where word separation can be considered as a problem solved in many cases. But for Arabic writing, several works accredit the difficulty of this task. As a word analysis is interesting to show the power of the reading approaches, there is a gap between their modeling and their extraction. We think that the extraction will gain to be made by gathering PAWs through a mathematical formalism like HMM.

Globally we observe that few researches try to take into account the whole problematic of the Arabic script. Thus the word segmentation problem is mainly eluded, the PAW-level global recognition was the object of very few works, and several segmentation-based

approaches made the irrelevant hypothesis of a pure letter segmentation.

Some other problems, like elongations and vertical ligatures, are often cited in the Arabic script description, but less often taken into account in the work itself.

Thus it seems that the main experience brought from Latin works concerns the models and the features, but not the problems encountered and the processes followed to solve them. In fact, many specific problems were raised, but a lot of works consist to try another set of models, features, methods, coming from Latin works.

References

[1] N. Essoukri Ben Amara and F. Bouslama, Classification of Arabic script using multiple sources of information : State of the art and perspectives, in Int. Journal on Document Anal. And Recognition, IJDAR, 2003, 5 :195-212.

[2] L. M. Lorigo and V. Govindaraju, Offline Arabic Handwriting Recognition: A survey, in IEEE Trans. on Pat. Anal. and Mach. Int. (PAMI), vol. 28, n. 5, pp. 712-724, may 2006.

[3] A. Amin, Off-line Arabic Character Recognition: the state of the art, in Pattern Recognition, vol. 31, n. 5, pp. 517-530, 1998.

[4]
http://www.islamicart.com/main/calligraphy/styles/kufi.htm

[5] J. L. McClelland and D. E. Rumelhart, An interactive activation model of context effects in letter perception, in Psychological Review, 88: pp. 375-407, 1981.

[6] J. L. McClelland and D. E. Rumelhart, Distributed memory and the representation of general and specific information, in Journal of Experimental Psychology: General, pp.159-188, 1985.

[7] T. Steinherz, E. Rivlin and N. Intrator, Off-line cursive script word recognition: a survey, IJDAR (1999) 2: 90-110.

[8] S. Srihari, H. Srinivasan, P. Babu and C. Bhole, Handwritten Arabic Word Spotting using the CEDARABIC Document Analysis System, Symposium on Document Image Understanding Technology, The Marriott Inn and Conference Center, University Maryland University College, Adelphi, Maryland November 2-4, 2005.

[9] C. I. Tomai, B. Zhang and S. N. Srihari, Discriminatory Power of Handwritten Words for Writer Recognition. Proceedings of the 17th International Conference, 23-26 Aug. 2004 Page(s):638 - 641 Vol. 2.

[10] N. Farah, M. T. Khadir and M. Sellami, Artificial neural network fusion: Application to Arabic words recognition, ESANN'2005 proceedings - European Symposium on Artificial Neural Networks, Bruges (Belgium), 27-29 April 200

[11] M.M.M. Fahmy and S. Al Ali, "Automatic Recognition of Handwritten Arabic Characters Using Their Geometrical Features," Studies in Informatics and Control J., vol. 10, 2001.

[12] W.F. Clocksin and P.P.J. Fernando, "Towards Automatic Transcription of Syriac Handwriting," Proc. Int'l Conf. Image Analysis and Processing, pp. 664-669, 2003.

[13] H. Almuallim, S. Yamagochi: "A method of recognition of Arabic cursive handwriting", Pattern recognition, vol. 9, Nr 5, pp. 715–722, September 1987.

[14] I.S.I. Abuhaiba, M.J.J. Holt, and S. Datta, "Recognition of Off-Line Cursive Handwriting," Computer Vision and Image Understanding, vol. 71, pp. 19-38, 1998.

[15] N. Ben Amara, and A. Belaid, Printed PAW recognition based on planar hidden Markov models Proceedings of the 13th International Conference on Pattern Recognition, 25-29 Aug. 1996 Page(s):220 - 224 vol.2.

[16] Ch. Choisy and A. Belaïd, Coupling of a local vision by Markov field and a global vision by Neural Network for the recognition of handwritten words, in ICDAR'03, Edinburgh, 3-6 August, pp. 849-953, 2003.

[17] P. Burrow. Arabic Handwriting Recognition, Report of Master of Science School of Informatics, University of Edinburgh 2004.

[18] H. Miled, C. Olivier, M. Cheriet, Y. Lecourtier: "Coupling observation/letter for a Markovian modelisation applied to the recognition of arabic handwriting". Proc. of ICDAR'97, pp. 580–583, Ulm, Germany, 1997.

[19] M. Pechwitz and V. Maergner, HMM Based Approach for Handwritten Arabic Word Recognition Using the IFN/ENIT- Database, 7th ICDAR, Vol. 2 p. 890-894, 2003.

[20] X. D. Huang, K.F. Lee and H. Hon, "On Semi-continuous Hidden Markov Modeling, " in Proc. of the IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 689--692, April 1990.

[21] A. W. Senior et A. J. Robinson. An Off-Line Cursive Handwriting Recognition System. PAMI, 20(3) :308–321, March 1998.

[22] R. Bippus. 1-Dimensional and Pseudo 2-Dimensional HMMs for the Recognition of German Literal Amounts. ICDAR'97, vol 2, pages 487–490, Ulm, Germany, Aug. 1997.

[23] S. Snoussi Maddouri, H. Amiri, A. Belaïd and Ch. Choisy, Combination of Local and Global Vision Modeling for Arabic Handwritten Words Recognition, in Eighth IWHFR, pp. 128-132. 2002.

[24] L. Souici, and M. Sellami, "A hybrid approach for Arabic literal amounts recognition," AJSE, the Arabian Journal for Science and Engineering, Volume 29, Number 2B, pp: 177-194, October 2004

[25] M. Côté, E. Lecolinet, M. Cheriet and C. Y. Suen, Building a perception based model for reading cursive script, in ICDAR, vol. II, pp. 898-901, 1995.

[26] J. Ruiz-Pinales and E. Lecolinet, A new perceptive system for the recognition of cursive handwriting, ICPR, vol. III, pp. 53-56, 2002.

[27] B. Zhang and S. N. Srihari, "Binary vector dissimilarity measures for handwriting identification," Proceedings of the SPIE, Document Recognition and Retrieval , pp. 155–166, 2003.

[28] B. Al-Badr and R. M. Haralick, A segmentation-free approach to text recognition with application to Arabic text, in Int. Journal on Document Anal. And Recognition, IJDAR, 1998, 1:147-166.

[29] A. Amin and W. Mansoor, Recognition of Printed Arabic Text using Neural Neworks, in ICDAR'97, volume II, pp 612-615, 1997

[30] H. Nishida. An approach to integration of off-line and on-line recognition of handwriting. Pattern Recognition Letters, 16(11):1213–1219, 1995.

[31] H. Miled and N. E. Ben Amara, Planar Markov Modeling for arabic Writing Recognition : Advancement State. ICDAR 2001, volume I, pp 69-73, Seattle, 2001.

[32] J. Trenkle and A. Gillies and E. Erlandson and S. Schlosser and S. Cavin, "Advances in Arabic Text Recognition", Proceedings of the Symposium on Document Image Understanding Technology, Columbia, Maryland, 2001.

[33] M. S. Khorsheed and W. F. Clocksin, Spectral features for Arabic word recognition, ICASSP'00, page(s): 3574-3577, vol.6, 2000.

[34] M. Z. Khedher and G. Abandah, Arabic Character Recognition using Approximate Stroke Sequence, Arabic Language Resources and Evaluation Status and Prospects, LREC2002, Las Palmas de Gran Canaria, 1st June 2002

[35] A. J. Alnsour and L. M. Alzoubady, Arabic Handwritten Character Recognized by Neocognitron Artificial Neural Network, University of Sharjah Journal of Pure and Applied Sciences, vol. 3, n°2, june 2006.

[36] A. Asiri and M. S. Khorsheed, Automatic Processing of Handwritten Arabic Forms Using Neural Networks, Transactions on Engineering, Computing and Technology, vol. 7, august 2005.

[37] J. Cowell and F. Hussain, A fast recognition system for isolated Arabic character recognition IEEE Information Visualization IV2002 conference July 2002, London

[38] M. Fakir, M. M. Hassani and C. Sodeyama, On the recognition of arabic characters using Hough transform technique, Malaysian Journal of Computer Science, vol. 13, n°2, 2000

[39] A. Gillies and E. Erlandson and J. Trenkle and S. Schlosser, "Arabic Text Recognition System", Proceedings of the Symposium on Document Image Understanding Technology, Annapolis, Maryland, 1999.

[40] T. Sari, L. Souici, and M. Sellami, Off-line Handwritten Arabic Character Segmentation Algorithm: ACSA, Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02), p 452, Niagara-on-the-Lake, Canada, 2002.

[41] M. S. Jelodar, M. J. Fadaeieslam, N. Mozayani and M. Fazeli, A Persian OCR System using Morphological Operators, The Second Worl Enformatika Conference, WEC'05, pp 137-140, February 25-27, Istanbul, Turkey, 2005.

[42] Ch. Choisy and A. Belaid, Cross-learning in analytic word recognition without segmentation, in Int. Journal on Document Anal. And Recognition, IJDAR, 4(4): 281-289, 2002.

[43] L. Souici, N. Farah, T. Sari, and M. Sellami, "Rule Based Neural Networks Construction for Handwritten Arabic City-Names Recognition", Proceedings of the 11th International conference Artificial Intelligence: AIMSA 2004, Varna, Bulgaria, September 2004, , LNAI 3192, pp: 331-340, Springer, ISSN: 0302-9743.

[44]

http://home.ican.net/~galandor/littera/syn_cor1.htm