Cursive Arabic Handwriting Recognition System Without Explicit Segmentation Based on Hidden Markov Models
Mouhcine RABI1*, Mustapha AMROUCH1, Zouhair MAHANI2

1Laboratory IRF-SIC, faculty of sciences IbnZohr University, Agadir, Morocco
2 High School of Technology, IbnZohr University, Agadir, Morocco

*mouhcineh@gmail.com
m.amrouch@uiiz.ac.ma
zouhir.mahani@gmail.com

Abstract
In this paper we present a system for offline recognition cursive Arabic handwritten text which is analytical without explicit segmentation based on Hidden Markov Models (HMMs). Extraction features preceded by baseline estimation are statistical and geometric to integrate both the peculiarities of the text and the pixel distribution characteristics in the word image. These features are modelled using hidden Markov models. The HMM-based classifier contains a training module and a recognition module. The training module estimates the parameters of each of the character HMMs uses the Baum-Welch algorithm. In the recognition phase, feature vectors extracted from an image are passed to a network of word lexicon entries formed of character models. The character sequence providing the maximum likelihood identifies the recognized entry. If required, the recognition can generate N best output hypotheses rather than just the single best one. To determine the best output hypotheses, the Viterbi algorithm is used. The experiments on images of the benchmark IFN/ENIT database show that the proposed system improves recognition.

Keywords
Recognition; handwriting; Arabic text; HMMs.

I INTRODUCTION

The recognition of cursive Arabic handwriting is an active area of pattern recognition research. The variability of words, letter shapes are context sensitive, inter and intra word spaces, the cursive nature of Arabic handwriting, the skew and slant of characters and words makes the construction of offline recognition system a challenging task. Researches have tried various approaches for text recognition employing various techniques for pre-processing, feature extraction and classification [A.Lawgali 2015].

The subject of this article concerns the recognition of cursive Arabic handwriting [M.T Parvez 2013] [AL-Shatnawi 2011]. Several systems are available based on two approaches; a global approach that considers the word as non-divisible base entity avoiding the segmentation process and its problems. This approach is reliable and applicable for vocabularies of limited size. Against, the analytical approach is based on the decomposition of the word sequence into characters or graphemes proceeding by a segmentation phase. The latter can be explicitly based on a priori division of the image into subunits (letters or grapheme) or implicitly based on a recognition engine to validate and rank the segmentation hypothesis. The approach used in our system is analytical based on implicit segmentation; segmentation and recognition are carried out jointly.
The first step of a handwriting recognition system after preprocessing is the extraction features. The objective of this phase is the selection of primitives relevant for the next steps of classification and recognition. The performance of a recognition handwritten system largely depends on the quality and the relevance of the extracted features. In our system after the baselines estimation, the extracted features are statistics acting on the densities of pixels and structural extracted from the representation of the character shapes. Hidden Markov models (HMMs) are used for classification [S. Azeem 2013] [AlKhateeb 2011] [A. Maqqor 2014]. There are many reasons for success of HMMs in text recognition including avoidance of the need to explicitly segmentation. In addition, HMMs have sound mathematical and theoretical foundations. Each word is described by a model built by concatenating the models of the component character. The system performs training and recognition of words and characters.

![Diagram](Image)

Figure 1 presents the synopsis of the proposed system.

The remainder of this paper is organized as follow. Section 2 presents a detailed description of the features extraction preceded by baselines estimation. Section 3 is focused on classification step. The performance of the recognition system has been experimented on the benchmark database IFN/ENIT and the obtained experimental results are shown and analysed in section 4. The paper finally concludes with some conclusions and perspectives.

### II EXTRACTION FEATURES

#### 1.1 Baseline Estimation

The goal is to find, for a given word, the positions of the two following parallel lines (Figure 2):

- Lower baseline (LB),
- Upper baseline (UB).
These baselines divide the image into lower, upper and middle zone.

Several techniques exist for estimating these two lines, horizontal projection histogram, the Hough transform, the minima of the bottom contours [Farooq et al. 2005], the neighborhood approach and components, PCA (Principal Component Analysis) [B. Su, X. Ding 2013] [O. Morillot 2013].

The approach used is based on the horizontal projection curve that is computed with respect to the horizontal pixel density, knowing that the skew and slant correction of words are made in preprocessing step to harmonize the course of the sliding windows in the extraction features. LB corresponds to the maximum of the projection profile curve, then, the algorithm scans the image from top to bottom to find the upper baseline, which corresponds to the first line with a projection value higher or equal to the average row density. Thus, the handwritten variability of word in both zones, upper and lower are considered.

1.2 Extraction Features

The features extraction method used in our system is inspired by work of El-Hajj [El-Hajj 2009] with some modifications; the used technique has shown excellent results in several researches [A.-L. B-Bernard 2012] [T. Bluche 2015]. The features extraction stage consists of extracting a sequence of characteristics vector by dividing the word image into vertical frames. The sliding windows are shifted in the direction of writing (right to left). The width of each window is a parameter to set, the height of a window varies according to the dimension of the word image.

In each window we extract a set of 28 features representing the distribution features based on foreground pixels densities and concavity features. Each window is divided into a fixed number of cells. Some of these features are extracted from specific areas of the image delimited by the word baselines.

In our experimentation the parameters are set to n = 20 cells and the width = 8 pixels. This leads to a total of Nf = 28 to calculate in each frame.

Let:
- n(i) : the number of foreground pixels in cell i
- r(j) : the number of foreground pixels in the jth row of a frame.
- b(i) : the density level of cell i : b(i)=0 if n(i)=0 else b(i)=1.

The extracted features are the following:
- f1: density of foreground (black) pixels.
  \[
  f_1 = \frac{1}{H \times W} \sum_{i=1}^{n_r} n(i)
  \]
- f2: number of transitions black/white between two consecutive cells.
  \[
  f_2 = \sum_{i=2}^{n_r} |b(i) - b(i-1)|
  \]
f3: difference’s position of gravity centers of foreground pixels in two consecutive frame (current and previous)

\[ f3 = g(t) - g(t-1) \]

f4: normalized vertical position of the center of gravity of the foreground pixels in the whole frame with respect to the lower baseline.

\[ f4 = \frac{g - L}{H} \]

f5, f6: represent the density of foreground pixels over and under the lower baselines.

\[
\begin{align*}
    f5 &= \frac{\sum_{j=L+1}^{H} r(j)}{H.w} ; \\
    f6 &= \frac{\sum_{j=1}^{L-1} r(j)}{H.w}
\end{align*}
\]

f7: number of transitions black/white between two consecutive cells of different density levels above the lower baseline.

\[ f7 = \sum_{i=1}^{n} |b(i) - b(i-1)| \]

f8: zone to which the gravity center of black pixels belongs (lower zone f8=3, middle zone f8=2, upper zone f8=1)

f9, ..., f14: the concavity features are defined as:

\[ f9 = \frac{C_{left-up}}{H} ; \]

\( C_{left-up} \): the number of background pixels that have neighbor black pixels in the two directions (left and up)
The same applies to f9, ..., f14 in six directions left-up, up-right, right-down, down-left, vertical and horizontal.

f15, ..., f20: the baseline dependent features related to the core zone are defined as:

\[ f15 = \frac{CM_{left-up}}{d} ; \]

\( CM_{left-up} \): the number of background pixels in the configuration left-up
The same applies to f16, ..., f20 in six directions left-up, up-right, right-down, down-left, vertical and horizontal.

f21, ..., f28: represent the density of foreground pixels in each vertical column in a frame.

In each frame 28 features vector are extracted, these features are statistical and geometric to integrate both the peculiarities of the text and the pixel distribution characteristics in the word image, which capture the type of strokes (curved, oriented, vertical, and horizontal).

III MODELLING

2.1 Hidden Markov Models

A Hidden Markov Model (HMM) is a doubly stochastic process with an under-lined stochastic process (Markov chain) that is not observable (it is hidden), but can only be observed
through another set of stochastic processes that produce the sequence of observed symbols [Rabiner, 1989]. In order to define an HMM completely, following elements are needed, \( \lambda = (N, M, A, B, \Pi) \):

- \( N \): The number of states of the model,
- \( M \): The number of observation symbols in the alphabet. If the observations are continuous then \( M \) is infinite,
- \( A \): A set of state transition probabilities,
- \( B \): A probability distribution in each of the states,
- \( \Pi \): The initial state distribution.

The use of HMM aims to resolve three problems:
- The Evaluation
  Given an HMM \( \lambda \) and a sequence of observations \( O=o_1,o_2,\ldots,o_T \), what is the probability that the observations are generated by the model, \( P\{O|\lambda\} \)?
- The Decoding
  Given a model \( \lambda \) and a sequence of observations \( O=o_1,o_2,\ldots,o_T \), what is the most likely state sequence in the model that produced the observations?
- The Training
  Given a model \( \lambda \) and a sequence of observations \( O=o_1,o_2,\ldots,o_T \), how should we adjust the model parameters \( \{\Pi, A, B\} \) in order to maximize \( P\{O|\lambda\} \)?

2.2 Character and word models

The used approach is analytical and based on character modeling by HMM. Each character model has a right-left topology and parameters: number of hidden states, state transition probabilities and observation probabilities.

There is no specific theory to set these parameters, so the solution is empirical. Many considerations can be taken into account in setting these parameters and in particular the technique used in the generation of sequences of observations. In our system we used a model with four states for each character with three transitions for each state (Figure 3).

![Figure 3. Character HMM topology](image)

Word model is built by concatenating the appropriate character models (figure 4).

![Figure 4. HMM model for Arabic word: مساحة](image)

IV. EXPERIMENTATIONS AND RESULTS

In order to investigate the potential of our system for offline cursive handwriting recognition, the benchmark database IFN/ENIT is used [Peschwitz 2003], that contains a total of 26459 handwritten words of 946 Tunisian town/villages names written by different writers. We used
the toolbox HTK (Hidden Markov Model Toolkit[2006]) to model the characters and words.

The table below shows the experimental results of our system compared to other recognition systems using the same benchmarking database IFN/ENIT, divided into four sets a, b, c for training and d for testing:

<table>
<thead>
<tr>
<th>System</th>
<th>Models</th>
<th>Recognition rate %</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Kessentini et al 2010]</td>
<td>Multi Stream HMM</td>
<td>79.80</td>
</tr>
<tr>
<td>[Alkhatteeb et al. 2011]</td>
<td>HMM and dynamic Bayesian network</td>
<td>86.73</td>
</tr>
<tr>
<td>[Parvez et al. 2013]</td>
<td>FATF with set medians</td>
<td>79.58</td>
</tr>
<tr>
<td>[A.Maqqor 2014]</td>
<td>Multiple Classifiers</td>
<td>76.54</td>
</tr>
<tr>
<td>Proposed System</td>
<td>HMM</td>
<td>87.93</td>
</tr>
</tbody>
</table>

Table 1 shows the results of recognition rates for various offline systems recognition of cursive Arabic handwritten text using various models and the same database to compare rates and infer the effectiveness of the proposed method.

[Alkhatteeb et al. 2011] are presented a comparative study of approaches for recognizing handwritten Arabic words using Hidden Markov Models (HMM) and Dynamic Bayesian Network (DBN) classifiers, and the recognition rate achieved was 86.73%. [Kessentini et al 2010] [Parvez et al. 2013] [A.Maqqor 2014] are presented systems using respectively multi-stream Hidden Markov Models, Fuzzy Attributed Turning Function (FATF) with set-median; the recognition rate for the results of the systems mentioned does not exceed 80%. Whereasthe proposed system outperforms the results and achieves 87.93%.

Finally, the performance of handwritten Arabic recognition systems is significantly improved using our system based on HMMs. It remains to boost the rate using annexes improvements (Post-Processing: language models)

V. CONCLUSION & PERSPECTIVES

In this paper, we present a recognition system of Arabic cursive handwriting based on hidden Markov models. The extracted features are based on the densities of foreground pixels, concavity and derivative features using sliding window, some of these features depend on baselines estimation. The modelling proposed has improved recognition, and shown encouraging results to be perfect later. Many points are yet to be achieved, firstly modeling a character allows deformations related to its context (next and previous character). To account possible deformations, contextual modeling of characters is opted. The word is no longer seen as a succession of independent characters, but as a sequence of characters in context. Word models are the concatenation of context-dependent characters models: the trigraph, this modeling will allow building more accurate and more efficient models. Taking into account the characters environment allows more precise and more effective models to be built. However, this implies a multiplication of HMM parameters to be learned, it would be the focus of our next work. Then language models will be incorporated to refine and improve the results and lead to a more efficient system.
References


S. Young, al., the HTK Book V3.4 Cambridge University Press, Cambridge UK, 2006


Pechwitz M. and Maergner V., "Baseline estimation for Arabic handwriting words”, in Proc. of the 8th IWFHR 2002, pp 479-484, Canada, August (2002).
