Cursive Arabic Handwriting Recognition: Combination of Multi-HMMs classifiers based on oriented fragmentation

Ramy EL-HAJJ, Chafic MOKBEL and Laurence LIKFORMAN

University of Balamand
Tripoli - Lebanon

ENST PARIS
Paris - France

April 06
• **Introduction**
  • Arabic handwriting recognition.
  • HMM-based classifiers

• **Reference System : HMM-Based**
  • Baseline Detection
  • Feature Extraction
  • HMM-Based Classifier

• **Proposed System : Multi-HMMs Combination**
  • Two-Stages Structure
  • Post-Treatment

• **Experiments and results**

• **Conclusion and perspectives**
Arabic Handwriting recognition

- cursive handwriting
- dots and diacritical marks
- variable and context sensitive shapes
- different writing styles
- writing skew/slant
- inter/intra word spaces, character ligatures of variable length
- character ligatures on the baseline of variable length
HMM–based Arabic handwriting recognition

- Global Approaches
  - one model per word (small lexicon)
  - training from word images

- Analytical Approaches
  - one model per character
  - word models: concatenation of character models
  - joint character segmentation and recognition
  - segmentation
Reference system Diagram
Baseline Detection

- lower and upper baselines
- based on the result of projection on the vertical axis
- from baseline position: extraction of baseline dependent features (presence of descenders and ascenders)
Feature Extraction

- **sliding window technique**
  - vertical overlapping windows (frames)
  - window width fixed (system parameter)
  - window height depends on each word image.
  - a feature vector is calculated for each frame

*Word image divided into vertical frames (here without overlap)*
Feature Extraction

- Each frame is divided into cells
- We extract on each frame 26 features
- Two types of features are considered:
  - **Distribution features**: based on foreground (black) pixel densities
  - **Concavity features**: based on foreground pixel configurations
  - Derivative is added as complementary feature
- For each type of feature, a subset is baseline dependent
Distribution features

• $f_1$ is the density of foreground (black) pixels:

$$f_1 = \sum_{i=1}^{n_c} n(i)$$

- $n(i)$: number of foreground pixels in the $i$th cell,
- $n_c$: number of cells

• $f_2$ represents the number of transitions between two consecutive cells of different density levels:

$$f_2 = \sum_{i=2}^{n_c} |b(i) - b(i-1)|$$

- $b(i) = 0$ if $n(i) = 0$ else $b(i)=1$. 
Distribution features

- $f_4$ to $f_{11}$: 8 features that represent the densities of foreground pixels for each vertical column of pixels in a frame:
Distribution features

- $f_{12}$ is the vertical position of the center of gravity of the foreground pixels in a frame with respect to the lower baseline

\[
f_{12} = \frac{g - L}{H}
\]

- $f_{13}$ and $f_{14}$: The density of foreground pixels over and under the lower baseline in a frame

\[
f_{13} = \frac{\sum_{j=L+1}^{H} r(j)}{H \cdot w}
\]

\[
f_{14} = \frac{\sum_{j=1}^{L-1} r(j)}{H \cdot w}
\]
Distribution features

• $f_{15}$: the number of transitions between two consecutive cells of different density levels above the lower baseline

$$f_{15} = \sum_{i=k}^{n} |b(i) - b(i-1)|$$

• $k$: is the cell that contains the lower baseline

• $f_{16}$: represents the zone to which the gravity center of the foreground pixels belongs with respect to baselines
Derivative feature

• $f_3$: difference in the vertical position of gravity centers of foreground pixels in the current frame $t$ and in the previous one

$$f_3 = g(t) - g(t - 1)$$

$$g = \frac{\sum_{j=1}^{H} j \cdot r(j)}{\sum_{j=1}^{H} r(j)}$$
Concavity features

- provide local concavity information and stroke direction
- 5 configuration types, 3x3 window
- Each normalized concavity feature:

  \[ f_{17-21} = \frac{\text{#white pixels (background) in configuration}}{H} \]

  with \( H \) : height of the frame
Concavity features

f22 – f26 = \#white pixels in configuration in the corezone
\[ d \]

with \( d \) : height of the corezone
HMM-Based Classifier: character modeling

- Analytical approach
- Parameters (topology, number of states, transitions, number of Gaussians) obtained by cross validation
- 4 states, right-left topology
- Mixture of 3 Gaussian distributions with diagonal covariance matrices for each state
- 159 character models
- Models for characters with additional marks
HMM-Based Classifier: word modeling

- concatenating appropriate character models
- space models

The HMM model for the Arabic word "ما رتب"
HMM-Based Classifier

General purpose HMM classifier: The HCM full toolkit

- Compile HMM networks
- Training/Adaptation
- Recognition with beam search

HCM has a text definition C-like interface and produces (fixed point) models

Training:
- Estimates character-HMM parameters on the training set of the database.
- Segmental Expectation Maximization (EM) algorithm is used.

- The EM algorithm consists of iteratively estimating the model parameters:
  - Each iteration is formed of two steps:
    - **Expectation E-step:** Data completion based on the model parameters.
    - **Maximization M-step:** Estimate the model parameters using the completed data
Recognition

- Network of lexicon entries formed of character models
- Words are described in terms of characters constituents
- The character sequence providing the maximum likelihood identifies the recognized entry
- The Viterbi algorithm is used
  - HCM provides N-best solutions
Experiments: Reference System

- Benchmark database IFN/ENIT
  - Lexicon: 946 Tunisian town/village handwritten names
  - 26459 word images
  - divided into four subsets a, b, c, and d

- Two feature sets:
  - Set Fw: 16 features baseline-independent
  - Set Fb: 26 features (baseline dependent+independent)

- Cross validation methodology: training on 3 sets (among a,b,c,d), recognition on the remaining set
# Results: Reference System

## The Complete Database

<table>
<thead>
<tr>
<th>Test</th>
<th>Training data sets</th>
<th>Test data set</th>
<th>Rec. rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>b, c, d</td>
<td>a</td>
<td>83.04 %</td>
</tr>
<tr>
<td>2</td>
<td>a, c, d</td>
<td>b</td>
<td>83.10 %</td>
</tr>
<tr>
<td>3</td>
<td>a, b, d</td>
<td>c</td>
<td>82.47 %</td>
</tr>
<tr>
<td>4</td>
<td>a, b, c</td>
<td>d</td>
<td>83.31 %</td>
</tr>
</tbody>
</table>

Classification Problems - Examples

- Vertical ligatures

- Descenders and Ascenders overlap

- Diacritical mark’s position
Proposed system

General Architecture of proposed System
Proposed system

- Integration of information about possible inclination in handwritten
- Vertical and slanted (Oriented) frames
- Three angles of inclination (-α  0  +α)

![Diagram showing angles of inclination.](image)
Proposed system

- For each inclination, an HMM-Based classifier proposes a Top-4 list of candidates

- Post-Treatment: Voting module
  - Fusion lists of candidates proposed by HMM-Recognizer
  - Re-arrangement of candidates
Proposed system

List 1

List 2

List 3

scores
dakhaniyah 15.83
bata 15.51
shaba 15.29
kana 14.71

scores
kana 17.13
bata 16.96
shaba 16.87
dakhaniyah 16.45

scores
kana 18.68
bata 17.74
shaba 17.22
dakhaniyah 17.01

scores de combinaison
bata 51.15
kana 49.58
shaba 49.17
dakhaniyah 32.28
## Evaluation and Results: Proposed System

- Data base: subset of 21480 images
- Dictionary: 459 Classes (town name)

<table>
<thead>
<tr>
<th>Test Data</th>
<th>Recognizer</th>
<th>Recognition Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Top1</td>
<td>Top2</td>
</tr>
<tr>
<td>Sub set</td>
<td>Size</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>5352</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Recognizer</th>
<th>Top1</th>
<th>Top2</th>
<th>Top3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference System</td>
<td>88.85</td>
<td>91.95</td>
<td>93.32</td>
</tr>
<tr>
<td>Proposed System</td>
<td>90.58</td>
<td>94.46</td>
<td>95.38</td>
</tr>
</tbody>
</table>
Conclusion

- Reference System for automatic Cursive Handwritten texts recognition.
  - Analytical Approach, Segmentation-free, HMM-Based
  - Robust Features (language independent)
  - ICDAR 2005 Arabic Handwritten Recognition Competition,

- Proposed System
  - Two-stages structure: Recognition – Post-Treatment
  - Oriented Fragmentation
  - Performance Enhancement
Perspectives

- Many points are yet to be achieved such as
  - Apply the new method on the whole IFN/ENIT Database
  - Examination of new Primitives
  - Enhancement of post-treatment stage:
    - Apply the independent features on a Latin handwriting database
References


References


References


