

Husni A. Al-Muhtaseb, Sabri A. Mahmoud, and Rami S. Qahwaji,  
"Recognition of Off-line printed Arabic text Using Hidden Markov Models",  
Signal Processing, Volume 88, Issue 12, December 2008, pp. 2902-2912.

**Keywords:** Arabic Text Recognition, Hidden Markov Models, Feature Extraction, Omni Font Recognition



Contents lists available at ScienceDirect

Signal Processing

journal homepage: [www.elsevier.com/locate/sigpro](http://www.elsevier.com/locate/sigpro)



## Recognition of off-line printed Arabic text using Hidden Markov Models

Husni A. Al-Muhtaseb<sup>a,\*</sup>, Sabri A. Mahmoud<sup>a</sup>, Rami S. Qahwaji<sup>b</sup>

<sup>a</sup> Information and Computer Science Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

<sup>b</sup> Electronic Imaging and Media Communications Department, University of Bradford, Bradford, UK

### ARTICLE INFO

#### Article history:

Received 1 January 2008

Received in revised form

12 June 2008

Accepted 16 June 2008

Available online 27 June 2008

#### Keywords:

Arabic text recognition

Hidden Markov Models

Feature extraction

Omni font recognition

### ABSTRACT

This paper describes a technique for automatic recognition of off-line printed Arabic text using Hidden Markov Models. In this work different sizes of overlapping and non-overlapping hierarchical windows are used to generate 16 features from each vertical sliding strip. Eight different Arabic fonts were used for testing (viz. Arial, Tahoma, Akhbar, Thuluth, Naskh, Simplified Arabic, Andalus, and Traditional Arabic). It was experimentally proven that different fonts have their highest recognition rates at different numbers of states (5 or 7) and codebook sizes (128 or 256).

Arabic text is cursive, and each character may have up to four different shapes based on its location in a word. This research work considered each shape as a different class, resulting in a total of 126 classes (compared to 28 Arabic letters). The achieved average recognition rates were between 98.08% and 99.89% for the eight experimental fonts.

The main contributions of this work are the novel hierarchical sliding window technique using only 16 features for each sliding window, considering each shape of Arabic characters as a separate class, bypassing the need for segmenting Arabic text, and its applicability to other languages.

© 2008 Elsevier B.V. All rights reserved.