AIN SHAMS UNIVERSITY

FACULTY OF ENGINEERING

Computer Engineering and Systems Department

Offline Handwritten Arabic Text Recognition using Hidden Markov Models

A Thesis submitted in partial fulfillment of the requirements of a Master of Science degree in Electrical Engineering Computer Engineering and Systems Department

by

Ahmed Hussein Sayed Metwally

Bachelor of Science degree in Electrical Engineering

Computer Engineering and Systems Department

Faculty of Engineering, Ain Shams University, 2013

Supervised By

Prof. Dr. Hazem Mahmoud Abbas

Dr. Mahmoud Ibrahim Khalil

Cairo, 2019

AIN SHAMS UNIVERSITY

FACULTY OF ENGINEERING

Computer Engineering and Systems Department

Offline Handwritten Arabic Text Recognition using Hidden Markov Models

By

Ahmed Hussein Sayed Metwally Bachelor of Science in Electrical Engineering Computer Engineering and Systems Department Faculty of Engineering, Ain Shams University, 2013

Examiners' Committee:

Title, Name and Affiliation	Signature
Prof. Dr. Mohsen Abd Elrazek Rashwan Faculty of Engineering, Cairo University	.....
Prof. Dr. Hussein Ismael Shaheen Faculty of Engineering, Ain Shams University	.....
Prof. Dr. Hazem Mahmoud Abbas Faculty of Engineering, Ain Shams University	.....
Dr. Mahmoud Ibrahim Khalil Faculty of Engineering, Ain Shams University	.....

Statement

This thesis is submitted as a partial fulfillment of Master of Science in Electrical Engineering, Faculty of Engineering, Ain shams University. The author carried out the work included in this thesis, and no part of it has been submitted for a degree or a qualification at any other scientific entity.

Ahmed Hussein Sayed Metwally

Signature:

Date:

Researcher Data

Name: Ahmed Hussein Sayed Metwally

Date of Birth: 17/05/1991

Place of Birth: Kingdom Saudi Arabia

Last academic degree: Bachelor of Science Field of specialization: Electrical Engineering University issued the degree : Ain Shams University Date of issued degree : 2013

THIS PAGE INTENTIONALLY LEFT BLANK

Abstract

The complexity of Arabic Letters lies in the fact that each letter in each different position (start, end, middle, and isolate) is represented with a different shape and style of writing. But the similarity occurs between different letters in different positions, which raises conflict when it comes to recognizers and classifiers. Those facts, in addition to the cursive nature of Arabic language, along with variations in writing style, methods, and fonts, makes Arabic Handwriting a fairly complex task to perform.

In this research, a new approach to Handwriting recognition is introduced. The method involves the training of a separate HMM for every Arabic letter in the alphabet in each of their various positions. The method followed is based on diacritics removal, along with similar letter grouping, to reduce the total number of models to be trained to improve the overall efficiency of the system. The system also uses a set of hybrid high performance features to help the HMM identify the main characteristics of each letter and help identify the differences between them. And finally after HMM has performed its recognition phase, the post-processing algorithm is used to improve on the recognition

rate of the HMM, reaching a recognition rate of around 87%. The proposed system is trained and tested using the IFN/ENIT database which contains tens of thousands of images handwritten in Arabic from different writers to provide the needed variation for improved training process and non-biased recognition.

Keywords: Arabic Handwriting Recognition, Hidden Markov Models, structural features, Concavity Features, Distribution Features, Post-processing.

Acknowledgment

All praise is due to Allah, Most Merciful, the Lord of the Worlds, Who taught man what he knew not. I would like to thank God Almighty for bestowing upon me the strength and patience to finalize this research.

I wish to express my thankfulness and gratitude to my supervisors, Prof. Dr. Hazem Abbas and Dr. Mahmoud Khalil for their patient guidance, encouragement, insightful thoughts and useful discussions. I feel extremely lucky to have supervisors who cared so much about my work, who were always helpful when it comes to discussions, responding to my questions, and always guiding me in the correct path. Thank you for being role models for me and thank you again for your guidance and follow up without which this work wouldn't have seen the light.

I would also like to thank my family for their prayers, support and continuous encouragement. Thank you for always pushing me to continuously do my best.

Thanks to all of my friends, and co-workers for their support and encouragement.

Finally, I want to thank the Examiners' Committee for their time and effort while reviewing this thesis. And for providing suggestions to further help improve the overall content and presentation of this thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents