بسم الله الرحمن الرحيم

شبكة المعلومـــــات الجامعية

التوثيق الالكتروني والميكروفيلم

شبكة المعلومات الجامعية

جامعة عين شمس

التوثيق الالكتروني والميكروفيلم

﴿ قـــســـم ﴾

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأفلام قد اعدت دون آية تغيرات

يجب أن

تحفظ هذه الأفلام بعيداً عن الغبار

في درجة حرارة من 15 – 20 مئوية ورطوبة نسبية من 20-40 %

To be kept away from dust in dry cool place of
15 – 25c and relative humidity 20-40 %

بعض الوثائق الأصلية تالفة

بالرسالة صفحات

لم ترد بالأصـــل

Mansoura University
Faculty of Engineering
Computer and Systems Engineering Department

Intelligent Expert System
for
Articulate Arabic Text Machine Reader

A Thesis
Submitted for Partial Fulfillment of the
Degree of Master of Science
in
Computer and Systems Engineering

Submitted By

Mohamed Fathi Hamed ALRahmawy

B.Sc., Electronics Engineering
Mansoura University, 1997
Research student at Computer and Systems Engineering Department
Mansoura University

Supervised By

Prof. Dr. Ali Ibrahim ALDesouki

Faculty of Engineering
Computers & Systems Engineering Department
Mansoura University, Mansoura

Dr. Hesham Arafat Ali

Faculty of Computers and Information
Information Systems Department
Mansoura University, Mansoura

Mansoura, 2001

Mansoura University
Faculty of Engineering
Computer and Systems Engineering Department

Intelligent Expert System
for
Articulate Arabic Text Machine Reader

A Thesis
Submitted for Partial Fulfillment of the
Degree of Master of Science
in
Computer and Systems Engineering

Submitted By

Mohamed Fathi Hamed ALRahmawy

B.Sc., Electronics Engineering
Mansoura University, 1997
Research student at Computer and Systems Engineering Department
Mansoura University

Supervised By

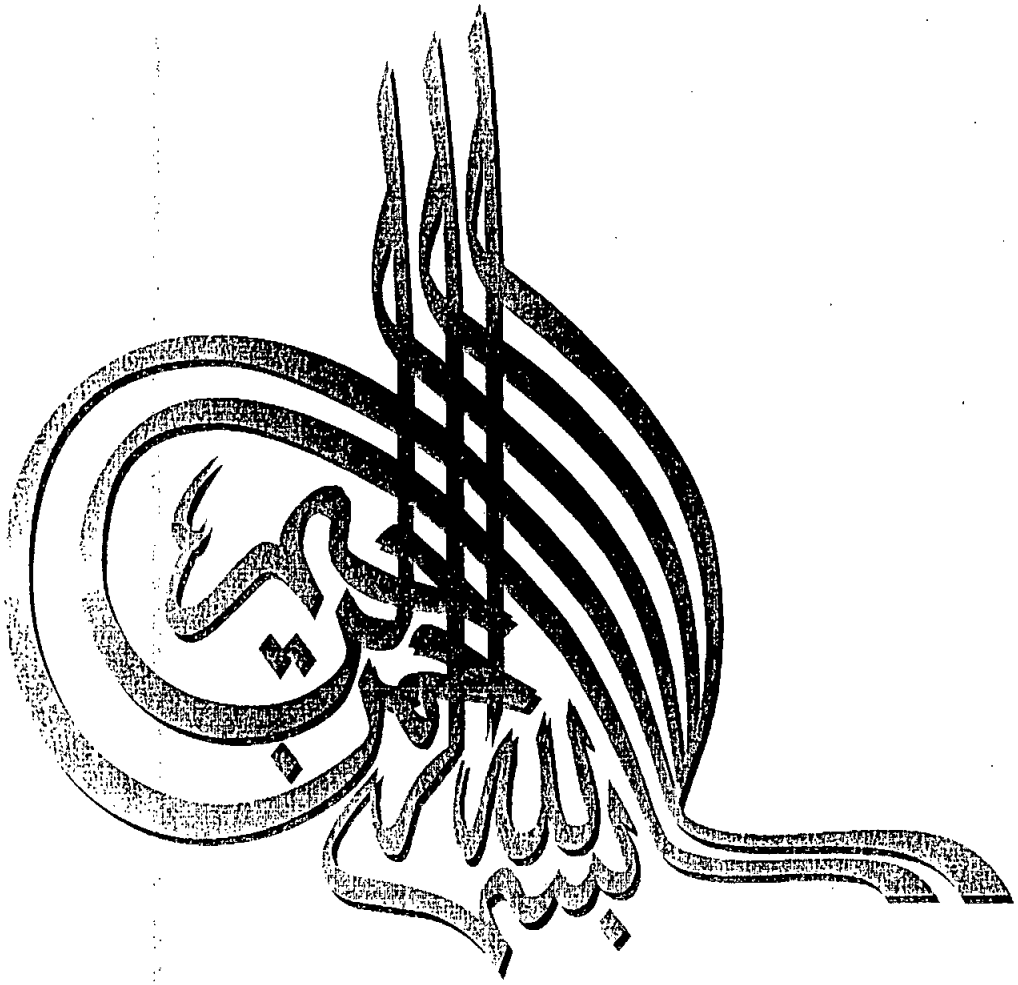
***Prof. Dr.* Ali Ibrahim ALDesouki**

Faculty of Engineering
Computers & Systems Engineering Department
Mansoura University, Mansoura

***Dr.* Hesham Arafat Ali**

Faculty of Computers and Information
Information Systems Department
Mansoura University, Mansoura

Mansoura, 2001


بسم الله الرحمن الرحيم

Supervisors

***Dissertation Name*:**

Intelligent Expert System

for

Articulate Arabic Text Machine Reader

***Researcher Name*:**

Eng: Mohammad Fathy Hamed AL-Rahmawy

Supervisor	Position
Pof. Dr. Ali Ibrahim ALDesooki	Professor at Computers & Systems Dept. Faculty of Engineering Mansoura University
Dr. Hesham Arafat Ali	Assist Prof. at Information Systems Dept. Faculty of Computers & Information Systems Mansoura University

Supervisor	Signature
Pof.Dr. Ali Ibrahim ALDesooki	
Dr. Hesham Arafat Ali	Hesham Arafat A L.

Abstract

Mohammed Fathi Hamed ALRahmawy

Intelligent-Expert System

for

Articulate Arabic Text Machine Reader

Master of Science Dissertation
Mansoura University, 2001

This thesis aims mainly to build a fast and efficient Arabic OCR system using object-oriented programming technology in order to build an Articulating machine reader for the printed Arabic writing.

Hence, the basic difficulties and the different characteristics of the Arabic text recognition problem are outlined. Then, different stages of OCR systems are reviewed and the basic approaches used in each stage are studied and the previous work in Arabic OCR is reviewed.

Also, basic concepts of Neural Networks and its most general models are reviewed. Then, a summary of back propagation algorithm for learning is presented as the most widely learning algorithm. The virtues and limitations of Back-Propagation Learning are studied.

Next, The details of the algorithms used in implementing the proposed system are studied. Where, novel algorithms for document preprocessing and extracting the chain-coded inner and outer contours of the subwords of the document and representing them as objects are presented. Then, locations of these objects are analyzed in order to segment them in separate lines. Also, a new Arabic text segmentation algorithm is presented for segmenting the chain coded contours of the Arabic subwords into chain coded objects of the primitives of the characters (sub-characters) constituting these subwords. Then, a novel fast and efficient algorithm for extracting the central-moments features out of the chain-coded upper and lower outer contours of the segmented primitives is used in order to improve the feature extraction rate.

For the classification of the primitives, a two-stage hybrid recognition system is implemented for the classification of the segmented primitives. The hybrid system uses two Neural networks in its first stage. The used neural networks are embedded within the system as objects and linked with its objects for clustering the primitives to be recognized into one of the predefined clusters. Then, in the second stage a set of classifiers (one classifier for each cluster) that use statistical, structural and heuristic rules of Arabic writing are implemented for the final classification of the primitives and building the characters.

Also, a novel method for recombining the recognized sub-words into words using in-between spaces and language rules is presented.

Finally, An Arabic word-based articulation sub-system is presented for articulating either the recognized text or simply from a text file.

Keywords: *OCR, Pattern Recognition, Neural networks, Image Processing, Over Segmentation, preprocessing, Chain Coding, Moments, Object Oriented Programming, Contour Processing, Hybrid Recognition System.*

Dedication

I would like to present this simple work to my mother for her great effort and patience in bringing me up and for her moral support in hard times of finishing this work. I'd like also to dedicate this work to the soul of my father who taught me to always put all my dependence on **Allah** and to **Allah**, and how to always keep trust in myself. Finally, I present this work also to my sister and to her two little children Yara and Yousif for their encouragement.

Acknowledgements

Firstly and lastly, I thank **Allah** for helping and guiding me in completing this thesis.

Then, I would like to express my great gratitude to Prof. Dr. Ali ELDesouki for his great support, advises, revising the thesis and for his long patience with me.

I also thank Dr. Hesham Arafat for his thankworthy effort and for his valuable advises and for all the materials and references he gave to me.

I would like to thank my fellows and the staff of both the Computer and System Engineering Department and Mansoura Univ. Computer Center for their thankworthy support and encouragement.

I would like also to thank all my friends with no exception for their encouragement.

Finally, I express my deep thanks and great gratitude to my family, especially my mother, for their great patience and moral support during the most important stages of the thesis.

CONTENTS