

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGEMENT	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	xi
	LIST OF FIGURES	xiii
	LIST OF ABBREVIATIONS	xv
	LIST OF APPENDICES	xvi
1	INTRODUCTION	
	1.1 Problem Background	1
	1.2 Statement of the Problem	3
	1.3 Research Aim	4
	1.4 Objectives	4
	1.5 Research Scope	5
	1.6 Significance of the Research	6
	1.7 Organization of the Report	6
2	LITERATURE REVIEW	
	2.1 Introduction	7
	2.2 Common Characteristics of Arabic Script	10

2.3	Character Recognition	15
2.4	Arabic Character Recognition Stages	16
2.4.1	Preprocessing and Representation	18
2.4.1.1	Vertical and Horizontal Projection	18
2.4.1.2	Thinning (skeleton extraction)	22
2.4.1.3	Contour Tracing	24
2.4.1.4	Baseline Detection	26
2.4.2	Segmentation	26
2.4.3	Feature Extraction	29
2.5	Arabic Character Recognition Techniques	30
2.5.1	Neural Network Technique	31
2.5.2	Hidden Markov Models (HMMs)	32
2.5.3	Genetic Algorithm	34
2.6	Discussion	38
2.7	Summary	40
3	METHODOLOGY	
3.1	Introduction	45
3.2	Operational Framework	46
3.2.1	Phase 1: Planning Phase and Literature Review	47
3.2.2	Collecting Data	47
3.2.3	Data Analysis	49
3.2.4	Storing Resulting Samples in the Database	50
3.2.5	Preprocessing for the Image	52
3.2.6	Representation	52
3.2.7	Feature extraction	54
3.2.8	GA	60
3.2.8.1	The fitness function	60
3.2.8.2	Selection reproduction operator	62
3.2.8.3	Crossover operator	62
3.2.8.4	Mutation operator	63

	3.2.8.5 Peak concatenation function	64
	3.2.9 Programming of Techniques	64
	3.2.10 Evaluation	65
	3.3 Instrumentation	65
	3.4 Summary	67
4	EXPERIMENTAL RESULTS AND DISCUSSION	
4.1	Introduction	68
4.2	Available Arabic OCR tools	68
4.3	Representation Results and Analysis	71
	4.3.1 Results of Building the Database	72
	4.3.2 Segmentation Performance	73
	4.3.2.1 Results of Segmenting the Image into Lines	73
	4.3.2.2 Results of Segmenting the Lines to Words or Sub-words	76
	4.3.2.3 Results of Segmenting Words to Peaks or Characters	77
	4.3.3 Results of Thinning the Contour	80
	4.3.4 Results of Detecting the Baseline	81
4.4	Results of the Recognition Algorithm	82
	4.4.1 Feature Extraction Results	82
	4.4.2 Genetic Operation Results	86
4.5	Comparing our Algorithm with Other Algorithms	92
	4.5.1 GA Systems	92
	4.5.2 ICDAR 2009 Arabic Handwriting Recognition Competition	93
4.6	Discussion	95
4.7	Summary	97
5	CONCLUSION	
5.1	Introduction	98
5.2	Findings	99
5.3	Contributions of the Study	99
5.4	Future Works	100

5.5	Summary	102
	BIBLIOGRAPHY	103
	APPENDICES A-F	111 – 130

LIST OF TABLES

TABLE NO	TITLE	PAGE
2.1	Comparison of different scripts (Amin, 1997)	10
2.2	Arabic alphabet shapes	12
2.3	Components of OCR recognition	11
2.4	Summary of GAs in character recognition	39
2.5	Summary of the literature review on AOCR techniques; Where H stand for Horizontal projection and V stand for Vertical projection	41
3.1	The groups of writers	48
3.2	Number of words, sub-words, and letters in one form	51
3.3	Number of words, sub-words, and letters for the writers in each group	51
3.4	Comparison between the proposed database and AHD/USM	52
3.5	Fitness function of the character ن	61
3.6	Fitness function of the character م	61
3.7	List of software needed	66
3.8	List of hardware used	66
4.1	Readiri Pro 11's recognition of printed text	69
4.2	Numbers of lines images resulted from segment the paragraphs	75
4.3	Accuracy rate of segmentation method	78

4.4	Examples of segmented and over-segmented words	79
4.5	Examples of unsegmented words	79
4.6	Examples of the poor preprocessing experiment in the feature extraction stage.	83
4.7	Examples of the excellent preprocessing experiment in the feature extraction stage.	84
4.8	Examples of the for all experiment in the feature extraction stage.	85
4.9	Results from all experiments in the feature extraction stage	85
4.10	Examples of poor preprocessing experiment in the recognition stage	87
4.11	Examples of the excellent preprocessing experiment in the recognition stage.	88
4.12	Examples of All experiments in the recognition stage	89
4.13	Results of all experiments in the recognition stage	91
4.14	Comparing the GA systems	93
4.15	Comparing our system with the results of the ICDAR 2009 competition	94

LIST OF FIGURES

FIGURE NO	TITLE	PAGE
1.1	Structure of a single population evolutionary algorithm	4
2.1	Arabic characters differing with dots or hamza	13
2.2	Example of Arabic sub-words	14
2.3	Dissimilar styles and fonts for Arabic wording	15
2.4	Arabic character recognition stages	17
2.5	Horizontal and vertical projections (Mohammed, 2006)	19
2.6	A word image and its skeleton	22
2.7	Genetic algorithm	35
3.1	Operational Framework	46
3.2	Printed text paragraph	50
3.3	Features of words, sub-words, and characters	55
3.4	Detecting the baseline by the horizontal projection profile: (a) handwritten word, (b) horizontal projection profile to detect the baseline.	55
3.5	Feature vector representing the character ـ	57
3.6	Feature extraction algorithm	59
3.7	Feature vectors of ـ and ن	61
3.8	Example of reproduction operator	62
3.9	Example of crossover operator	63
3.10	Example of mutation operator	63
4.1	Readiris Pro 11 Recognition of printed text	70

4.2	Printed text paragraph	70
4.3	Reorganization of Readiri Pro 11 from Figure 4.2	71
4.4	The collected characters, words, sub-words, and paragraphs in our database from handwriting and printing text.	72
4.5	Segmenting the paragraph into lines by horizontal projection profile: (a) printed text paragraph, (b) horizontal projection profile used to segment the paragraph into lines.	74
4.6	Segment the paragraph to line, (a) the horizontal projection profile of the paragraph image ,(b) the paragraph image ,(c) image of first line segmented from paragraph ,(d) image of second line segmented from paragraph ,(e) image of third line segmented from paragraph, (f) image of fourth line segmented from paragraph	75
4.7	Segmentation of the line to words and sub-words by the vertical projection profile: (a) printed text line segmented from paragraph, (b) horizontal projection profile of the segmented line, (c) vertical projection profile of the line segmented into words and sub-words.	76
4.8	Success of the segmentation method	78
4.9	Clusters in the skeleton: (a) the main image, (b) the skeleton image.	80
4.10	Results from all experiments in the feature extraction stage	86
4.11	Relation between population size and recognition rate in the excellent preprocessing experiment.	91

LIST OF ABBREVIATIONS

ACR	-	Arabic Character Recognition
BMP	-	Bitmap Picture
CEDR	-	Center of Excellence for Document Analysis and Recognition
CR	-	Character Recognition
GAs	-	Genetic Algorithms
HWR	-	Handwriting Recognition
OCR	-	Optical Character Recognition
OACR	-	Optical Arabic Character Recognition
UR	-	UnRecognized
R	-	Recognized
US	-	UnSegmented

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	Form	111
B	Form of one writer	116
C	Examples of paragraph images	118
D	Examples of Experimentation of Feature Extraction Algorithm	119
E	Examples of Experimentation of Genetic Algorithms' Recognition	123
F	Time Table	129