

**Universiti Teknologi MARA**

**DEVELOPMENT OF TEXT  
EXTRACTION AND RECOGNITION  
PROTOTYPE USING ADAPTIVE  
RESONANCE THEORY 1 (ART1)  
NEURAL NETWORK**

**IFFARINI BINTI IDRIS**

Thesis submitted in fulfillment of the requirements for  
**Bachelor of Science (Hons) Intelligent Systems**  
**Faculty of Information Technology And**  
**Quantitative Sciences**

MAY 2008

## **ACKNOWLEDGEMENT**

**By the name of Allah, Most Gracious and Most Merciful**

Alhamdulillah, praise and thank to Allah because of His Almighty and His utmost blessing, I was able to finish my final year project within the time duration given.

This research would not have been completed without the help, support, encouragement and involvement from many people. Here I would like to express my gratitude to those who have helped me, either directly or indirectly.

Firstly, a very special thanks to my supervisor, Assoc. Prof. Zaidah Ibrahim for her help and advice throughout the whole process in making this research a success. Her guidance and wise supervision have benefited me greatly.

Secondly, I would like to thank all my lecturers in FTMSK, and to my dedicated lecturer for Intelligent System Project (ITS690), Miss Norzaidah Mohd Noh for the ideas, guidance and support in completing this research.

A lot of thanks to my lovely family for their love, support, understanding, trust and advice that they have given. To all my friends especially Noor Amera Azliza Nor Azman for their help, support and advice, thank you very much.

## **TABLE OF CONTENTS**

<b>CONTENT</b>	<b>PAGE</b>
DECLARATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ALGORITHMS	xi
LIST OF EQUATIONS	xii
ABSTRACT	xiii

### **CHAPTER ONE: INTRODUCTION**

1.0	Introduction	1
1.1	Project Overview	1
1.2	Problem Statement	2
1.3	Project Objective	3
1.4	Project Scope	4
1.5	Project Significant	4
1.6	Conclusion	5

### **CHAPTER 2: LITERATURE REVIEW**

2.0	Introduction	6
2.1	Document Processing	6
2.2	Document Processing Task	8
2.2.1	Pre-Processing	8
2.2.2	Page Layout Analysis	8
2.2.2.1	Text Image Restoration	10

2.2.2.2	Skew Detection	11
2.2.3	Text Extraction	11
2.2.3.1	Image Digitization	12
2.2.4	Character Recognition	13
2.2.4.1	Identification of touching character	14
2.2.4.2	Location of cutting points	15
2.3	Character Recognition	15
2.4	Neural Network and Character Recognition	18
2.4.1	Artificial Neuron	19
2.5	Techniques used by other researchers	20
2.5.1	Text Extraction and Recognition Techniques	20
2.5.2	Character Segmentation Techniques	23
2.5.3	Backpropagation Neural Network	24
2.6	Techniques Used	26
2.6.1	Mathematical Morphological Technique	26
2.6.2	Segmenting using Labeling Technique	27
2.6.3	Adaptive Resonance Theory 1 (ART1)	27
2.6.4	The Network for ART1	28
2.7	Application of Text Extraction and Recognition	30
2.8	Conclusion	31

### **CHAPTER 3: METHODOLOGY**

3.0	Introduction	33
3.1	Problem Assessment and Research Study	37
3.2	Knowledge Acquisition	37
3.2.1	Literature Search	38
3.2.2	Data Collection	39
3.3	Prototype Design	41
3.3.1	Design Interface for the Prototype	41

## **ABSTRACT**

Character recognition system can contribute tremendously towards the advancement of automation process and can be useful in many other applications such as Data Entry, Document Processing and Cheque Verification. In this research, a prototype of text extraction and recognition using Adaptive Resonance Theory 1 (ART1) was proposed. For this project, several sets of images were collected from magazines and text books. In prototype design, the interface of ART1 and the ART1 neural network architecture were designed. The pre-process part of this prototype was developed using MATLAB and the recognition part was developed using C++. During the pre-processing stage, images were converted to binary image. Then, the title of the document images was extracted using Mathematical Morphological technique and the characters were segmented using labeling technique. After the pre-processing stage, each of the pixels value that represent the character will be the input to the ART1 network for character recognition process. ART1 neural network has proven to give good performance with 65.7 % recognition rate. A comparative study was conducted between ART1 and Backpropagation Neural Network (BPNN) to compare their recognition performances. BPNN is unable to meet the performance goal because of insufficient number of training data. In conclusion, ART1 is better than BPNN when the number of training data is small.