

**UNIVERSITI TEKNOLOGI MARA**

**JAWI SUB-WORD RECOGNITION  
SYSTEM USING WINDOW-BASED  
SEGMENTATION-FREE APPROACH**

**ROSLIM BIN MOHAMAD**

Thesis submitted in fulfillment  
of the requirements for the degree of  
**Doctor of Philosophy**

**Faculty of Computer and Mathematical Sciences**

July 2017

## ABSTRACT

The presence of features such as cursive, diversity of writing styles and sizes of characters in a Jawi text, ligature, and vertical overlapping make the recognition of Jawi handwritten text to be difficult. For the recognition system based on model development using windows, the presence of such features cause the resulting model to be less consistent, where it produces a different sequence of primitive structures of words/sub-words from the same lexicon. To overcome the inconsistency model problem, a handwritten Jawi text recognition system based on a sub-word model has been developed. The proposed modeling technique which is known as Selection Segmentation-Free (SSF) separates core and connection structure of a sub-word into a different window. The resulting window will go through a selection process to determine the windows that will be used to represent the sub-word model. In order to increase accuracy and efficiency of the representation feature, two categories of features which are known as primary and secondary features were extracted from each of the selected windows. Primary feature were extracted using Window Code Representation (WCR) technique from main structure. Secondary feature for supporting the primary feature were extracted from dot and main structure. For the experiment purposes, a total of 1200 sub-words of 80 lexicons were used. Each lexicon is randomly selected and divided into three sets. Three experiments to evaluate the performance of SSF, WCR and combination of primary and secondary feature techniques were conducted. The three techniques are combined to represent the proposed system and compared with the comparison system introduced by Remon (2009). Comparison result shows that the recognition rate of proposed system (84.8%) is better than comparison system (79.1%).

## ACKNOWLEDGEMEN

Alhamdulillah, First of all I would like to thank the Almighty for being able to complete this thesis successfully.

A million thanks to Prof. Dr. Hj. Mazani bin Hj. Manaf who acts as my main supervisor for this Doctor of Philosophy (Ph.D) study. Also to both my co-supervisors, Dr. Rose Hafsah Abd. Manaf and Assoc. Prof. Dr. Muhammad Faidzul Nasruddin for their support and guidance in completing this thesis.

I specially thank to my parents and family whose encouragement and support has been a constant source of inspiration during completing this thesis.

Thank you very much

## TABLE OF CONTENTS

	<b>Page</b>
<b>CONFIRMATION BY PANEL OF EXAMINERS</b>	i
<b>AUTHOR'S DECLARATION</b>	ii
<b>ABSTRACT</b>	iii
<b>ACKNOWLEDGEMENT</b>	iv
<b>TABLE OF CONTENTS</b>	v
<b>LIST OF TABLES</b>	x
<b>LIST OF FIGURES</b>	xi
<b>LIST OF ABBREVIATION</b>	xvi
<b>CHAPTER ONE : INTRODUCTION</b>	
1.1 Introduction	1
1.2 Research Background	3
1.3 Problem Statement	6
1.4 Research Questions	7
1.5 Research Objectives	8
1.6 Research Scope	9
1.7 Research Significances	8
1.8 Thesis Outline	10
1.9 Summary	11
<b>CHAPTER TWO : LITERTURE REVIEW</b>	
2.1 Introduction	13
2.2 Development of an Offline Handwritten Text Recognition System	14
2.2.1 Preprocessing	16
2.2.2 Segmentation	19
2.2.2.1 Segmentation-based methods	20

2.2.2.2	Segmentation-free methods	24
2.2.3	Feature extraction	29
2.2.4	Classification	32
2.2.4.1	Multilayer Perceptron	34
2.2.4.2	Random Forest	36
2.2.4.3	Waikato Environment for Knowledge Analysis (WEKA)	37
2.3	Jawi Writing Characteristics	39
2.4	The Development of Offline Jawi Handwriting Recognition System	43
2.5	Implicit Segmentation and Feature Extraction Technique were studied	51
2.5.1	Implicit segmentation	52
2.5.2	Feature extraction	53
2.7	Summary	54

## **CHAPTER THREE : RESEARCH METHODOLOGY**

3.1	Introduction	56
3.2	Research Conceptual Model	56
3.3	Proposed Research Framework	58
3.3.1	Phase 1: Preliminary Study	60
3.3.2	Phase 2: Data Collection	60
3.3.2.1	Database Background	61
3.3.2.2	Data Selection	62
3.3.2.3	Extra Data Collection	64
3.3.2.4	Data Collection for Experiment	64
3.3.3	Phase 3: Algorithm Design and Construction.	65
3.3.3.1	Preprocessing	66
3.3.3.2	Implicit Segmentation	69
3.3.3.3	Feature extraction	69
3.3.3.4	Classification	71
3.3.4	Phase 4: Performance Analysis and Result Analysis	71
3.3.4.1	Experiment I (E1)	72