JAWI RECOGNITION SYSTEM

NUR AZIELA BINTI MANSOR

This thesis is submitted as partial fulfillment of the requirements for the award of the Bachelor of Electrical Engineering (Hons.) (Electronic)

Faculty of Electrical & Electronics Engineering
Universiti Malaysia Pahang

NOVEMBER, 2010

Signature : _____

Author : NUR AZIELA BINTI MANSOR

Date : 14 NOVEMBER 2010

ACKNOWLEDGEMENT

First of all I am grateful to Allah S.W.T for blessing me in completion my final year project (PSM) with success in achieving my objectives in finishing this project.

Secondly I would also like to take this opportunity to thank my supervisor Pn Nor Farizan Bt Zakaria for guiding and helping me throughout my final year project, thesis and other courses.

I also very thankful to have full support from my parents and my family for their love, courage, dedication and support in completing this project. Also do not forget to my love who always willingly assist and support me throughout my journey of education, support me and willing to hear my problem in completing this project. you are deserve my wholehearted appreciation

My sincere appreciation to the lecturers of Faculty of Electrical and Electronics Engineering who have put in effort to the lectures and always nurture and guide us with precious advices. Thank you for sharing those experiences.

Finally, also do not forget to thank to all my friends who always willingly assist and support me throughout my journey of education, you all deserve my wholehearted appreciation that have given me advice and encourage in completing my project and contributed by participating in the survey and collecting the Jawi character. Thank you very much to all and May Allah bless you.

ABSTRACT

Character recognition plays an important role in the modern world. It can solve more complex problem and makes humans' job easier. Jawi is one of the important character that we used in our daily life. Jawi script is an important Malay heritage that has been in general, replaced by the Roman script drastically. From a dominant writing in Malay world, the usage of Jawi is confined mostly in Islamic religious context nowadays. As an initiative to encourage the learning of Jawi, this research proposed Jawi Character Recognition system using Neural Network and Supervised Learning method. The aim of this research is to develop software that able to recognize Jawi character. To improve the recognition of the character, the system uses neural network training algorithm called Supervised Learning to receive new character pattern in order to strengthen the weight of the pixels. In this project, it design and train network used Radial Basis Function (RBF) with backpropagation Neural Network. This Jawi Character recognition system begins with image processing and then the output image is trained using backpropagation algorithm. Backpropagation network learns by training the input, calculating the error between the real output and target output, propagates back the error to network and modify the weight until the desired output is obtain. The system will training and recognition system will be test to ensure the system can recognize the pattern of the character.

ABSTRAK

Pengecaman aksara memainkan peranan penting dalam dunia moden. Hal ini dapat menyelesaikan masalah yang lebih kompleks dan membuat pekerjaan manusia lebih mudah. Jawi adalah salah satu aksara penting yang kita gunakan dalam kehidupan seharian. Huruf Jawi adalah warisan Melayu yang penting yang telah secara umum, secara drastik telah digantikan oleh huruf Roman. Dari tulisan yang dominan di dunia Melayu, penggunaan Jawi terhad terutamanya dalam konteks agama Islam saat ini. Sebagai inisiatif untuk menggalakkan pembelajaran Jawi, kajian ini mencadangkan Sistem Pengecaman Huruf Jawi menggunakan Rangkaian Neural Network dan kaedah bimbing Belajar. Tujuan kajian ini adalah untuk membangunkan perisian yang dapat mengenal huruf Jawi. Untuk meningkatkan pengecaman huruf, sistem ini menggunakan algoritma latihan rangkaian Neural Buatan disebut Pembimbing Belajar untuk menerima pola watak baru dalam rangka memperkuat berat piksel. Dalam projek ini, rangkaian ini direka bentuk dan dilatih menggunakan Rungsi Pangkalan Radial(RBF) dengan algoritma Penghantaranbalik dalam Rangkaian Neural. Sistem Pengecaman Huruf Jawi ini bermula dengan pemprosesan imej dan kemudian imej yang dikeluarkan dilatih dengan algoritma Penghantaranbalik. Algoritma Penghantaranbalik mempelajari melalui input data dengan melatih input data untuk menghitung ralat antara keluaran sebenar dengan keluaran sasaran, dan menghantar kembali kesalahan atau ralat tersebut ke dalam rangkaian dan mengubahsuai pemberat sehingga mendapat keluaran sasaran yang dikehendaki. Sistem ini akan dilatih dan sistem pengecam akan diuji untuk memastikan sistem dapat mengenali pola watak.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

Jawi is one of the earliest characters used for writing Malay. Its development is linked with arrival of Islam in Nusantara. Jawi is an adapted Arabic character for writing Malay language. It is an important Malay heritage that has been disregarded nowadays and the usage is only seen in a religious context. Jawi characters consist of 29 Arabic characters along with some extra characters unique to Jawi to accommodate Malay language. Al-Quran, the holy book of Islam, is recited in Arabic language using Arabic characters. The relationship between Jawi and Arabic characters shows the importance of being Jawi literate in order for a person to be able to recite Al-Quran verses in performing the prayer, which is a mandatory obligation in Islam.

The aim of this research is to develop software that able to recognize the letter of Jawi characters. This system can recognize jawi character from “Alif” to”Nya” and functions as a method to guide the people to recognize the jawi character. This project mainly using MATLAB software to develop a programming

and uses Graphical User Interface (GUI) to perform the system and make an analysis for recognizing Jawi character.

Basically, in this recognition system, this project uses Artificial Neural Network using backpropagation method, in MATLAB software to recognize and train the jawi character because this technique is more efficient to recognize the pattern of recognition.

1.2 Problem Statement

There are several factors why Jawi had been neglected nowadays. There are identified as the followings.

1. Lacking of reading material written in Jawi script available in the market.

   Nowadays, English language had become the main spoken and written language among countries in the world, as well as in Malaysia. On the other hand, Jawi materials had been decreasing in an alarming state. This is proven by the discontinued production of “Utusan Melayu”, the only Malay newspaper written in Jawi script in our country due to lacking of sales compared to other newspapers written in other language. Thus, this situation is a sign that this script had become an endangered script and it will soon become historical without any initiatives in gaining back its popularity.

2. The status of Jawi in Malaysian educational system and the implication to future generation

The reality in our educational system until today is that the reading capability is only focused on Roman script. As we know, Jawi is not the main subject in the school nowadays. This is one of the factors why Jawi has been neglected. Student just focuses on their exam subject to score. That why many Malay people are illiterate in Jawi character and do not know how to reading in Jawi.

Malay students in an Islamic-based school, they have a very strong knowledge of Jawi script compared to government school students. The Malay students in a government school were exposed to Jawi script only in the subjects related to Islamic religion. Meanwhile, students in an Islamic-based school used Jawi scripts in most of their subjects. This is because it is compulsory for them to take the academic subjects, which are very related to Islamic religion such as Arabic Language and Al-Quran studies. As a result, the Islamic-based school students produced more Jawi literate students than the government school.

Based on the factor above, Jawi Character Recognition System is develop in order to tackle the people in learning Jawi and recognizes the Jawi character.

1.3 Objectives

The objectives of this project are:

i. To recognize jawi character for the recognition analysis using Neural Network
ii. To develop software that able to recognize Jawi character.

1.4 Scope of Project

The scopes of this project are listed below:

i. The system only recognize single letter jawi character
ii. Built program in MATLAB using Neural Network to recognize the image.
iii. The image used is an offline grayscale image jawi character as a database
iv. Use Graphic User Interface (GUI) in MATLAB to perform this system
v. Design And Train the Network with backpropagation in Neural Network

CHAPTER 2

LITERATURE REVIEW

2.1 Background of Jawi

Jawi is one of the earliest characters used for writing Malay. Its development is linked with arrival of Islam in Nusantara. Jawi is adapted from Arabic characters, which consists of 29 characters. Arabic is the written language of the Quran, the holy book of Islam. Arabic language originated in Saudi Arabia in pre-Islamic times, and spread rapidly across the Middle East. Jawi has been widely used in Malay Peninsular since the 17th century until the British invented the romanised Malay script when they colonized Malay Peninsular in the 18th century. English language had subsequently greatly influenced the spelling structure of the Malay language until it was standardized in the post 1973 years [Mashkuri, Zainab, Rohana and Nor Edzan, 2005]. Thereafter, the Romanised Malay or "Rumi" had been widely used, including in the government sectors. Jawi on the other hand, are seen occasionally in certain areas and certain contexts, particularly with regards to Islamic religion only. Today the Jawi font is mainly used for Islamic religious documents and texts. There are factors that threatened even these usages of Jawi. One dominant factor is the growing reluctance among local publishers to publish religious books for the public mainly in Jawi.[1]

Table 2.1: Basic Jawi Character

	Arabic Characters	In English written as		Arabic Characters	In English written as
1	ا	Alif	15	ض	Dhad
2	ب	Ba	16	ط	Tha
3	ت	Ta	17	ظ	Zho
4	ث	Tsa	18	ع	Ain
5	ج	Jim	19	غ	Ghain
6	ح	Ha	20	ف	Fa
7	خ	Kho	21	ق	Qaf
8	د	Dal	22	ک	Kaf
9	ذ	Dzal	23	ل	Lam
10	ر	Ra	24	م	Mim
11	ز	Zai	25	ن	Nun
12	س	Sin	26	و	Wau
13	ش	Syin	27	ه	Ha
14	ص	Shad	28	ء	Hamzah
			29	ي	Ya

Table 2.2: Seven Character Unique to Jawi

	Jawi characters	In English written as		Jawi characters	In English written as
1	چ	Cha	5	ڽ	Nya
2	ڠ	Nga	6	ۏ	Va
3	ڤ	Pa	7	ﻻ	LamAlif
4	ک	Ga			

2.2 Preprocessing

In the processing stage, commonly its concentrate on removal of non-useful information such as noise, skew detection and correction. To remove the noise as isolated pixels for any given pixel, the checking of the existence of a neighboring pixel in all the possible eight directions is needed (Figure 2.1). In this stage, there are three steps involve to recognize the character which is filtering, image enhancement and binarization. [2] Binarization is the image conversion into binary image. [1] In many application involving character recognition, the input are first binarized to form a two level image based on the threshold value. It is common to preprocess the scanned image for image enhancement to eliminate the presence of noise and other type of distortions that occur during the scanning process.

Image enhancement is the basic way to change the brightness and contrast of the image. By enhancing the image, the intensity of the image can be adjusted. In this preprocessing stage, the image is enhanced, resized and binarized to make the image clearer and more accurate. It is necessary to employ the non linear technique for processing the character images prior to binarization.[3,4]

There are several methods available for thresholding image to produce binary image. An experimental performance evaluation of several such technique can be found in[3]. These methods include fixed global threshold, otsu threshold and other techniques. The input grayscale pixels are denoted b x_i ε {0, 1}. The corresponding output binarization pixel are denoted by b_iε{0,1} where 0 refer to 'black' and 1 refer to 'white'[4]

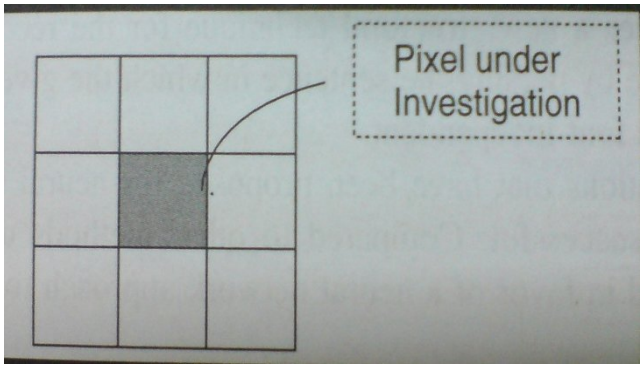

Figure 2.1: Eight-Connected Component

Image enhancement is necessary to remove noise by filtering the image, adjusting the contrast of the image and enhancing the edge of the character before binarization process and training the character recognition. [3].

In this project, the global image threshold using Otsu's method will be chosen as the technique to binarized the character image. This method finds the global threshold that minimized the intrasclass variance of the resulting black and white pixels of the image. This method also standard binarization technique and was implemented in MATLAB function "graythresh". [4, 6].

2.3 Artificial Neural Networks (ANNs)

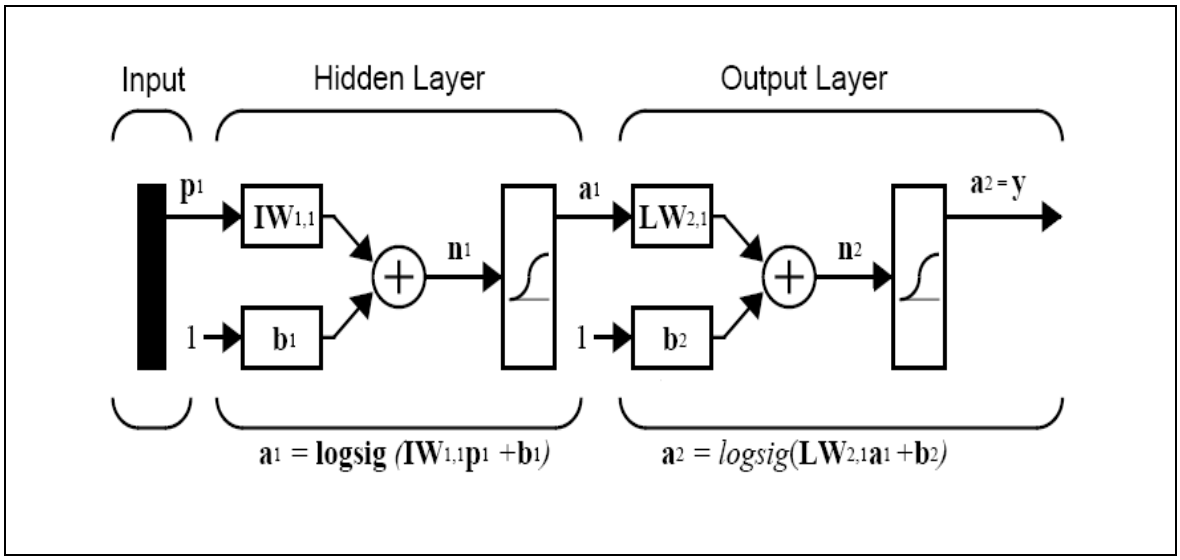

Figure 2.2: Neural Network Architecture [9]

Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements.

Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output as shown in Figure 2.3. There, the network is adjusted, based on the comparison of output and target, until the network output matches the target. Typically many such input or target pairs are needed to train a network. [7].

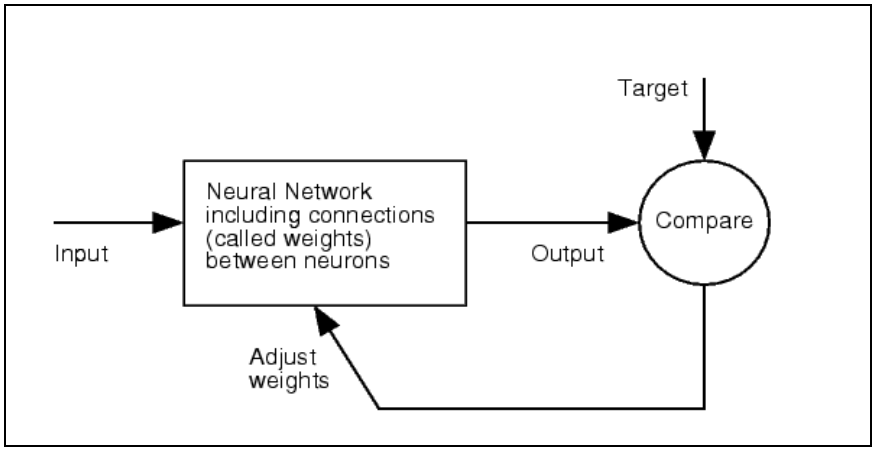

Figure 2.3: Neural Network Training

Neural networks have been trained to perform complex functions in various fields, including pattern recognition, identification, classification, and speech, vision, and control systems.

Today neural networks can be trained to solve problems that are difficult for conventional computers or human beings. This Neural Network can be implementing in image recognition application.

2.3.1 Multilayer Neural Networks

Multilayer neural networks are feedforwardANN models which are also referred to as multilayer perceptrons. The addition of a hidden layer of neurons in the perceptron allows the solution of nonlinear problems such as the XOR, and many practical applications (using the backpropagation algorithm).

However, the difficulty of adaptation of the weights between the hidden and input layers of the multilayer perceptrons has dampened such architecture during the sixties. With the discovery of the backpropagation algorithm by Rumelhart, Hinton

and Williams in 1985, the adaptation of weights in the lower layers of multilayer neural networks are now possible.

The researchers proposed the use of semilinear neurons with differentiable activation functions in the hidden neurons referred to as logistic activation functions (or sigmoids) which allows the possibility of such adaptation.

2.3.2 Advantages and Disadvantages of Artificial Neural Networks (ANNs)

ANNs is a system that takes the operation of biological neural networks as conceptual basis as it is an emulation of biological neural system. Despite the disadvantages that it is made with, it performs certain tasks that a program made for a common microprocessor is unable to perform. In other words, a neural network can perform tasks that a linear program cannot.

When an element of the neural network fails, its parallel nature enables it to continue without any problem. Besides, it learns and does not need to be reprogrammed. Thus, it can be implemented in any application without any problem. However, the neural network needs training prior to its operation. Its architecture is different from that of a microprocessor; therefore, it needs to be emulated.[8]

In addition, high processing time is required for large neural networks. Artificial neural networks can have different architectures that consequently require different types of algorithms, but it is relatively simpler than to be a complex system.

2.3.3 Mathematical Modeling of ANNs from Biological Model

A biological nervous system consists of neurons as the basic signaling units where each neuron is a discrete cell whose several processes arise from its cell body. The ANNs emerged as circuits that could perform computational tasks with biological neurons as basic conceptual components.

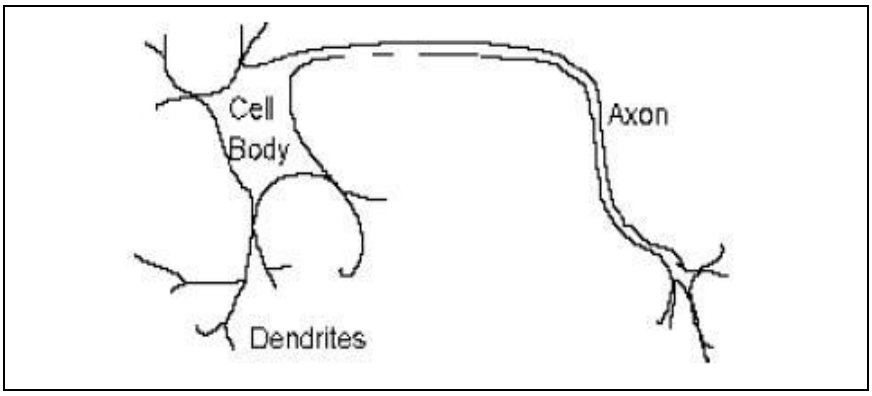

Figure 2.4: Biological model (Neuron)

The neurons or cells as shown in Figure 2.4 are modeled as processing units where the area of contact between two physically non-touching neurons is called synapse where in this synaptic cleft electric signals are sent through chemical interactions. In a functional model, the synapses are modeled as weights and their values note the connection strength between an input and a neuron. [9]

The inputs are modified by their respective weights before linear combination takes place whereby they are summed up by an adder. Then, an activation function will control the amplitude of the neuron output to a range between 0 and 1, or, -1 and 1. This is mathematically described in the Figure 2.5 below according to the formula shown in Figure 2.6.

$$v_k = \sum_{j=1}^{p} w_{kj} x_j$$

Figure 2.5: Mathematical Formula for adjust weight

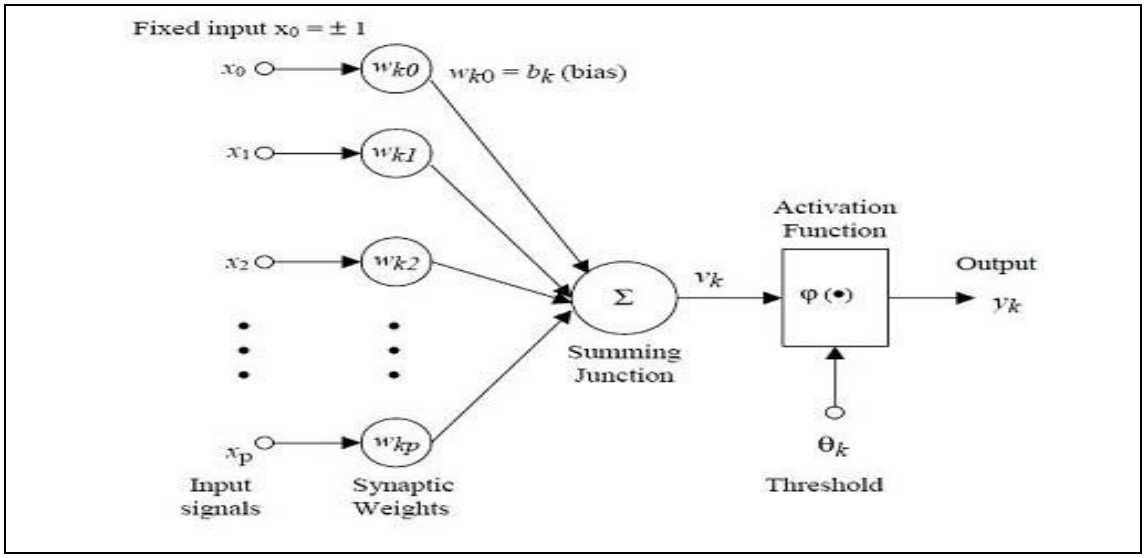

Figure 2.6: Mathematical model (ANNs)

The neuron output, yk, is the outcome of some activation function on the value of vk. In short, an artificial neural network is a pool of simple processing units that communicate by sending signals to each other over a large number of weighted connections. Apart from adjusting the weight, each processing units receive input from neighbours or external sources to compute an output signal which is propagated to other units. [2, 9]

There are three types of units in neural systems: input units which receive data from outside the neural network, output units which send data out of the network and hidden units whose input and output signals remain within the network.