DEVELOPMENT OF HANDWRITTEN CHARACTER RECOGNITION BY USING ARTIFICIAL NEURAL NETWORK

LISA UMAMI BINTI ABDUL KHALIK

This Report Is Submitted In Partial Fulfillment of Requirement for the Bachelor Degree of

Electronic Engineering (Wireless communication)
Faculty of Electronic and Computer Engineering Universiti Teknikal Malaysia Melaka

JUNE 2015

© Universiti Teknikal Malaysia Melaka

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

FAKULTI KEJURUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : DEVELOPMENT OF HANDWRITTEN CHARACTER RECOGNITION BY USING ARTIFICIAL NEURAL NETWORK.

Sesi Pengajian : 2014/2015 Semester 2

Saya LISA UMAMI BINTI ABDUL KHALIK mengaku membenarkan laporan Sarjana Muda ini disimpan di Perpustakaan dengan syarat-syarat kegunaan seperti berikut:

1. Laporan adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan laporan ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (√) :

☐	SULIT*	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
☐	TERHAD*	(Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
☐	TIDAK TERHAD	

Disahkan oleh:

(TANDATANGAN PENULIS)

Alamat : Kg. Bukit Rahmat,84700,Gerisik,Ledang,Johor.

Tarikh: 9/6/15

Disahkan oleh:

(COP & TANDATANGAN PENYELIA)
ROSMAN ABD RAHIM
Lecturer
Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka
Hang Tuah Jaya
76100 Durian Tunggal, Melaka

Tarikh: 9/6/2015

© Universiti Teknikal Malaysia Melaka

"I hereby declare that this is the results of my own paper except for quotes as cited in the references.

Signature :

Author : LISA UMAMI BINTI ABDUL KHALIK

Date : 8 JUNE 2015

© Universiti Teknikal Malaysia Melaka

"I hereby declare that I have read this report and in my opinion this report is sufficient in terms of the scope and quality for the award of Bachelor Degree of Electronic and Computer Engineering (Wireless Communication) with Honours."

Signature :

Supervisor's Name : MISROSMAN BIN ABDUL RAHIM

Date : 8 JUNE 2015

© Universiti Teknikal Malaysia Melaka

This project and research work is dedicated to my beloved parents, for their devoted caring throughout my life, my beloved brothers and sisters, who give the inspiration to me and also my friends for their encouragement. Without their support this report may not have been done.

© Universiti Teknikal Malaysia Melaka

ACKNOWLEGDEMENT

Alhamdulillahirabbil „alamin washolatu wassalamu‟ala asrafil anbiyai wal mursalin, wa‟ala alihi waaskhabihi ajma‟in. praise to Allah S.W.T the most gracious and merciful for giving me the strength and wisdom in finishing Final Year Project (FYP) and report for the session 2015 as well as possible. First and foremost, I would like to express my sincere gratitude to both of my supervisor Mr. Muhammad Noorazlan Shah Bin Zainudin and Mr. Rosman bin Abdul Rahim, for his patience, motivation, enthusiasm, and massive knowledge. Their guidance helped me in all the time of research and writing of this report. I could not have imagined having a better advisor and supervisor for my Final Year Project (FYP).

During completing this Final Year Project (FYP) and report, I face a lots of problems and difficulties. However, I manage to crossed it all with patience and complete it appropriately. The journey in completing Final Year Project (FYP) and report really teach me in terms of leadership, proper time management, communication skill, technical knowledge, soft skill and so many more. The advice and encouragement from supervisor is really inspired me to do the best.

My sincere thanks also goes to my family members and friends also became a part of in completing the Final Year Project (FYP) and report, which help me in giving some rational idea and supports. Finally, I would like to thank those concerned directly and indirectly in completing the Final Year Project (FYP) and report successfully. Thank very much to you all.

© Universiti Teknikal Malaysia Melaka

ABSTRACT

The constant development of computer tools leads to a requirement of easier interfaces between the man and the computer one of the example is handwritten character recognition. Handwritten character recognition is the ability of computer to receive and interpret the handwritten input images from sources such as document, photograph or touch-screen and transform it to machine readable and editable format. In this project the system implemented using the MATLAB software which using tool of Graphic User Interface (GUI). There are several steps involved in this system. The first step is preprocessing. The input image which consists of the handwritten text will be loaded and will go through some process like conversion to grayscale and binary image and image filtering. Next, the image will be segmented to individual character bounding plotting the bounding block and then it will be extracted according to respective to their feature vector. Finally, the classification process will take place and the extracted character will be classified by using the Artificial Neural Network. The GUI is created to uses windows, icons, and menus to bring out commands, such as opening, deleting, and moving files and to complete each steps.

© Universiti Teknikal Malaysia Melaka

ABSTRAK

Pembangunan berterusan alat komputer membawa kepada keperluan antara muka yang lebih mudah antara manusia dan komputer salah satu contoh adalah pengenalpastian watak tulisan tangan. Pengenalpastian watak tulisan tangan ialah keupayaan komputer untuk menerima dan mentafsir imej input tulisan tangan dari sumber seperti dokumen, gambar atau skrin sentuh dan mengubahnya ke mesin format yang boleh dibaca dan disunting. Dalam projek ini sistem yang dilaksanakan dengan menggunakan perisian MATLAB yang menggunakan alat Antara Muka Pengguna Grafik (GUI). Terdapat beberapa langkah yang terlibat dalam sistem ini. Langkah pertama ialah pra pemprosesan. Imej input yang terdiri daripada teks tulisan tangan akan dimuatkan dan akan melalui beberapa proses seperti penukaran kepada skala kelabu dan imej binari dan penapisan imej. Seterusnya, imej tersebut akan dibahagikan kepada bounding watak individu merancang blok bounding dan kemudian ia akan diekstrak mengikut masing-masing untuk vektor ciri mereka. Akhir sekali, proses klasifikasi akan berlaku dan watak yang diekstrak akan diklasifikasikan dengan menggunakan Rangkaian Neural Buatan (Artificial Neural Network). GUI ini bertujuan untuk kegunaan window, ikon dan menu untuk membawa keluar arahan, seperti membuka, memotong, dan fail yang bergerak dan untuk menyelesaikan setiap langkah.

© Universiti Teknikal Malaysia Melaka

TABLE OF CONTENT

CHAPTER	CONTENT	PAGE
	PROJECT TITLE	**i**
	REPORT VERIFICATION STATUS FORM	**ii**
	DECLARATION	**iii**
	SUPERVISOR DECLARATION	**iv**
	DEDICATION	**v**
	ACKNOWLEDGEMENT	**vi**
	ABSTRACT	**vii**
	ABSTRAK	**viii**
	TABLE OF CONTENT	**ix**
	LIST OF TABLES	**xiii**
	LIST OF FIGURE	**xiv**
	LIST OF ABBREVIATIONS	**xvi**
	LIST OF APPENDICES	**xvii**
I	**INTRODUCTION**	
	1.1 Project background	**1**
	1.2 Problem Statement	**2**
	1.3 objectives	**4**
	1.4 scope	**4**
	1.5 Project significant	**5**

© Universiti Teknikal Malaysia Melaka

1.6	Expected Output	**5**

II LITERATURE REVIEW

2.1	Introduction to Handwritten Character Recognition	**6**
2.2	The Proposed Handwritten Character Recognition Program and Methodology	**7**
2.3	Data Collection	**10**
2.4	Pre-processing	**10**
	2.4.1 Noise removal	**11**
	2.4.2 Binarization	**11**
	2.4.3 Normalization	**12**
2.5	Segmentation	**12**
2.6	Feature Extraction	**13**
2.7	Feature Classification	**14**
	2.7.1 Artificial Neural Network	**14**
	2.7.1.1 Multilayer Perceptron Network (MLP)	**15**

III METHODOLOGY

3.1	Introduction	**17**
3.2	Project planning	**18**
3.3	System Requirement Specifications	**19**
	3.3.1 Hardware and Software Requirements	**19**
	3.3.2 High Level Specification	**19**
3.4	Flow Chart	**20**

© Universiti Teknikal Malaysia Melaka

3.4.1 Pre-processing **22**
3.4.1.1 Binarization **22**
3.4.1.2 Edge Detection **23**
3.4.1.3 Morphological Operation **23**
3.4.2 Segmentation **24**
3.4.3 Feature Extraction **25**
3.4.3.1 Feature Extraction Based on Character Geometry **25**
3.4.3.2 Feature Extraction Based on Gradient Function **29**
3.4.4 Classification **31**
3.4.4.1 Artificial Neural Network **31**

IV RESULT AND DICUSSION

4.1 Introduction **34**
4.2 Result **34**
4.2.1 Input Image **35**
4.2.2 Processing Image **36**
4.2.3 Training the dataset **39**
4.2.3.1 Artificial Neural Network **39**
4.2.4 Feature Extraction **42**
4.2.5 Testing and Classification **43**
4.2.6 Performance **46**
4.2.7 System GUI **48**
4.3 Discussion **51**

© Universiti Teknikal Malaysia Melaka

V CONCLUSION AND FUTURE EXTRACTION

5.1 Conclusion **52**

5.2 Future Suggestion **53**

REFERRENCES **54**

APPENDIX A **56**

APPENDIX B1 **58**

APPENDIX B2 **61**

APPENDIX C **65**

© Universiti Teknikal Malaysia Melaka

LIST OF TABLES

NO. OF TABLE	TITLE	PAGE
3.1	Minimum requirement	**19**
4.1	The result of training session	**40**
4.2	The testing-phase result	**43**

© Universiti Teknikal Malaysia Melaka

LIST OF FIGURES

NO.OF FIGURE	TITLE	PAGE
2.1	The basic step of recognition system	**8**
2.2	The architecture of handwritten recognition	**9**
2.3	Pre-processing result	**9**
2.4	Skeletonization of English character	**11**
2.5	Pixel P and 8-neighbor	**14**
2.6	Multilayer Perception Network (MLPN)	**15**
3.1	Flow chart of Optical Character recognition	**21**
3.2	Normalization of image	**24**
3.3	Universe of Discourse	**25**
3.4	Starters are rounded	**26**
3.5	Intersections	**27**
3.6	Minor starters	**28**
3.7	Sobel masks for Gradient	**29**
3.8	Directions of chain codes	**30**

© Universiti Teknikal Malaysia Melaka

3.9	Typical Neural Network	**32**
3.10	Neural Network	**33**
3.11	Sample Input & Output	**33**
4.1	Some of the training-sets	**35**
4.2	Some of the test-sets	**35**
4.3	Before pre-processing	**36**
4.4	Converted to gray scale image	**37**
4.5	Converted to binary image	**37**
4.6	Edge detection	**37**
4.7	Image dilation	**38**
4.8	Image filling	**38**
4.9	After pre-processing	**38**
4.10	The layers in the Neural Network	**39**
4.11	Confusion matrix	**41**
4.12	Feature extraction	**42**
4.13	Comparison of the recognition accuracy for individual characters	**46**
4.14	Comparison of the recognition accuracy of input words image for testing	**47**
4.15	Home page	**48**
4.16	Uploading image	**49**
4.17	Neural Network Training	**49**
4.18	Extracting Characters	**50**
4.19	Output file	**50**

© Universiti Teknikal Malaysia Melaka

LIST OF ABBREVIATIONS

HCR	Handwritten Character Recognition
OCR	Optical Character Recognition
US	United States
PDA	Personal Digital Assistant
ANN	Artificial Neural Network
MLPN	Multilayer Perception Network
GUI	Graphical User Interface
MSE	Mean square Equation

© Universiti Teknikal Malaysia Melaka

LIST OF APPENDICES

NO.OF APPENDIX	TITLE	PAGE
APPENDIX A	Coding for pre-processing	**57**
APPENDIX B1	Coding for training	**59**
APPENDIX B2	Coding for testing	**61**
APPENDIX C	Full coding in the GUI interface	**64**

© Universiti Teknikal Malaysia Melaka

© Universiti Teknikal Malaysia Melaka

© Universiti Teknikal Malaysia Melaka

CHAPTER 1

INTRODUCTION

1.1 Project background

Handwritten Character recognition has been one of the most fascinating and challenging research areas of image processing and pattern recognition in the recent years [4]. HCR is the capability of a computer to be given and interpret intelligible handwritten input for example digital cameras as well as other devices. HCR system involves a number of phases that are preprocessing, feature extraction, and classification. Handwritten Character recognition can be classified into two elements which are on-line and off-line character recognition.

Off-line character recognition includes the automatic conversion of text in an image into letter codes which can be workable within computer and text-processing

© Universiti Teknikal Malaysia Melaka

applications. The data acquired by this form is considered a static representation of handwriting. Off-line character recognition is somewhat challenging, as different people have different handwriting patterns. On-line character recognition comprises of the automatic conversion of text as it is written on a particular digitizer or PDA, in which a sensor senses the pen-tip movements and also pen-up/pen-down switching. This type of data is known as digital ink and can be perceived as a digital outline of handwriting. The obtained signal is transformed into letter codes which are workable within computer and text-processing applications.

This study introduces the principle stages of HCR system and the classification process for recognizing a handwritten character. That process will be analyzing using Artificial Neural Network. The objectives of this project is to take handwritten English characters as input, process the character, and then train the proposed algorithm, to recognize the character.

This project is targeted at study the algorithms in recognizing characters of English language. This project is limited to English characters only. This is because English is the internationally used language. It can be more deeply established to recognize the characters of different languages.

1.2 Problem Statement

Optical character recognition is a part of biometric system. Since 1929, number of character recognition systems have been develop and are use for even commercial purpose also [1]. Several applications including mail sorting, bank processing, document reading and postal address recognition, bank checking process [1] require off-

© Universiti Teknikal Malaysia Melaka

line handwriting systems. Working in Postal service need us to decode and deliver something like 30 million handwritten envelopes every single day. The challenges are to do mail-sorting that ensure all those millions of letters reach their destinations. When it comes to processing more human kinds of information, it is a hard task for computer to do because we have to “communicate” to them through relatively crude devices such as keyboards and mice so they can figure out what we want them to do [10].

In the US and most of other countries, bank cheques are preprinted with the account number and the check number in special ink and format; as such, these two numeric fields can be easily read and processed using computerized techniques. However, the amount fields on a filled-in check is usually read by human eyes, and involves significant time and cost, especially when one considers that over 50 billion checks are processed per annum in the US alone. The ”read” check is difficult to deal with, in term of time and cost [11]. In order to overcome this, we have to develop algorithms that can make computer understand the handwriting and to do the bank check task.

By referring to the stated problem, we have chosen the handwritten recognition as a solution to these problem. But, to implement this task we also face any problem that are :

1. Machine simulation of human functions has been a very challenging research field since the advent of digital computers

2. HCR is a challenging problem since there is a variation of same character due to the change of fonts and sizes.

3. The differences in font types and sizes make the recognition task difficult and resulting the recognition of character process become not accurate.

© Universiti Teknikal Malaysia Melaka

1.3 Objectives

The objectives of the project are:

1. To improve the techniques or method for HCR from variation of font and size.

2. To study and improve the algorithm in order to minimize the error when analyzing the extracted features of the handwriting image to achieved a greater accuracy.

3. To analyze the performance of Artificial Neural Network.

1.4 Scope

The scope of this project consist of discovering the handwritten English character recognition algorithm. There are different techniques of handwritten recognition which include a series of procedures that perform to pre-processing, character segmentation, feature extraction and classification. Besides, gathering handwriting image database from the computer will be carried out to collect the datasets for training and testing. The size of all image database will be set as 30x30 pixels. This is to make sure all the database image readily available for the pre-processing steps.

© Universiti Teknikal Malaysia Melaka

In classification part, the proposed technique is Artificial Neural Network. The characters will be analyze using proposed algorithm in term of the classification accuracy rate. The simulation will then being tested by using MATLAB software.

1.5 Project significant

This project intends to help researcher, public and industries to understand the concept of handwritten recognition and to choose the best technique for the accuracy of the system. The example of application of handwritten recognition can be seen in bank cheques checking process, document reading, and postal code or address recognition as well as have potential in reading aid for the disable. The wide use of this system or application needs us to improve the current system in order to get better application of the system in the future.

1.6 Expected output

The expected output of this project is we can see the overall step by step process of this system. In the final result, we can see how much percent of the proposed classification technique can produce to state whether the result is valid or not. At the end of this project, the three techniques will be applied using Artificial Neural Network.

© Universiti Teknikal Malaysia Melaka