FEATURE ENHANCEMENT FOR EXTRACTING ON-LINE ISOLATED HANDWRITTEN CHARACTERS

MUHAMMAD FAISAL ZAFAR

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy

Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia

AUGUST 2006

Dedicated to my beloved father (late, 7 Dec. 2005)

ACKNOWLEDGEMENT

All the praise for The **Almighty Allah SWT** for His Blessings and uncountable awards towards me in spite of all of my weaknesses and serious faults. Hundreds of millions **Darood** upon our **Holy Prophet** (Sallallaho alaihi wasallam) being a trivial Ummati.

I wish to thank and express my deepest gratitude and appreciation to the following individuals and organization that enable and motivate me in completing my study and research:

Associate Professor Dr. Dzulkifli Mohamad who greatly helped me in every way I needed to go through this study. His personal kindness, skill, patience and guidance has been given and coloring my daily life. As my supervisor, he also becomes my parent.

Associate Professor Daut Bin Daman, for his encouragement, guidance, moral support and friendship. Associate Professor **Dr. Siti Mariyam Bt. Hj. Shamsuddin** who always was kind and helpful being Head of Department. **En. Razib Othman** who is my lovely friend and well wisher. All respectable teachers of GMM for their valuable support all the time.

Professor Dr. Rahmalan and Professor Dr. Safaai Deris from **School of Graduate Study,** for their full support and cooperation in my thesis submission.

I'm also thankful to my friends Abdul Majid, Anjum Iqbal, Adnan yonus, Imran Ghani, Adil Khattak, Kashif Saleem, Mubarak Ali, Mahmoud Ali, Abdurrahman, Adil Ali, Khalid Sb and Jehan. They were always available for my help whenever I needed. It will be dishonest if I don't thank to all staff of our faculty who gave their best cooperation during my research work.

My fellow postgraduate students Chu Kai Chuan, Lai Chui Yen, Moon Tan, Ijat, Ferhan, Fadni, Ijam, Masroor and so many others should also be recognized for their valuable suggestions and support.

My parents, **Muhammad Hanif** (late) and **Razzia Beghum**, who gave me a real love, pray, support, and all they have. Special pray for my beloved father who passed away at my final stage of thesis submission. Together with my brothers and sisters, they have been supporting me morally in ups and downs. And to my wife **Laila Khalid** who gives a meaningful love and care to me. The sweet company of my kids, **Jowairia**, **Saleem** and **Laiba** also kept me fresh and alive.

I have no words to pay thanks to a great personality, **Mian Muhammad Bashir Hayat**, who actually the driving force behind all of my efforts and successes and his affection is the real assets for me in this world as well as the next.

ABSTRACT

The study of online handwriting recognition has gained an immense interest among the researchers especially with the increase in use of the *personal digital assistant* (PDA). The large number of writing styles and the variability between them make the handwriting recognition a challenging area to date. The present tools for modelling are not sufficient to cater for the various styles of human handwriting. Furthermore, the techniques used to get appropriate features, architecture and network parameters for online handwriting recognition are still ineffective. The success of any recognition system depends critically upon how far a set of appropriate numerical attributes or features can be extracted from the object of interest for the purpose of recognition. Thus the aim of this research work is to propose novel feature extraction methods to facilitate a system or device to achieve satisfactory online handwriting recognition. Two new simple and robust methods based on annotated image and sub-character primitive feature extractions have been proposed. The selection of features is based mainly on their effectiveness. Using the proposed techniques and a neural network based classifier, several experiments were carried out using the UNIPEN benchmark database. The techniques are independent of character size and can extract features from raw data without resizing. The maximum recognition rates achieved are 94% and 92% for annotated image and sub-character primitive methods respectively.

ABSTRAK

Kajian pengecaman tulisan tangan semakin mendapat perhatian para penyelidik, khususnya apabila penggunaannya telah diaplikasikan di dalam peralatan keperluan era baru seperti *personal digital assistant* (PDA). Kepelbagaian gaya tulisan dan kewujudan beberapa pembolehubah yang boleh mempengaruhi gaya tulisan menjadikan pengecaman tulisan tangan satu bidang kajian yang agak mencabar pada hari ini. Peralatan pemodelan yang sedia ada pada hari ini masih tidak mampu menangani kepelbagaian gaya tulisan tangan manusia. Tambahan pula teknik untuk mendapatkan parameter kesesuaian ciri, senibina dan rangkaian untuk mengecam tulisan tangan secara atas talian masih juga kurang berkesan. Keberkesanan suatu sistem pengecaman adalah bergantung sepenuhnya kepada sejauhmana set ciri atau sifat numerik yang sesuai dapat diekstrak daripada objek yang hendak dicam. Oleh itu, tumpuan utama kajian ini adalah untuk mencadangkan kaedah baru pengekstrakan ciri bagi membantu sistem atau alat untuk mendapatkan satu pengecaman tulisan tangan secara atas talian yang lebih berkesan. Dua kaedah baru yang mudah dan tegar berasaskan pengekstrakan ciri imej teranotasi dan primitif sub-huruf telah dibangunkan. Pemilihan ciri dilakukan hanya berdasarkan kepada keberkesanan. Dengan menggunakan kaedah yang telah dibangunkan ini bersama pengelas rangkaian neural, beberapa pengujian telah dilakukan dengan menggunakan data daripada pangkalan data piawai UNIPEN. Teknik ini didapati tidak terhad kepada saiz huruf dan mampu mengesktrak ciri daripada data mentah tanpa perlu pensaizan semula. Kadar pengecaman tertinggi yang telah dicapai adalah 94% untuk imej teranotasi dan 92% untuk primitif sub-huruf.