Handwritten Computer Program Recognition, Compilation & Execution Application

M.K Wanigapura

149234E

FACULTY OF INFORMATION TECHNOLOGY,

UNIVERSITY OF MORATUWA

2017

Handwritten Computer Program Recognition, Compilation & Execution Application

M.K Wanigapura

149234E

A DISSERTATION

SUBMITTED TO THE

FACULTY OF INFORMATION TECHNOLOGY,

UNIVERSITY OF MORATUWA, SRI LANKA

FOR THE PARTIAL FULFILLMENT OF THE REQUIREMENT

OF THE MASTERS DEGREE OF

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

JUNE 2017

Declaration

I hereby declare that the project work entitled "Handwritten Computer Program Recognition, Compilation & Execution Application", submitted to the university of Moratuwa, Sri Lanka, is a record of an original work done by me, under the guidance of my Supervisor Senior Lecturer Mr.Saminda Premaratne. This project work is submitted in the partial fulfillment of the requirement for the award of the degree of Master of Science in Information Technology. The results embodied in this report have not been submitted to any other University or Institution for the award of any degree or diploma. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Supervisor

Mr.Saminda Premaratne…… Date: / /

(Senior Lecturer – University of Moratuwa)

Student

M.K Wanigapura…… Date: / /

Dedication

I would like to dedicate this thesis to my supervisor, senior lecturer **Mr.Saminda Premaratne** who gave me tremendous support and motivation throughout the entire process of the research.

Acknowledgement

I would like to express my special appreciation and thanks to my advisor, Senior Lecturer Mr.Saminda Premaratne, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research student. Your advice on both research as well as on my career has been invaluable.

Besides my supervisor, I would like to thank, Prof. Asoka S. Karunananda Dean of Kothalawala Defense University. The quality of this thesis and research have improved by using comprehensive guidelines of Prof. Karunananda. I would also like to acknowledge the rest of lectures in the department of IT at university of Moratuwa, Sri Lanka for their encouragement.

Last but not the least, I would like to thank my colleagues and family for supporting me spiritually throughout and my life in general.

Abstract

Make a computer program that does automatically recognize a handwritten computer program, compile and execution is extremely difficult. The reasons for these are the various types, shapes of handwritten characters of different peoples. This paper provides an accurate handwritten character recognition method, which blend with the image processing and the training for handwritten characters. Further, automatic error correction of handwritten program, compilation and execution of the program, will be discussed in detail throughout the paper.

Recognition of handwritten characters has been a research challenged. Handwritten character recognition belongs to the family of optical character recognition performing automatic identification. Among other issues in optical character recognition, handwritten identification, accuracy or the correctness is the main research issues.

This research has developed a system for identification of handwritten characters. The solution can identify the handwritten characters with a higher level of accuracy. This solution can deal with any images captured by a digital camera or scanned images of handwritten characters. The algorithm applies on preprocessed image with handwritten characters. The preprocessor has used standard image processing techniques and the trainer is used to train for the specific handwritten characters.

Table of Contents

1 Introduction 1

1.1 Prolegomena 1

1.2 Background and motivation 2

1.3 Problem statement 2

1.4 Hypothesis 3

1.5 Objectives 3

1.6 Blend of Image Processing & Training approach 3

1.7 Structure of the thesis 4

1.8 Summary 4

2 Developments and Challenges in Handwritten Character Identification 5

2.1 Introduction 5

2.2 Categorization of OCR 5

2.3 Essential Image Processing Techniques used in OCR 6

2.4 Supporting Tools and Techniques in OCR 11

2.5 Problem Definition 12

2.6 Summary 13

3 Image processing and Tesseract OCR engine in handwritten computer program recognition 14

3.1 Introduction 14

3.2 Image processing – Preprocessing 14

3.3 Tesseract OCR engine 14

3.4 Summary 16

4 Approach 17

4.1 Introduction 17

4.2 Hypothesis 17

4.3 Users 17

4.4 Input 17

4.5 Output 18

4.6 Process 18

4.6.1 Image Acquisition 19

4.6.2 Image Preprocessing 19
4.6.3 Language Training 22
4.6.4 Character Recognition 23
4.6.5 Post Processing and program execution 23
4.7 Features 24
4.8 Summary 24
5 Design 25
5.1 Introduction 25
5.2 Top Level Architecture 25
5.3 Interface Module 26
5.4 Preprocessing module 26
5.5 Language Training module 27
5.6 Character Recognition Module 27
5.7 Post processing and execution module 28
5.7.1 Format Code 28
5.7.2 Check Brackets 28
5.7.3 Check & Correct Syntax 28
5.8 Summary 29
6 Implementation 30
6.1 Introduction 30
6.2 Overall Solution 30
6.3 Implementation of the interface module 30
6.4 Implementation of the preprocessing module 32
6.4.1 Gray Scaling 32
6.4.2 Thresholding 33
6.4.3 Noise Removal 33
6.4.4 Thinning 33
6.4.5 Skewing 33
6.5 Implementation of Training Module 34
6.6 Implementation of Post processing and Execution module 36
6.7 Summary 37
7 Evaluation 38

7.1 Introduction 38
7.2 Participants 38
7.3 Testing Environment 38
7.4 Data collection 38
7.5 Data Analysis 39
7.6 Summary 43
8 Conclusion 44
8.1 Introduction 44
8.2 Overview of the research 44
8.3 Major Findings 44
8.4 Future work 45
8.5 Summary 45
References 47
Appendix 50

Table of Figures

Figure 1 - Tesseract OCR engine 15
Figure 2 - schematic diagram of the recognition system 19
Figure 3 - Image Pre-processing Steps 20
Figure 4 - Image feed for OCR 23
Figure 5 - Top Level Architecture 25
Figure 6 - Design of Pre-processing module 26
Figure 7 - Design of Language Training module 27
Figure 8 - character recognition module 27
Figure 9 - Design of post processing Module 28
Figure 10 - User Interface for image acquisition and pre-processing 31
Figure 11 - User Interface for OCR Training 31
Figure 12 - User Interface for post-processing 32
Figure 13 - Box File Editor 35
Figure 14 –Gearating font properties 35
Figure 15 – Check bracket in post-proccessing 36
Figure 16 - Check Syntax in post-proccessing 37
Figure 17 – Character recognition parentage 39
Figure 18 – Average character recognition rate 40
Figure 19 – Recognized character count for different users 40
Figure 20- Average recognition rate for different users 41
Figure 21 – Character recognition on case sensitivity 41
Figure 22 – Average character recognition on case sensitivity 42
Figure 23 – Character recognition count 42
Figure 24 – Average character recognition count 43

Table of Tables

Table 1 - Average recognition rate 45
Table 2 - Average Character Count 45