


FEATURE BASED INDEXING OF HAND-WRITTEN TEXT IMAGES

NAGARAJAH KUNTHARSHAN

 This dissertation was submitted to the Department of Computer Science and Engineering of the University of Moratuwa in partial fulfillment of the requirement for the Degree of M Sc in Computer Science

University of Moratuwa, Sri Lanka
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

March 2010

96425

Abstract

Identity management system which is maintaining a manual database system to store its data is facing a big challenge of searching a particular entry from the database if a query rose where the date of birth of the person is not known. Person Identification department, as an example, the existing database is in the form of Pre-designed hand written physical cards with the details. The cards are served separated by sex and ordered by date of birth.

In case of a person whose date of birth is not known, there is a challenge of going through the cards one by one to spot the exact record. So, person registration department decided to computerize the database and found out the following. Since the hand writing is very poor there is no chance of using character recognition software. The better option could be manually enter the data into a database, but quotations submitted for the tender called for this task was much higher than the potential level of the department.

Finally, feature extraction method is found out as an ideal solution for this task. In this approach all the cards are scanned and saved in the system using batch scanning. Each file is pre-processed and the number of characters in the name is saved to the database as index for the corresponding scanned image. The search operation on the database based on the number of characters of the name will list down the name of the possible card and the corresponding card also will be fetched from the saved location and previewed. Among the cards that are identified, user needs to find the cards manually.

This will narrow down the search. System will fail in counting the number of characters in the name if there is no space left between two characters or if one character is split "into two by mistake. To handle these challenges, a new intelligent algorithm should be generated with the ability to understand the order or the pattern of occurrences of the characters and make the decision based on them

Declaration

I, Nagarajah Kunatharshan hereby declare that the work included in this dissertation in part or whole has not been submitted for any other academic qualification at any institution.

N. Kunatharshan

Mr. N. Kunatharshan

Submitted for examination

UOM Verified Signature

DR. Chathura De Silva

(Supervisor)



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Acknowledgements

Having completed the MSc. Degree program at the Department of Computer Science and Engineering, University of Moratuwa, I would like to take this opportunity to thank those who helped me a lot in completing it at this level.

First of all, I would like to thank my supervisor, Dr. Chathura De Silva for the supervision and guidance despite his busy schedules. Also I would like to thank Ms. Visaka Nanayakkara, Dr. Gihan Dias and Dr. Sanath Jayasena for conducting the subject Research Seminar, which helped me a lot when browsing through the published research papers.

I should be grateful to all the members of the staff of the Department of Computer Science and Engineering for giving valuable feed back during the progress review sessions.

Finally, I would like to thank my family for supporting me in every aspect in completing my degree program and the research successfully.

Thank you all!



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Kunatharshan N

(06/8279)

Table of Contents

Declaration.....	2
Abstract.....	3
Acknowledgements.....	4
Table of Contents.....	5
List of Figures.....	7
1. Introduction.....	8
2. Literature Review	11
2.1 Edges Detection:	11
2.2 Lines Detection – Hough Transform:	12
2.3 Image Registration with Template matching	13
2.4 Cross-correlation.....	14
2.5 Open and Close.....	14
2.6 Erosion	15
2.7 Dilation	15
2.8 Opening.....	16
2.9 Closing	16
2.10 Color Scale Conversion – gray scale.....	17
3. Materials and Methods	18
3.1 Adding images to the system	19
3.1.1 Template image:.....	19
3.1.2 Printed Character Removal:	20
3.1.3 Removal of lines:.....	21
3.1.4 Closing:	23
3.1.5 Character counting:	24
3.2 Finding images.....	24
4. Observations	25

 University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

5. Analysis and Discussion of Results..... 26

 5.1 Binary image conversion: 26

 5.2 Closing: 27

 5.3 Line removal: 28

6. Conclusions and Recommendations for Future Research 29

7. References..... 30

8. Annexure..... 32



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

List of Figures

Figure 3.1: Flowchart of the model	18
Figure 3.2: Scanned image (Eyes and ID Numbers are masked in order to keep the confidentiality)	19
Figure 3.3: Template image	20
Figure 3.4: Cropped Image	21
Figure 3.5: Name surroundings after the removal of printed characters	21
Figure 3.6: Generated line image.....	22
Figure 3.7: Scanned image after the line removal	23
Figure 3.8: Name surroundings after the removal of lines	23
Figure 3.9: After the closing operation.....	23
Figure 3.10: After the noise removal.....	24
Figure 4.1: Processed segment of hand written characters.....	25
Figure 4.2: Processed segment of printed characters.....	25
Figure 5.1: Result with binary image	26
Figure 5.2: Result with gray image.....	27
Figure 5.3 close with 'disk' (radius 2) morphological structuring element	27
Figure 5.4: close with 'line' (length 10, angle 90 degrees) morphological structuring element	27
Figure 5.5: close with 'line' (length 3, angle 90 degrees) morphological structuring element	28