COMPUTER AIDED ANALYSIS OF HANDWRITTEN DEVANAGARI DOCUMENTS FOR FORENSIC APPLICATIONS

NIVEDITA YADAV

AMARNATH AND SHASHI KHOSLA SCHOOL OF INFORMATION TECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY DELHI
SEPTEMBER 2014

COMPUTER AIDED ANALYSIS OF HANDWRITTEN DEVANAGARI DOCUMENTS FOR FORENSIC APPLICATIONS

by

NIVEDITA YADAV

Amarnath and Shashi Khosla School of Information Technology

Submitted

in fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2015

I dedicate this thesis :

To my parents for unconditional love and support

To my husband, Purushotham, without whose caring support it would not have been possible

Certificate

This is to certify that the thesis titled "**Computer Aided Analysis of Devanagari Documents For Forensic Applications**" being submitted by **Ms. Nivedita Yadav** to the Amarnath and Shashi Khosla School of Information Technology, Indian Institute of Technology, Delhi, for the award of the degree of Doctor of Philosophy, is a record of bona-fide research work carried out by her under our guidance and supervision. In our opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted to any other university or institute for the award of the degree or diploma.

Prof. Santanu Chaudhury
Professor
Department of
Electrical Engineering
Indian Institute of Technology Delhi
New Delhi - 110 016

Prof. Prem Kumar Kalra
Professor
Department of
Computer Science and Engineering
Indian Institute of Technology Delhi
New Delhi - 110 016

DATE
NEW DELHI

Acknowledgements

A doctoral degree is the highest point of formal education tree and I am thankful to all those institutions that have contributed towards my education. From Kendriya Vidyalaya Sainik Vihar, University of Delhi to National Institute of Criminology and Forensic Sciences and IIT Delhi. I am very fortunate to have performed my PhD work at IIT Delhi. Therefore, there are many people to thank for their part in my success. This journey would have not been possible without the love, guidance and support of many people. Each person has had a significant impact on my graduate career that I will not forget.

The research for this thesis was carried out in research group of Multimedia Lab at Indian Institute of Technology Delhi. The funding in the form of Junior research fellowship came from Directorate of Forensic Science Services (DFSS), Ministry of Home Affairs, Government of India. I am grateful for their support.

This thesis would have not been possible without the support of two crucial persons, my thesis supervisors Prof. Santanu Chaudhury and Prof. Prem Kumar Kalra for giving me this opportunity to carry out this research project under their guidance. Its been an honour to work with you both. Prof. Chaudhury's guidance and constructive criticisms helped me develop a better understanding of the subject. Prof. Kalra is incredibly organised and a great problem solver, both these qualities were helpful in moving this project forward. I appreciate their vast knowledge and skill in many areas and assistance in writing reports. Over the years, both have given me superb scientific guidance.

I must also acknowledge Prof. Sanjiva Prasad, Coordinator School of IT, for encouragement and support. I would also like to thank my thesis committee members, Prof. K. K. Biswas, Prof. Subodh Kumar and Dr. Sumantra Dutta Roy, for their contributions to this work. My thesis committee members guided and helped me to improve the manuscript of this thesis. I am thankful to them. I would like to thank Shri Mohinder Singh (Retired Director Government Examiner of Questioned Documents, Hyderabad) for his valuable suggestions on Primitive selection work. I would like to express my appreciation for Dr Rajesh Kumar (FSL Delhi), his insightful questions usually provoke interesting scientific discussions. He deserves many thanks for his continued interest and guidance throughout my work.

From the Multimedia lab, I would like to acknowledge Usha Mam and Deepti for all the administral help and support. I truly appreciate the time and help of all those colleagues who rolled up their sleeves for providing timely help. I am thankful to my colleagues, fellow PhD students, my juniors for all those helpful discussions and coffee breaks. I also enjoyed discussions with my fellow PhD students very much. The Multimedia lab members are a wonderful group of people who I have enjoyed working with and learning from. I am also thankful to Shri Rajesh (school of IT) for all the administral help.

I am grateful for having had opportunity to travel to conferences and learn from the leading experts in my field. My learning network consists of the people from the international research community with whom I interacted at conferences as well as of fellow PhD students from various disciplines who shared their best practices to stay productive and sane during the stressful years with me online through blogs and Twitter discussions. I am thankful to them for introducing pomodoro techniques and silva method

of mind relaxation. My sincerest gratitude and appreciation to them.

The individuals I have met in IITD that I consider friends are too numerous to name. There are a few, however, that cannot go unmentioned, I would specifically like to recognise Rajeshwari, Rachna, Pooja, Dipti, Rupi, Nisha, Sunita and Poornima. These friends have been there for me when the challenges seemed too great to overcome.

Finally, I would like to acknowledge friends and family who supported me during my time here. First and foremost I would like to thank my parents,Deepak, Gita, Deepika, Nannu and Keshu for constant love and support. I would not be who am I today without you all. Special thanks to my best friends Vydehi, Shelly, Sarita, Aashish, Abhinav, you all have been my best cheerleaders. Last but not least, I would like to thank my husband, Purushotham, for his patience, encouragement and unwavering love.

Nivedita Yadav

Abstract

The studies on handwriting are quite old. The variations in handwritings, both inter-writer and intra-writer have made this popular research area in the field of pattern recognition. This thesis addresses the problem of automated forensic handwriting analysis for handwritten Devanagari documents. All Indian scripts are very different from English. They are based on Devanagari script which is among the top five widely spoken languages in the world. We present a systematic examination of Devanagari documents for different forensic applications like forgery detection, writer recognition, writer verification, writer recognition using multiple feature combination, multiple word based forgery detection and discriminative primitive selection for Devanagari text.

We present a new database "DevLipi" on Devanagari for forensic analysis of handwriting. We have collected 10 pages from each writer and they were asked to write one page per day. This was done to capture as many variations as we can in the handwriting to study the inter-writer and intra-writer variations. We have a collection of 400,000 handwritten words.

We also propose a technique for selection of most discriminatory primitives of Devanagari script. We use feature selection method to select a set of primitives, which can help in better and faster discrimination of writers. Both identification and verification models can be used for the forensic analysis of handwriting. In our study, we have considered the handwritten documents which cover a large range of natural variations. Discrimination studies based on words and characters are reported for English,

Arabic etc in current literature. However, such discrimination of handwritten words based study is missing for Devanagari documents.

We have also investigated skilled forgery based on multiple words. In this study, we take samples from a group of subjects where we ask them to write words which contain discriminative primitives, in their natural handwriting. We also take another group of subjects where we ask them to forge the collected genuine handwriting documents to make a forged document that is convincingly similar. The collected handwriting samples are digitised and we extract word level HOG features and compute shape descriptors. By using these features we discriminate between authentic and forged handwriting samples using SVM classifier.

Finally, we present a novel MKL based identity determination framework which provides a feature combination framework for the forensic document anaysis. In this framework, set of novel features are introduced, which are optimally combined by the proposed MKL formulation in kernel space for improving the identification results. We consider both the aspects of identity determination, i.e., identification and verification. The proposed MKL formulation is applied for both the problems where a novel Genetic algorithm based solution is presented for optimizing the MKL formulation.

Contents

List of Figures

List of Tables

Glossary

FDE	Forensic Document Examination
CEDAR	Center of Excellence for Document Analysis and Recognition
EER	Equal Error Rate
FAR	False Acceptance Rate
GSC	Gradient, Structural Concavity
IBM	International Business Machine
KNN	K Nearest Neighbor
MKL	Multiple Kernel Learning
HOG	Histogram of Oriented Gradients
RBF	Radial Basis Function
RBF-SVM	Support Vector Machine with RBF kernel
ROC	Receiver Operating curve
ROI	Region of Interest
SFS	Sequential Forward Selection
SVM	Support Vector Machine
2-D	Two Dimensional
VSC	Video Spectral Comparator
SD	Shape Descriptor
FMT	Fourier Mellin Transform
EHOG	Envelope Histogram of Oriented Gradients
SC	Shape Context
DPI	Dots per inch
FMT	Fourier Mellin Transform
SC	Shape Context
CFT	Computational featureset