

EXPLOITING MULTIMEDIA CONTENT: A MACHINE LEARNING BASED APPROACH

by

EHTESHAM HASSAN

Department of Electrical Engineering

Submitted

in fulfillment of the requirements of the degree of

Doctor of Philosophy

to the



INDIAN INSTITUTE OF TECHNOLOGY DELHI

NOVEMBER 2012

Certificate

This is to certify that the thesis titled “**EXPLOITING MULTIMEDIA CONTENT: A MACHINE LEARNING BASED APPROACH**” being submitted by **EHTESHAM HASSAN** to the Department of Electrical Engineering, 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 him 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 any degree or diploma.

Prof. Madan Gopal

Department of Electrical Engineering

India Institute of Technology Delhi

New Delhi - 110016

Prof. Santanu Chaudhury

Department of Electrical Engineering

India Institute of Technology Delhi

New Delhi - 110016

To my family

Acknowledgements

First of all my ethereal indebtedness and recognition goes to the Almighty that He steers me in my whole walk of life.

The successful completion of this thesis is due to the mode of supervision, timely encouragement and efficient guidance received from my supervisors Professor Madan Gopal, and Professor Santanu Chaudhury. My deep appreciation and bounteous praise goes to my supervisors for their patience and assistance in settling my fuzzy ideas, confusions and short comings. The thesis would not achieved its present form without their continuous erudite help and conscientious support. My heartfelt thanks and regards to Santanu sir for all his financial support and freedom which never let me bother for any publication charge and other requirements. I acknowledge with gratitude the benevolence of my seniors, friends and the staff of Control lab and Multimedia lab who have been ever helpful and supportive during the course of this thesis. Last but not the least my benevolent gratitude, credit affirmation goes to my family members for their plunk for, plump for and scrupulous sustainability and unsparing and unstinted counselling, steering and guidance throughout the course of this research project and in whole walk of life.

Abstract

This thesis explores use of machine learning for multimedia content management involving single/multiple features, modalities and concepts. We introduce shape based feature for binary patterns and apply it for recognition and retrieval application in single and multiple feature based architecture. The multiple feature based recognition and retrieval frameworks are based on the theory of multiple kernel learning (MKL). A binary pattern recognition framework is presented by combining the binary MKL classifiers using a decision directed acyclic graph. The evaluation is shown for Indian script character recognition, and MPEG7 shape symbol recognition. A word image based document indexing framework is presented using the distance based hashing (DBH) defined on learned pivot centres. We use a new multi-kernel learning scheme using a Genetic Algorithm for developing a kernel DBH based document image retrieval system. The experimental evaluation is presented on document collections of Devanagari, Bengali and English scripts.

Next, methods for document retrieval using multi-modal information fusion are presented. Text/Graphics segmentation framework is presented for documents having a complex layout. We present a novel multi-modal document retrieval framework using the

segmented regions. The approach is evaluated on English magazine pages. A document script identification framework is presented using decision level aggregation of page, paragraph and word level prediction. Latent Dirichlet Allocation based topic modelling with modified edit distance is introduced for the retrieval of documents having recognition inaccuracies. A multi-modal indexing framework for such documents is presented by a learning based combination of text and image based properties. Experimental results are shown on Devanagari script documents.

Finally, we have investigated concept based approaches for multimedia analysis. A multi-modal document retrieval framework is presented by combining the generative and discriminative modelling for exploiting the cross-modal correlation between modalities. The combination is also explored for semantic concept recognition using multi-modal components of the same document, and different documents over a collection. An experimental evaluation of the framework is shown for semantic event detection in sport videos, and semantic labelling of components of multi-modal document images.

Table of Contents

Acknowledgements	iii
Abstract	v
List of Figures	xv
List of Tables	xix
1 Introduction	1
1.1 Scope and Objective	2
1.2 Major Contributions of the Thesis	7
1.3 Layout of the thesis	11
2 Learning for Multimedia Content Management: Revisited	15
2.1 Analysis of Multiple Feature based Recognition and Retrieval	16
2.1.1 Multiple Feature based Retrieval	19
2.2 Multimedia Database Indexing	20
2.3 Hashing based Indexing Schemes	22
2.4 Learning for Indexing and Ranking	25

2.5	Concept driven Content Management	27
2.6	Multimedia Retrieval and Analysis Systems	31
2.6.1	Document Image Retrieval	32
2.6.2	Video Analysis and Retrieval	35
2.7	Motivation for the Present Work	36
3	Multiple Features for Recognition of Binary Patterns	41
3.1	Introduction	41
3.2	Related Works	44
3.3	Feature Extraction	49
3.3.1	Fringe Map (FM)	49
3.3.2	Histogram of Oriented Gradients (HOG)	52
3.3.3	Shape Descriptor (SD)	53
3.3.4	Modified Shape Descriptor (MSD)	58
3.4	Multiple Kernel Learning for Character/Symbol Classification	59
3.4.1	Binary MKL Problem Formulation	61
3.4.2	DAG based Classifier Design	63
3.5	Experimental Evaluation and Discussion	65
3.5.1	Character/Primitive Recognition	65
3.5.2	Symbol Recognition	76
3.6	Conclusions	80
4	Word based Document Image Indexing and Retrieval	83

4.1	Introduction	83
4.1.1	Analysis of Feature Representations for Word Images	85
4.2	Overview of the Document Indexing Framework	89
4.3	Shape based Feature Representation for Word Images	91
4.3.1	Extension of Shape Descriptor for Word Image Representation	91
4.4	Distance based Hashing for Indexing	96
4.4.1	Distance based Hashing	97
4.4.2	Pivot Object Selection	99
4.4.3	Locality Sensitivity Analysis of Distance based Hashing Functions	100
4.4.4	Hierarchical DBH	103
4.5	Experimental Results and Discussion	105
4.6	Multi Probe Hashing in DBH Framework	116
4.6.1	Step-wise Multi-probing in Distance Based Hashing	117
4.6.2	Success Probability Estimation	118
4.6.3	Performance Evaluation	120
4.7	String like Word Representation for Document Image Indexing	122
4.7.1	Word Image Representation	124
4.7.2	Document Indexing using Edit distance based hashing	126
4.7.3	Experimental Evaluation	127
4.8	Conclusions	129

5 Learning for Document Image Indexing with Multiple Features 131

5.1	Introduction	131
5.2	Distance based Hashing in Kernel Space	133
5.2.1	Proposed Kernel based DBH	133
5.3	Multiple Kernel Learning for Hashing	136
5.3.1	Optimization Problem Formulation	137
5.3.2	Genetic Algorithm based Optimization Framework for Multiple Kernel Learning	139
5.3.3	Preliminary Evaluation with MNIST dataset	142
5.4	Document Image Indexing Using Combinations of Features	150
5.4.1	Feature Description	152
5.4.2	Retrieval Results	154
5.5	Conclusions	161
6	Multi-modal Information Integration for Document Retrieval	163
6.1	Introduction	163
6.2	Methods for Multi-modal Document Image Retrieval	166
6.2.1	Existing Text/Graphics Segmentation Methods for Document Analysis	166
6.2.2	Existing Script Identification Methods	168
6.2.3	Methods Addressing the Recognition Inaccuracies for Document Re- trieval	172
6.3	Separation Framework for Multi-coloured Text/Graphics	173
6.3.1	Scheme for Document Image Segmentation	174

6.3.2	Experimental Evaluation	181
6.3.3	Multi-modal Retrieval of Document Images having Embedded Graphics	183
6.4	Script based Segmentation of Document Image	185
6.4.1	Overall Framework	187
6.4.2	Features Extraction	190
6.4.3	Script Identification at Block Level	192
6.4.4	Script Identification at Word Level	197
6.4.5	Results and Discussion	198
6.5	LDA based Searching for OCR'ed Text	202
6.5.1	Overall Framework: Document Indexing and Retrieval	205
6.5.2	Details of Indexing the OCR'ed Documents	207
6.5.3	Experimental Validation	208
6.6	Word based Multi-modal Document Image Indexing	210
6.6.1	Experimental Evaluation	213
6.7	Conclusions	215
7	Concept Learning for Multimedia Content Handling	217
7.1	Introduction	217
7.2	MKL based Concept Learning	219
7.2.1	Feature Description	221
7.2.2	Annotation Model Architecture	225
7.2.3	Experimental Results	226

7.3 MKL for LSCOM Concept Recognition	229
7.4 MKL based Feature Combination for Concept driven Retrieval	230
7.4.1 Image Feature Description	231
7.4.2 MKL Details and Results	232
7.5 Multi-modal Concept linkage using Conditioned Topic Modelling	236
7.5.1 Conditioned Topic Learning for Multi-modal Retrieval	237
7.5.2 Experimental Results and Discussion	241
7.6 Multi-modal Concept Recognition	243
7.6.1 Proposed Event Detection Framework	245
7.6.2 Experimental Results	248
7.6.3 Concept Recognition of Multi-modal Document Images	253
7.7 Conclusions	255
8 Conclusions	257
8.1 Summary of the Contributions	258
8.2 Scope of Future Work	260
Bibliography	263
A Locality Sensitive Hashing	293
B Relevance Vector Machine for Classification	295
C Conditional Random Fields	301
D Latent Dirichlet Allocation	303

Publications	306
Biography	309