

Certificate

It is certified that the work contained in this thesis entitled “Enhancing the efficiency of text-image based tasks through multimodal solutions” by “Deepanwita Datta” is original and has been carried out under our supervision and that it has not been submitted elsewhere for a degree. It is further certified that the student has fulfilled all the requirements of Comprehensive Examination, Candidacy and SOTA for the award of Ph.D. Degree.

Thesis Supervisor

Professor Sanjay Kumar Singh

Dept. of Computer Science and
Engineering

Indian Institute of Technology (BHU),

Varanasi, India

Thesis Co-Supervisor

Dr. C Ravindranath Chowdary

Dept. of Computer Science and
Engineering

Indian Institute of Technology (BHU),

Varanasi, India

December 2017

DECLARATION BY THE CANDIDATE

I, Deepanwita Datta certify that the work embodied in this thesis is my own bonafide work and carried out by me under the supervision of Prof. Sanjay K. Singh and Dr. C Ravindranath Chowdary from December-2013 to December-2017, at the Department of Computer Science and Engineering, Indian Institute of Technology (BHU) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully lifted up any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, theses, etc., or available at websites and included them in this thesis and cited as my own work.

Date :

Signature of the Student

Place :

(Deepanwita Datta)

CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of my/our knowledge.

Signature of Supervisor

(Prof. Sanjay K. Singh)

Signature of Head of Department

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis : **Enhancing the efficiency of text-image based tasks through multimodal solutions**

Name of the Student : **Deepanwita Datta**

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University), Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the “Doctor of Philosophy”.

Date:

Signature of the student

Place:

(Deepanwita Datta)

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author’s personal use provided that the source and the Institute’s copyright notice are indicated.

ACKNOWLEDGEMENTS

“I can no other answer make, but, thanks, and thanks.”

–William Shakespeare

Thank you God for this life and its bountiful reserves whose fruits I continue to enjoy and cherish.

Firstly, I would like to express my sincere gratitude to my advisor Prof. Sanjay Kumar Singh for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. Besides my advisor, I would like to thank to my co-advisor Dr. C Ravindranath Chowdary and the rest of my thesis committee: Dr. R. S. Singh, Dr. B. Biswas and Dr. S. Das for their insightful comments and encouragement, but also for the hard question which incited me to widen my research from various perspectives.

My sincere thanks also goes to Dr. S. Pal, Prof. K. K. Shukla and Prof. A. K. Tripathi who provided their continuous support to me. I express my gratitude to Dr. R. P. Meena and Mr. Shashikant Singh who gave access to the laboratory and research facilities. Without their precious support it would not be possible to conduct this research.

I thank my fellow labmates for the stimulating discussions and for all the fun we have had in the last four years. In particular, I am grateful to Dr. D. Nandi for providing me the first glimpse of research. I am also thankful to the numerous anonymous reviewers, whose inputs and insights helped in shaping the better part of this thesis.

Last but not the least, I would like to thank my family: to my husband Mr. Manajit Chakraborty for being supportive at every endeavour, to my father Mr. Sukumar Datta and mother Mrs. Prathama Datta for their endless affection, sacrifice, inspiration and guidance through out the journey of my life, to my sister Ms. Anindita Datta for her lively encouragement and making me happy every time I felt down.

Deepanwita Datta

Contents

Certificate	ii
Declaration by the Candidate	iv
Copyright Transfer Certificate	vi
Abstract	vii
Acknowledgements	x
Contents	xii
List of Figures	xvii
List of Tables	xix
Abbreviations	xxi
Symbols	xxiii
1 Introduction	1
1.1 Information Retrieval	1
1.2 Image Retrieval	2
1.2.1 Text Based Image Retrieval	3
1.2.2 Content Based Image Retrieval	4
Drawbacks	5
1.2.3 Multimodal Image Retrieval	6
1.2.4 Preliminaries	8
1.2.4.1 Relevance Feedback	9
1.2.4.2 Query Expansion	10

1.2.4.3	Keyphrase Extraction	11
1.2.4.4	Fusion	12
1.2.5	Measuring Metrics	13
1.2.5.1	Recall	13
1.2.5.2	Precision	14
1.2.5.3	F-measure	15
1.2.5.4	MAP	15
1.3	Thesis Motivation and Objective	16
1.4	Contribution	19
1.5	Thesis Organization	22
2	Literature Review	23
2.1	Literature Gap	30
3	Multimodal Retrieval	33
3.0.1	Motivation	35
3.1	Baseline Framework	36
3.1.1	Preprocessing	36
3.1.1.1	Text	36
3.1.1.2	Image	38
3.1.2	Indexing and Matching Score calculation	38
3.1.3	Learning Combination Parameters	39
3.1.4	Ranking	42
3.2	Query Reformulation	43
3.2.1	tf-idf based Query Expansion Model (TQEM)	45
3.2.2	KEA based Query Expansion Model (KQEM)	46
3.2.3	Mutual Information based Query Expansion Model (MIQEM)	49
3.3	Theoretical Comparison of Keyphrase Extraction Models	54
3.4	Experimental Setup	57
3.4.1	Dataset	57
3.4.2	Specifications	58
3.4.3	Evaluation Measures	59
3.5	Results and Analysis	59
3.5.1	Comparison of various approaches	60
3.5.2	Comparison of image retrieval results	66
3.6	Discussion	69
4	Topic Model based Keyphrase Extraction	71
4.0.1	Methodology	72
4.0.1.1	Parameter Declaration	76
4.0.2	Experimental Results and Analysis	76

4.0.3	Discussion	80
5	Automatic annotation for image retrieval	83
5.0.1	Motivation	85
5.1	Methodology	87
5.1.1	Graph Construction	87
5.1.2	Graph Clustering	90
5.1.3	Ranking and extraction of annotations	91
5.2	Experimental Setup	94
5.3	Results and Analysis	95
5.4	Discussion	98
6	Multimodal Classification	99
6.1	Methodology	101
6.1.1	Average combination	101
6.1.2	Weighted average combination	102
6.1.3	Hill Climbing for Weight Assignment	102
6.1.4	Extended Hill Climbing for Weight Assignment	104
6.2	Experimental Setup	106
6.3	Results and Analysis	108
6.4	Discussion	110
7	Conclusions	111
	List of Publications	115
	Bibliography	117

List of Figures

1.1	A diagrammatic overview of a multimodal IR system ¹	7
3.1	Snapshot of Topic set	37
3.2	High-level diagram of proposed multimodal retrieval model	43
3.3	Per Query Precision Comparison for Set 1	63
3.4	Per Query Precision Comparison for Set 2	63
3.5	Per Query Precision Comparison for Set 3	64
3.6	Per Query Precision Comparison for Set 4	64
3.7	Per Query Precision Comparison for Set 5	65
3.8	Comparison of retrieved images	67
4.1	Per Query Precision Comparison for Set 1	78
4.2	Per Query Precision Comparison for Set 2	78
4.3	Per Query Precision Comparison for Set 3	79
4.4	Per Query Precision Comparison for Set 4	79
4.5	Per Query Precision Comparison for Set 5	80
5.1	Describing a <i>concept</i>	85

List of Tables

3.1	Comparative overview of Keyphrase Extraction Models	56
3.2	Optimal Weight of Combination Parameters	60
3.3	Average Precision per query set and MAP for queries without narration	61
3.4	Average Precision per query set and MAP for queries with narration .	62
3.5	Convergence Test	65
4.1	Average Precision per query set and MAP	77
5.1	Theoretical comparison of various models	96
5.2	Annotation Results Comparison	97
6.1	Statistics of PASCAL VOC'07 dataset	107
6.2	Per class comparison of various combination approaches	109

Abbreviations

IR	I nformation R etrieval
TBIR	T ext B ased I mage R etrieval
CBIR	C ontent B ased I mage R etrieval
MIR	M ultimodal I mage R etrieval
FLDA	F isher L inear D iscriminated A nalysis
QE	Q uery E xpansion
LDA	L atent D irichlet A llocation
TF	T erm F requency
IDF	I nverse D ocument F requency
TQEM	T F- I D F B ased Q uery E xpansion M odel
KQEM	K E A B ased Q uery E xpansion M odel
MIQEM	M utual I nformation B ased Q uery E xpansion M odel
TMQEM	T opic M odel B ased K eyphrase E xtraction
SVM	S upport V ector M achine
MKL	M ultiple K ernel L earning
LSR	L east S quares R egression
HC	H ill C limbing
EHC	E xtended H ill C limbing

Symbols

t	Title
T	Text
n	Narrative
n_p	Positive part of the narrative
q_T	Text query
V	Visual features
V_{cime}	Visual cime feature
V_{telp}	Visual telp feature
V_{Surf}	Visual Surf feature
\mathcal{S}_T	Matching score of each text query-document pair
$\mathcal{S}_{V_{cime}}$	Matching score of cime features
$\mathcal{S}_{V_{telp}}$	Matching score of each telp features
$\mathcal{S}_{V_{Surf}}$	Matching score of each Surf features
δ_T	Combination parameter of text
$\delta_{V_{cime}}$	Combination parameter of cime feature
$\delta_{V_{telp}}$	Combination parameter of telp feature
$\delta_{V_{Surf}}$	Combination parameter of Surf feature
z	Topic
ξ	Damping factor
φ, ν	The conjugate Dirichlet priors

*To my family for being my source of inspiration and
providing me the courage to weather the storms
gracefully...*

