

DOCTORAL THESIS

Automatic semantic image annotation and retrieval

Wong, Chun Fan

Date of Award:
2010

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Automatic Semantic Image Annotation and Retrieval

WONG Chun Fan

A thesis submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Principal Supervisor: Prof. Clement H.C. LEUNG

Hong Kong Baptist University

August 2010

Abstract

As the number of Web images is increasing at a rapid rate, searching them semantically presents a significant challenge. Many raw images are constantly uploaded with few meaningful direct annotations of semantic content, limiting their search and discovery. As definite relationships exist between the type of scenes and image acquisition parameters, image metadata and parametric dimensions, it is possible to determine from these the semantic content of images. By exploiting the judgment and expertise that has gone into the image-capture process, and through the use of decision trees and rule induction, we have established a set of rules which allows the semantic contents of images to be identified. When jointly applied with ontology-based and contextual-based expansion approaches, they are able to produce a new level of meaningful automatic image annotation, from which meaningful high-level semantic image searches are performed. The system is evaluated quantitatively using more than 100,000 Web images with 1,000,000 associated human tags outside the training database. Experimental results indicate that this approach is able to deliver highly competent performance, attaining good recall and precision rates of sometimes over 80%. This approach enables an advanced degree of semantic richness to be automatically associated with images which could previously only be performed manually.

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iii
Table of Contents	v
List of Figures	xii
List of Tables	xxiv
Chapter 1 Introduction	1
1.1 Motivation and Significance	2
1.2 Semantic gap	4
1.3 Outline of This Thesis	7

Chapter 2	Literature review	10
2.1	Introduction	11
2.2	Interaction Between Users and Image Retrieval Systems	14
2.2.1	Query Space	15
2.2.2	Query Specification	17
2.2.3	Query Space Display	18
2.3	Visual Similarity	19
2.3.1	Image Segmentation	21
2.3.1.1	Fixed Partitioning	22
2.3.1.2	Segmentation by Region and Object	24
2.3.1.3	Saliency-based Image Segmentation	28
2.3.2	Features Extraction	32
2.3.2.1	Colour Image Processing	33
2.3.2.2	Image Processing for Local Shape	35
2.3.2.3	Image Texture Processing	37
2.3.3	Features Matching	39

2.4	Semantic Similarity	41
2.4.1	Object Ontology Models	46
2.4.2	Knowledge Database Models	47
2.4.3	Concept Distance Models	47
2.4.4	Semantic Concept from Metadata Models	49
2.4.5	Explicit Relevance Feedback Models	51
2.4.6	Implicit Indexing Models	54
2.5	Applications of Automatic Image Annotation	56
2.5.1	Statistical Approaches	60
2.5.1.1	Frequent Keyword Mining Models	61
2.5.1.2	Hierarchical Aspect Models	62
2.5.1.3	Word Co-occurrence Models	64
2.5.1.4	Machine Translation Models	67
2.5.1.5	Cross Media Relevance Models	68
2.5.1.6	Continuous-space Relevance Models	69
2.5.1.7	Latent Dirichlet Allocation Models	70

2.5.1.8	Scene-Oriented Models	73
2.5.1.9	Robust Non-parametric Density Estimation Models	76
2.5.1.10	Multi-resolution Hidden Markov Models	77
2.5.2	Classification Approaches	78
2.5.2.1	Non-negative Matrix Factorisation Models	78
2.5.2.2	Support Vector Machine Models	80
2.5.2.3	Cluster-based Retrieval Models	84
2.6	Applications of Query-by-example	85
2.7	Performance Measures	87
2.8	Summary	89
Chapter 3 Automatic Semantic Annotation Algorithms		90
3.1	Introduction	91
3.2	Correlating Scene Characteristics with Image Dimensions and Features	92
3.2.1	Scenes of Image	92
3.2.2	Image Metadata and Exchangeable Image File Format	96
3.3	Rule-Induction and Annotation Algorithms	98

3.3.1	Collection of Real-world Web Images	98
3.3.2	Rule-Induction	101
3.4	Experimental Validation	104
3.5	Comparison to Automatic Image Annotation Benchmark	111
3.6	Summary	116
Chapter 4	Ontology-based Expansion	121
4.1	Index Scoring and Ranking	125
4.2	Computation of Annotation Weights	126
4.3	Expansion Algorithms	129
4.3.1	Semantic Knowledge from WordNet	129
4.3.2	WordNet Distance	130
4.3.3	Normalised Google Distance	132
4.3.4	Wikipedia Distance	133
4.3.5	Flickr Distance	134
4.4	Systems Evaluation	135
4.5	Summary	146

Chapter 5	Contextual-based Expansion	156
5.1	Global Colour Distribution	158
5.2	HSV Colour Space	162
5.3	Summary	173
Chapter 6	MPEG-7 Structured Annotation Description Scheme	175
6.1	Generation of MPEG-7 Descriptions	177
6.2	Summary	183
Chapter 7	Summary and Conclusion	185
7.1	Summary	186
7.2	Contributions	188
7.3	The Future of Semantic Image Annotation	189
7.4	Conclusions	190
	Bibliography	191
	Appendices	227

Chapter A	Statistical Data	227
A.1	Detailed performance of semantic relatedness measures of WLVM, NGD and WD algorithms for concept “downtown”	227
A.2	Detailed performance of semantic relatedness measures of WLVM, NGD and WD algorithms for concept “sunrise and sunset”	240
A.3	Detailed performance of semantic relatedness measures of WLVM, NGD and WD algorithms for concept “wild bird”	249
A.4	Detailed performance of semantic relatedness measures of WLVM, NGD and WD algorithms for concept “wild beast”	254
Curriculum Vitae		260
Publications		261