



THE UNIVERSITY OF ADELAIDE

School of Computer Science

Efficient and Robust Image Ranking for Object Retrieval

Yanzhi Chen

December, 2013

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE
FACULTY OF ENGINEERING, COMPUTER & MATHEMATICAL SCIENCES

ABSTRACT

This thesis focuses on efficient and effective object retrieval from an unlabeled collection of images. The goal of object retrieval is to, given a query image depicting an object, return the dataset images containing that same object, quickly and accurately. Due to its simplicity and efficiency, it is common to use a “Bag-of-Words” (BoW) model in which each image is represented as a weighted vector of quantised features, known as visual words. Although the BoW retrieval system is efficient, the extraction and quantisation of local image features introduces errors into the retrieval results.

We build our retrieval system on the BoW model, proposing three kinds of method to improve the retrieval accuracy: *i*) refinement of BoW image representation; *ii*) refinement of image similarity; *iii*) retrieval result re-ranking. Firstly, a *visual thesaurus* structure is proposed to discover the spatial relatedness of visual words. Based on these, a spatial expansion method is able to enrich the original query with those spatially related visual words (enriched by a general thesaurus) and spatially related foreground words (enriched by an object-based thesaurus). Therefore, the BoW image representation is improved.

The second contribution improves the standard image similarity used in the BoW retrieval system such that the similarity between query/dataset images is better described. We do this by a cross-word image matching scheme, such that matching features mapped to different visual words are able to contribute to the similarity score.

Thirdly, we also aim at efficient result re-ranking methods to improve the initial retrieval results. We present two re-ranking methods in this thesis. A context based re-ranking method is based on the analysis of correlated subsets of image dataset, called “contexts”. Images that share contexts are weakly correlated to each other, and should therefore mutually influence each other’s ranking. The initial ranking scores are refined by this contextual information. We also present a ranking verification method that is able to extract a set of reliable query relevant images from the retrieved results and thus can be applied in a number of object

retrieval applications. Note that neither method needs to recover low level feature information or prior knowledge from the dataset. Instead, they utilize ranking information during run time.

We also revisit the definition of the object retrieval problem and propose a group-query method, in which the query is a collection of images depicting the same object instead of a single query image used in the traditional “query-by-example” methods.

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed

Date

03-11-2013

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors Dr. Anthony Dick and Dr. Xi Li. This thesis would not have been possible without their guidance, enthusiasm, and encouragement in the past four years. Warm thanks to Anthony, who has dedicated a lot of time to my thesis writing in these months. My thanks also go to Rhys Hill, who helps in preparing codes, offers advice and encouragement.

I am very grateful to people in ACVT for many interesting discussions and shared experiences. This includes: John Bastian (the awesomest knight with a shining armor), Zygmunt Szpak, Lachlan Fleming, Paul Sakrapee Paisitkriangkrai, Daniel Pooley and Alex Chicowski. I also thank Xue Zhou, who has visited ACVT for one year, for providing suggestion on my research work and my life! I am also very grateful to China Scholarship Council and the University of Adelaide for jointly funding my PhD study.

I would like to express my gratitude to my parents, for their support and understanding, and quiet patience in the past four years. Special thanks go to Guoge Han for discussing ways of research and future career, although we study totally different research areas (Computer Vision *v.s.* Ophthalmology). My final thanks go to friends back in Shanghai: Yuki, Yide, Kelly and Shuang. Their continual encouragement endeavors my PhD study.

TABLE OF CONTENTS

List of Figures	v
List of Tables	ix
Notation	xiii
Chapter I: Introduction	1
1.1 Problem statement	1
1.2 Why is it difficult ?	1
1.3 Contributions	3
1.4 Thesis outline	4
1.5 Publications	5
Chapter II: Literature review	9
2.1 Content-based image retrieval	11
2.1.1 Description of visual content	12
Survey of global features	13
Survey of local features	14
2.1.2 Visual content analysis	15
Query-by-example retrieval	17
2.1.3 A basic CBIR system	18
2.2 Visual image retrieval built on “Bag-of-Words” model	19
2.2.1 Why quantisation of local features?	20
2.2.2 Visual vocabulary building	21
Nearest neighbor search based on visual words	21
Scaling clustering	23
2.2.3 Efficient indexing by using text-retrieval methods	25
2.2.4 The BoW retrieval system architecture	26
2.3 Datasets, evaluation and the baseline system	29
2.3.1 Evaluation datasets	29

2.3.2	Evaluation criterion	30
2.3.3	The baseline retrieval system	33
2.4	Algorithms for retrieval performance improvement	35
2.4.1	Improvement of local region detector	35
2.4.2	Improvement of local feature descriptor	36
2.4.3	Improvement of visual vocabulary	37
2.4.4	Improvement of BoW representation	38
	Soft-assignment in quantisation	39
	Adding spatial information in the BoW representation	39
	Query expansion	40
2.4.5	Improvement of similarity measure	43
	Improvement of accuracy	44
	Improvement of scalability	44
2.4.6	Retrieval performance summary	46
2.5	Application of object retrieval	47
2.5.1	Image dataset mining	47
2.5.2	3D reconstruction from retrieved images	49
2.6	Summary	49
Chapter III: Building a visual thesaurus		51
3.1	Building a visual thesaurus	52
3.2	Spatial expansion based on a general thesaurus	55
3.3	Experimental results	57
3.4	Discussion	62
3.5	Building an object-based thesaurus	64
3.5.1	Automatic training data collection	65
3.5.2	Association of words in thesaurus	66
3.5.3	Spatial expansion based on an object-based thesaurus	67
3.6	Conclusion	69
Chapter IV: Improving the image similarity measure		73
4.1	Visual word re-weighting based on an object-based thesaurus	75
4.1.1	The visual word re-weighting scheme	75
4.1.2	Experimental results	80

4.1.3	Discussion	84
4.2	Spatially aware feature selection and re-weighting	84
4.2.1	Experimental results	86
4.2.2	Discussion	91
4.3	A cross-word matching measure via visual thesaurus	92
4.3.1	A cross-word image similarity measure	95
	Visual word weights from distances (offline)	96
	Cross-word matching similarity (online)	96
4.3.2	Inter-word distance measure	98
	The L_2 word distance	98
	The spatial co-occurrence distance	99
	A semantic distance measure	99
4.3.3	Experimental results	104
4.3.4	Discussion	111
4.4	Conclusion	114
Chapter V:	Context based re-ranking for object retrieval	117
5.1	Context based re-ranking	119
5.1.1	Random space partition	120
5.1.2	Context factor for re-ranking	123
5.2	Experimental results	124
5.2.1	Parameter setting	124
5.2.2	Effects of query expansion	127
5.3	Discussion	128
5.4	Conclusion	129
Chapter VI:	Ranking consistency for image matching and object retrieval	133
6.1	Ranking consistency similarity	137
6.1.1	Ranking consistency measures	137
6.1.2	Result re-ranking using ranking consistency	139
6.1.3	Fast approximate ranking list computation with list-wise min-Hash	141
6.2	Experiments	144

6.2.1	Experimental results of ranking verification	144
6.2.2	Discussion	150
6.3	Application I: Query expansion with ranking verification	151
6.3.1	Shortlist generation and query expansion models	152
6.3.2	Experimental results	153
6.4	Application II: Discovery of dataset images	155
6.4.1	Experimental results	155
6.5	Conclusion	157
Chapter VII:	Object retrieval with group-query	159
7.1	Forming a group-query and training samples	162
7.2	Average group-query retrieval	163
7.3	Discriminative group-query retrieval	164
7.3.1	Discriminative ranking function with linear SVM	164
7.3.2	Discriminative ranking function with boosting	166
7.4	Experiments	169
7.4.1	Experimental setup	169
7.4.2	Experimental results	170
7.4.3	Group-query with ranking verification	177
7.5	Conclusion	177
Chapter VIII:	Conclusion	179
8.1	Contributions	179
8.2	Summary	180
8.3	Future work	183
Bibliography		185

LIST OF FIGURES

1.1	Example of object retrieval.	2
1.2	Examples of image condition changes	3
1.3	Examples of query images of Oxford buildings.	7
2.1	Text-based image retrieval results.	10
2.2	Example of a successful recognition of object boundaries.	14
2.3	Example of affine invariant regions detected by Harris-affine, Hessian-affine and MSER.	16
2.4	Illustration of quantisation in the BoW model.	21
2.5	The baseline framework of BoW based retrieval system.	28
2.6	Illustration of precision-recall (PR) curve.	32
2.7	Top retrieval results of the baseline method.	34
2.8	Example of connected image groups: All souls and Radcliffe camera.	48
3.1	System framework of spatial expansion.	53
3.2	Example of a pair of frequently appearing visual words found by a general thesaurus.	54
3.3	Examples of feature correspondences returned by spatial expansion F_1	56
3.4	Examples of qualitatively examining of spatially expanded words.	59
3.5	Precision-recall (PR) curves of spatial expansion based on various general thesaurus.	60
3.6	Illustration of the accuracy of spatially expanded words.	61
3.7	Top retrieval results of the spatial expansion method (F_5).	63
3.8	Comparison of matched visual words.	64
3.9	Training pairs detected with automatic selection on the Oxford 5K and Paris 6K datasets.	66
3.10	Illustration of retrieval accuracy of spatially expanded words (F') included by various threshold.	68
3.11	Comparison of various spatial expansion to the baseline.	70

3.12	Top retrieval results of the spatial expansion method (F_{15}).	71
4.1	Examples of quantisation errors in feature space.	74
4.2	System framework of visual word re-weighting.	76
4.3	Examples of the distribution of the neighborhood of visual words.	78
4.4	Examples of visual word that appears in multiple images of the same building.	79
4.5	Evaluation of the retrieval performance with various training data size and re-weighting type.	82
4.6	Illustration of tf-idf score adaption	83
4.7	Top retrieval results of the visual word re-weighting method.	85
4.8	Illustration of the total association scheme.	86
4.9	Comparison of various spatial expansion to the baseline.	88
4.10	Illustration of detailed precision-recall (PR) curves of total association.	89
4.11	Top retrieval results of total association scheme.	92
4.12	System framework of cross-word matching.	95
4.13	Example of the L2 word distance.	98
4.14	Examples of visual words for the same topic.	100
4.15	Examples of images transitively connected by the geometric information.	101
4.16	Examples of small connected component.	102
4.17	Example of large connected component.	103
4.18	Illustration of a pair of visual words close in the semantic distance measure.	104
4.19	Retrieval performance comparison of five types of visual distance.	107
4.20	Comparison of mAP scores on all the 55 queries on the Oxford 5K and Paris 5K datasets.	108
4.21	Precision-recall (PR) curves for various distance measurement (Oxford 5K).	109
4.22	Precision-recall curves (PR) for various distance measurement (Paris 6K).	110
4.23	Top retrieval results of total association scheme.	113
5.1	Illustration of context based re-ranking.	118
5.2	System framework of context based re-ranking.	119
5.3	Illustration of random space partition via inverted file.	121

5.4	Retrieval results comparison of random space dimension.	125
5.5	Top retrieval results of context based re-ranking.	128
5.6	Examples of precision-recall (PR) curves of context based re-ranking.	130
6.1	Ranking consistency overview.	134
6.2	System framework of ranking consistency.	134
6.3	Examples of similarity score computed by Jacarrd similarity and RBO similarity.	139
6.4	Examples of ranking consistency similarity for a particular query <i>All souls 1</i>	140
6.5	Comparison of random collisions in standard min-Hash and list-wise min-Hash.	143
6.6	Ranking verification accuracy with approximate near neighbor search.	146
6.7	Ranking verification accuracy with various list-wise (v) min-Hash. . .	146
6.8	Illustration of the number of verified images <i>v.s.</i> the number of true positives obtained by ranking verification and spatial verification. . .	154
6.9	Clustering results by different image matching methods.	156
6.10	The average coverage of dataset images.	157
7.1	Illustration of group object retrieval method.	160
7.2	System framework of visual word re-weighting.	161
7.3	Retrieval results of noisy group-query (averaging group-query). . . .	164
7.4	Retrieval results of noisy group-query (averaging positive training data).	164
7.5	Example of training data samples used in discriminative ranking function query.	165
7.6	Illustration of training data separation by a linear SVM classifier. . .	165
7.7	Illustration of training data separation by a boosting classifier. . . .	167
7.8	Retrieval results of noisy group-query (discriminative ranking of group-query).	168
7.9	Retrieval results of noisy group-query (discriminative ranking of positive training data).	168
7.10	Top-6 retrieved results of using the linear SVM ranking function and our boosting-like ranking function with respect to the object landmark <i>Magdalen</i>	169
7.11	Precision-recall (PR) curves of individual query <i>v.s.</i> group-query . . .	171

7.12 Retrieval performance with different numbers of high quality query instances.	171
7.13 Retrieval performance with different numbers of low quality query instances.	172
7.14 Illustration of group-query retrieval results, with $M = 4$ high quality query instance.	175
7.15 Illustration of query retrieval results, with $M = 4$ low quality query instances used as group-query.	176
8.1 Summary of our proposed methods in the BoW based retrieval system.	183

LIST OF TABLES

2.1	The Keeper dataset.	27
2.2	Inverted file for the Keeper dataset.	27
2.3	Evaluation dataset summary	29
2.4	Text queries of Oxford and Paris datasets.	31
2.5	Examples of "good" and "junk" images from the Oxford 5K dataset.	31
2.6	Query examples of Caltech Categories and ImageNet datasets.	32
2.7	Homography transformation types used in spatial verification.	41
2.8	Retrieval performance summary of improvement methods (mAP).	47
3.1	Retrieval performance of spatial expansion on the Oxford 5K dataset.	58
3.2	Retrieval performance evaluation of spatial expansion on the Paris 6K dataset.	59
3.3	Comparison of spatial expansion F_5 to the state-of-the-art methods.	62
3.4	Retrieval results of spatial expansion based on an object-based thesaurus.	68
3.5	Comparison of spatial expansion to the state-of-the-art methods.	69
4.1	Evaluation of different training data size in visual word re-weighting on the Oxford 5K dataset.	81
4.2	Evaluation of different training data size in visual word re-weighting on the Paris 6K dataset.	81
4.3	The retrieval results of total association.	88
4.4	Comparison with methods requiring spatial consistency examination.	90
4.5	Average run time of retrieval methods.	90
4.6	Comparison of our methods to those that modify the baseline before the query is executed.	91
4.7	Comparison of total association to the state-of-the-art methods.	93
4.8	Retrieval performance with five types of visual distance.	106
4.9	Retrieval results of our inter-word distance measure method.	111
4.10	Results of different distance measure fusion.	112

4.11	Retrieval performance of cross-word distance measure method to the state-of-the-art methods.	114
5.1	Retrieval performance with different types of random space partition.	125
5.2	Retrieval performance with different threshold on context based re-ranking	126
5.3	Performance for online and offline expansion on context based re-ranking.	127
5.4	Computational cost comparison of spatial verification and context based re-ranking.	127
5.5	Comparison of context based re-ranking to the state-of-the-art methods.	131
6.1	Summary of various ranking list generation.	145
6.2	Ranking verification performance comparison on five datasets.	145
6.3	Ranking verification performance with different number of re-ranked images.	146
6.4	Ranking verification accuracy with varying number of hash functions M and list-wise min-Hash v	148
6.5	Comparison of various similarity measures used in min-Hash scheme.	149
6.6	Average memory usage of various re-ranking methods.	149
6.7	Average run time of various re-ranking methods.	149
6.8	Comparison of ranking verification to the state-of-the-art methods.	151
6.9	Example of shortlist images generated by spatial verification (R1) and ranking verification (R4).	152
6.10	Retrieval performance of query expansion (QE) combined with ranking consistency method.	154
6.11	The average coverage of dataset images.	156
7.1	Comparison of (maximum) individual and group-query	170
7.2	Retrieval performance with different discriminative ranking functions.	173
7.3	Average re-ranking CPU time	173
7.4	Retrieval results comparison between group-query and query expansion methods.	173
7.5	Comparison of group query to the state-of-the-art methods.	174
7.6	Retrieval performance of group-query with ranking verification.	177

8.1	Retrieval results summary.	181
8.2	Summary of retrieval results combination.	182

NOTATION

VISUAL WORDS

\mathbf{W}	visual vocabulary
w_i	visual word i
N	Visual vocabulary size
\mathbf{D}_i	the nearest neighborhood word set of visual word w_i
\mathbf{Q}	query words
\mathbf{W}_T	spatial related words
\mathbf{Q}'	spatial expansion words
\mathbf{W}_S	foreground words

IMAGES

\mathbf{V}	image dataset
V	image dataset size
$\text{loc}(w_i)$	image space location of a visual word w_i

THE BAG-OF-WORD IMAGE REPRESENTATION

\mathbf{v}_d	the visual word set of image d
\mathbf{q}	tf-idf query vector
\mathbf{d}	tf-idf image vector
$\tau(w_i)$	tf-idf weight of visual word w_i
$v_j(w_i)$	frequency of visual word w_i occur in image j
$\Psi(q, d)$	dot product image similarity between image d and query q

VISUAL THESAURUS

H	histogram set of visual thesaurus
F	a general thesaurus
F'	an object based thesaurus
Θ	visual distance measure
$\alpha(w_i)$	refined weight of visual word w_i
$\Gamma(q, d)$	cross-word similarity between image q and query d

CONTEXT BASED RE-RANKING

\mathcal{S}	collection of visual word vectors
\mathcal{C}	image groups for context re-ranking
c	a rank context
Ω	context factor
S	query relative words
$\Phi(q, d)$	contextual re-ranking similarity between image q and query d

RANKING CONSISTENCY

$J(i, j, h)$	Jaccard similarity between image i and image j at depth h
$R(i, j, h, p)$	RBO similarity between image i and image j at depth h with weight p

GROUP-QUERY

\mathcal{Q}	group-query
r	ranking list of retrieval results
\mathcal{R}	ranking list set
\mathcal{T}	training sample set