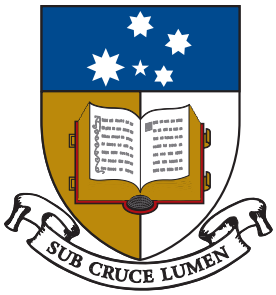

THE UNIVERSITY OF ADELAIDE

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

03 - 11 - 2013

Date

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

LIST OF TABLES

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
$\upsilon_j(w_i)$	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

$\mathbf{H}$	histogram set of visual thesaurus
$\mathbf{F}$	a general thesaurus
$\mathbf{F}'$	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

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

Ranking consistency

$J(i,j,h)$	Jacarrd 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

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