UNIVERSITI PUTRA MALAYSIA

SEMANTIC-BASED IMAGE RETRIEVAL FOR MULTI-WORD TEXT QUERIES

MOHSEN ZAND

FSKTM 2015 19

SEMANTIC-BASED IMAGE RETRIEVAL FOR MULTI-WORD TEXT QUERIES

By

MOHSEN ZAND

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

December 2015

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the Degree of Doctor of Philosophy

SEMANTIC-BASED IMAGE RETRIEVAL FOR MULTI-WORD TEXT QUERIES

By

MOHSEN ZAND

December 2015

Chair : Professor Shyamala A/P C. Doraisamy, PhD
Faculty : Computer Science and Information Technology

Catalyzed by the development of digital technologies, the amounts of digital images being produced, archived and transmitted are reaching enormous proportions. It is hence imperative to develop techniques that are able to index, and retrieve relevant images through user's information need. Image retrieval based on semantic learning of the image content has become a promising strategy to deal with these aspects recently. With semantic-based image retrieval (SBIR), the real semantic meanings of images are discovered and used to retrieve relevant images to the user query. Thus, digital images are automatically labeled by a set of semantic keywords describing the image content. Similar to the text document retrieval, these keywords are then collectively used to index, organize and locate images of interest from a database. Nevertheless, understanding and discovering the semantics of a visual scene are high-level cognitive tasks and hard to automate, which provide challenging research opportunities. Specifically, exploiting discriminatory features, handling the visual similarity between object classes and appearance diversity in each class, classification of low-level image visual features to appropriate semantic classes, comprehensively annotate images, and reliable indexing and ranking images through difficult queries are open issues to cope with. This study proposes new ideas to overcome these challenges.

First, a discriminatory image feature vector is generated using texture as a distinguishable visual feature. In the proposed method, the image texture which is extracted by the Gabor wavelet and the curvelet transforms in the spectral domain is encoded into polynomial coefficients. It not only provides rotation invariant features but also generates texture feature vectors with the maximum power of discrimination.

Second, a context-aware and semantic-consistent image descriptor is presented to exploit the image visual attributes in a contextual space. The high-level visual space is constructed by a Dirichlet process regardless of the semantic classes, and then, the posteriors are used to build the contextual space.

Third, the high-level feature similarities are employed to design a kernelized classification model which facilitates the reliable mapping of visual features to semantic concepts.

Fourth, with the proposed semantic label discovery and the kernelized classification model, more image annotations such as regional, subjective and latent labels are integrated for efficient image retrieval.

Finally, a probabilistic latent semantic indexing approach is proposed for a reliable multi-word image retrieval, where labeled images are represented by a finite mixture over latent topics. This structure enables multi-word querying and generates scalable indexing for ranked image retrieval based on the probability scores.

The effectiveness of the proposed approaches in the semantic-based image retrieval is demonstrated through comparisons in terms of precision and recall with state-of-the-art methods on the widely-used databases including ImageCLEF, MSRC, and others, which shows more efficient results. The validity of the overall SBIR with the whole components connected together is also evaluated on the SAIAPR TC-12 database, which obtains 26.8 in terms of mean average precision.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk Ijazah Doktor Falsafah

CAPAIAN IMEJ BERDASARKAN SEMANTIK BAGI KUERI TEKS PELBAGAI PERKATAAN

Oleh

MOHSEN ZAND

Disember 2015

Pengerusi : Profesor Madya Shyamala A/P C. Doraisamy, PhD
Fakulti : Sains Komputer dan Teknologi Maklumat

Bermangkinkan pembangunan pesat teknologi digital, jumlah imej digital yang terhasil, terarkib dan dihantar sudah menjadi terlalu banyak. Oleh yang demikian, teknik-teknik pengindeksan dan dapatan-semula imej perlu dihasilkan berdasarkan keperluan pengguna. Dapatan-semula berdasarkan pembelajaran semantik kandungan imej adalah satu strategi yang menjanjikan bagi menangani masalah ini. Dengan dapatan-semula imej berdasarkan semantic (SBIR), makna semantik sebenar imej dapat ditentukan dan digunakan bagi mendapatkan semula imej releven berdasarkan kueri pengguna. Ini membenarkan imej digital dilabel secara otometik dengan satu set kata kunci yang menerangkan kandungan imej berkenaan. Sama seperti dapatan semual dokumen teks, kata-kata kunci ini kemudian digunakan secara kolektif untuk tujuan pengindeksan, penyusunan semula dan pengesanan lokasi imej yang dikehendaki dari pangkalan data. Namun, memahami dan mengenalpasti semantik bagi scene visual merupakan satu tugasan pemikiran kognitif tahap tinggi dan sukar untuk diotomasi. Ini membuka ruang bagi peluang-peluang penyelidikan yang mencabar. Secara khusus, mengeksploitasi ciri-ciri pembezaan imej, pengendalian persamaan visual antara kelas-kelas objek dan kepelbagaian penampilan dalam setiap kelas, pengklasifikasian ciri-ciri visual tahap rendah imej kepada kelas semantik yang sesuai, anotasi imej secara komprehensif, dan pengindeksan dan pemberian kedudukan yang boleh dipercayai melalui kueri-kueri sukar, merupakan isu-isu terbuka untuk diatasi. Kajian ini mencadangkan ide-ide baharu untuk mengatasi cabaran-cabaran ini.

Pertamanya, satu vektor ciri imej yang membezakan, dijana menggunakan tekstur sebagai ciri visual. Dalam metod yang dicadangkan, ciri tekstur diekstrak menggunakan transformasi-transformasi wavelet Gabor dan curvelet dalam domain spektral dan kemudian dikod ke dalam bentuk pekali-pekali polinomial. Pendekatan ini bukan sekadar menawarkan ciri-ciri tidak berubah dari segi putaran, tetapi juga dapat menjana vektor-vektor ciri tekstur dengan kebolehupayaan pembezaan yang maksimum.

Kedua, satu penghurai yang sedar-konteks dan konsistent-semantik dibentangkan bagi mengeksploitasi sifat-sifat imej dalam ruangan kontekstual.

Ruangan visual tahap-tinggi ini dibina menggunakan satu proses Dirichlet, tanpa mengambil kira kelas semantik, dan kemudiannya posterioir-posterior digunakan bagi membina ruangan kontekstual.

Ketiga, persamaan-persamaan ciri tahap-tinggi digunapakai untuk merekabentuk satu model klasifikasi ber-kernel yang memudahkan pemetaan ciri-ciri visual kepada konsep-konsep semantik.

Keempat, dengan kaedah penemuan label-semantik dan model klasifikasi ber-kernel yang dicadangkan, anotasi-anotasi imej tambahan seperti pada rantauan, subjektif dan label-label terpendam dapat diintegrasi bagi dapatan semula imej yang cekap.

Akhirnya, satu pendekatan pengindeksan semantik terpendam bagi dapatan semula imej pelbagai-perkataan yang lebih efektif dicadangkan, di mana imej-imej terlabel diwakili oleh campuran terhingga melalui tajuk-tajuk terpendam. Struktur ini membenarkan kebolehan kueri pelbagai-perkataan dan menjana pengindeksan mampu-skala untuk dapatan semula imej berkedudukan berdasarkan markah kebarangkalian.

Keberkesanan pendekatan-pendekatan yang dicadangkan dibentangkan melalui perbandingan dengan metod-metod state-of-the-art sedia ada, berdasarkan terma-terma kepersisan dan penggilan balik. Pangkalan data imej digital terkenal digunakan iaitu ImageCLEF, MSRC dan lain-lain. Kesahihan keseluruhan komponen SBIR yang digabungkan juga dinilai menggunakan pangkalan data SAIAPR TC-12, di mana purata kepersisan pertengahan 26.8 diperolehi.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest praise and admiration to ALLAH, the most merciful, who has given me the strength, faith, confidence, and patience to complete this project despite all the challenges.

I must give my extremely regards and thanks to my parents who support me with their love and encouragement.

I owe my deepest gratitude to my supervisor, Dr. Shyamala Doraisamy, for sharing her interesting ideas, knowledge and experience, and for supporting me over the course of my studies. Her incredible enthusiasm, incisive criticism and keen eye for details have been a source of energy without which this work would not be the same. Her talent, diverse background, interest, teaching and research style have provided me an exceptional opportunity to learn.

I would also like to express my sincere thanks to the supervisory committee members, Dr. Alfian Bin Abdul Halin, and Dr. Mas Rina Binti Mustaffa, for their valuable suggestions and advices throughout this work. I'm grateful for careful comments on my writings, and for reviews of my papers and interesting discussions.

I certify that a Thesis Examination Committee has met on 10 December 2015 to conduct the final examination of Mohsen Zand on his thesis entitled "Semantic-Based Image Retrieval for Multi-Word Text Queries" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Fatimah binti Khalid, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Rahmita Wirza binti O. K. Rahmat, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Norwati Mustapha, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Guojun Lu, PhD
Professor
Federation University Australia
Australia
(External Examiner)

ZULKARNAIN ZAINAL, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 16 February 2016

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Shyamala A/P C. Doraisamy, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Alfian Bin Abdul Halin, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Mas Rina Binti Mustaffa, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

____
BUJANG BIN KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ________________________ Date: ___________________

Name and Matric No : Mohsen Zand GS36127

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ______________________________
Name of Chairman of Supervisory Committee: Shyamala A/P C. Doraisamy, PhD

Signature: ______________________________
Name of Member of Supervisory Committee: Alfian Bin Abdul Halin, PhD

Signature: ______________________________
Name of Member of Supervisory Committee: Mas Rina Binti Mustaffa, PhD

TABLE OF CONTENTS

				Page
ABSTRACT				i
ABSTRAK				iii
ACKNOWLEDGEMENTS				v
APPROVAL				vi
DECLARATION				viii
LIST OF TABLES				xii
LIST OF FIGURES				xiv
LIST OF ABBREVIATIONS				xvii
CHAPTER				
1	**INTRODUCTION**			1
	1.1	Semantic-based image retrieval		2
	1.2	Motivation		3
	1.3	Problem statement		5
	1.4	Objectives		6
	1.5	Contributions		6
	1.6	Thesis outline		7
2	**LITERATURE REVIEW**			9
	2.1	Feature extraction		9
		2.1.1	Color feature	10
		2.1.2	Shape feature	11
		2.1.3	Texture feature	11
	2.2	Image representation		14
		2.2.1	Representativeness issue	14
		2.2.2	Discrimination issue	14
		2.2.3	Quantization issue	16
	2.3	Automatic image annotation		17
		2.3.1	Multi-labelling annotation	17
		2.3.2	Image annotation as a machine translation task	18
	2.4	Semantic indexing and ranking		20
	2.5	Conclusion		21
3	**RESEARCH METHODOLOGY**			23
	3.1	Research overview		23
	3.2	Experimental design		25
	3.3	Test collection		25
	3.4	Evaluation parameters		27
	3.5	Conclusion		28
4	**TEXTURE CLASSIFICATION AND DISCRIMINATION FOR REGION-BASED IMAGE RETRIEVAL**			29
	4.1	Introduction		29
	4.2	Basic concepts		32
		4.2.1	Gabor wavelet filter	32
		4.2.2	Curvelet filter	33

		4.2.3 Fitting by polynomials	34
	4.3	Texture classification	35
	4.4	Results and Discussion	41
		4.4.1 Experiments on classifying textures in different rotations	41
		4.4.2 Experiments on the real image regions	44
	4.5	Conclusion	48
5	**CONTEXT-AWARE AND SEMANTIC-CONSISTENT IMAGE FEATURE EXTRACTION FOR AUTOMATIC IMAGE ANNOTATION**		49
	5.1	Introduction	49
	5.2	Visual space modelling	54
		5.2.1 A Dirichlet process for clustering visual space	54
		5.2.2 Variational inference for Dirichlet process	56
	5.3	Contextual space modelling	58
	5.4	Results and Discussion	61
	5.5	Conclusion	66
6	**SEMANTIC LABEL DISCOVERY FOR AUTOMATIC IMAGE ANNOTATION**		67
	6.1	Introduction	67
	6.2	High-level classification	70
	6.3	LDA model to enrich semantic labelling	77
	6.4	Results and Discussion	79
	6.5	Conclusion	83
7	**LATENT SEMANTIC LEARNING IN IMAGE RETRIEVAL FOR MULTI-WORD TEXT QUERIES**		84
	7.1	Introduction	84
	7.2	Latent semantic learning and indexing using RBM structure	87
	7.3	Inference and learning the model parameters	90
	7.4	Results and Discussion	92
		7.4.1 Experiments on ImageCLEF dataset	93
		7.4.2 Experiments on LabelMe dataset	94
	7.5	Conclusion	96
8	**CONCLUSION AND FUTURE WORK**		97
	8.1	Validation of the proposed methods in SBIR	97
	8.2	Summary and links to chapters	100
	8.3	Significant findings / contributions	100
	8.4	Recommendations	101
REFERENCES			102
BIODATA OF STUDENT			125
LIST OF PUBLICATIONS			126

LIST OF TABLES

Table		Page
3.1	The contingency table of retrieval results	27
4.1	The sub-band energies of the rotated textures in Figure 4.7 in one scale, and their rearranged energies	38
4.2	Coefficients of the fitted polynomials of the textures in Figure 4.7	39
4.3	Comparison of classification scores for different polynomial degrees	43
4.4	Comparison of classification scores in different rotation angles for different methods (best scores are indicated in bold)	43
4.5	The number of misclassified samples for each texture class and rotation angle for LBP and the proposed polynomial fitting method (P)	44
4.6	Classification rates for 20 different texture classes with polynomial Gabor, polynomial curvelet, fusion of Gabor and curvelet with polynomial and statistical curvelet methods	45
4.7	Comparison of discrimination power between different methods	47
5.1	The number of mixture components per visual feature for MSRC-21 (D1), and SAIAPR TC-12 (D2)	63
5.2	Annotation accuracy in the MSRC-21 (D1), and SAIAPR TC-12 (D2)	64
5.3	Average AUC in the MSRC-21 (D1), and SAIAPR TC-12 (D2)	65
6.1	$MAP_{annotation}$ results for different methods on the ImageCLEF dataset	81
6.2	Recognition rate for different methods	82
6.3	$MAP_{retrieval}$ results for the studied methods on the ImageCLEF dataset	83
7.1	MAP and BEP values of the studied methods for all single and multi-word queries on ImageCLEF database	94

7.2	MAP and BEP values of the studied methods for all single and multi-word queries on LabelMe database	95
8.1	MAP and BEP for the proposed method and Spyrou et al. (2014) on the SAIAPR TC-12 database	99

LIST OF FIGURES

Figure		Page
1.1	A general paradigm for SBIR	3
1.2	An alternative SBIR paradigm	3
3.1	The extended SBIR used as the research framework in this thesis	24
4.1	Pairs of different textures with similar statistical values obtained from curvelet sub-bands, which lead to the misclassification	31
4.2	The block diagram of the proposed method	32
4.3	Frequency spectrum coverage by a) Gabor b) Curvelet	34
4.4	a) A 'leaf' region image. b) One curvelet sub-band computed from image (a). c) Fitted polynomial of (b). d) Function and coefficients of the polynomial in (c)	35
4.5	a) An image region. b) Texture extracted by mirror padding method with wrong reflected edges in the texture. c) Improved mirror padding method	36
4.6	a) Regular shaped region. b) 40 Gabor wavelet sub-bands. c) Curvelet sub-bands in 4 scales	37
4.7	A texture image in the three rotation angles a) 0°, b) 15° and c) 30°	38
4.8	a) Two different image regions representing 'grass'. b) Polynomial coefficients of the region textures in (a)	40
4.9	Texture samples of the Outex_TC_00010 from Outex database (Timo Ojala & Maenpaa, 2002)	42
4.10	Comparison of classification rates for different texture classes in different methods	46
4.11	Image regions which are correctly classified based on their textures	47
4.12	Average retrieval results from ImageCLEF for different methods	48
5.1	Ambiguity in the visual words, (a) Visual diversity causes ambiguous visual words which are not representative of	51

the semantic classes, and (b) The lack of spatial context causes ambiguous visual words which are not discriminant.

5.2 The overall framework of the proposed CASC feature extraction 53

5.3 An image and its high-level visual features 56

5.4 (a) Co-occurrences of visual features. The bigger circles are more probable to be occurred concurrently. This information is captured from the images in the training set that include ‗sheep'. (b) features in (a), which shows higher probability of co-occurrences of the white color and fur texture with green color and rough texture. The visual feature naming is only for illustration purposes as they are actually integer values in our model. 60

5.5 Integrating visual context into the CASC descriptor 61

5.6 (a) High-level visual classes for the color feature in the new visual space, obtained from the MSRC-21 dataset. (b) High-level color extraction assigns the index of the closest color. It is achieved by the posterior probabilities. 63

5.7 Influence of the neighborhood level on the annotation accuracy 64

5.8 Annotation results on (a) MSRC-21 dataset, and (b) SAIAPR TC-12 dataset. The wrongly annotated tags are depicted by red color. The missed labels that supposed to be in the ground truth are shown with green color. 65

6.1 Similar images which surprisingly share similar labels 69

6.2 The framework of the proposed image annotation method 70

6.3 Two datasets separated by a hyperplane in the binary SVM classifier 72

6.4 The generative process of the LDA model 78

6.5 Annotation results on different images of the ImageCLEF dataset 82

7.1 a) The plate representation of the graphical model for standard LDA, and b) for the proposed method. Each node is a random variable, and the observed nodes are dashed. 88

7.2	Extension of the model in Figure 7.1 (b) using RBM structure for a vocabulary size 6, with 3 latent topics, and 2 images. The observed nodes are dashed.	88
7.3	Top 10 retrieved images for a, b, c) two-word query d) three-word query e) four-word query. The marked images are relevant to all the query words.	93
7.4	Precision-Recall graphs for (a) single-word queries and (b) multi-word queries	94
7.5	Latent topics from the multi-word query, -sea-building-sky ‘	96
8.1	Texture extraction. a) Original image, b) segmented image, c) a region of interest from (b) for texture extraction, d) extracted feature from (c), and e) polynomial coefficients of (d) as feature vector	97
8.2	Annotation results for different images of SAIAPR TC-12 database	98
8.3	Retrieved results for multi-word queries in the proposed method and Spyrou et al. (2014)	99

LIST OF ABBREVIATIONS

AIA	Automatic Image Annotation
AP	Average Precision
AUC	Area Under the ROC Curve
BEP	Break-Even Point
BoW	Bag-of-visual Words
CASC	Contextual-Aware and Semantic-Consistent
CBIR	Content-Based Image Retrieval
CCV	Color Coherence Vector
CLD	Color Layout Descriptor
CLP	Correlated Label Propagation
CM	Color Moments
CMRM	Cross Media Relevance Model
CRF	Conditional Random Field
CSD	Color Structure Descriptor
DCD	Dominant Color Descriptor
DCT	Discrete Cosine Transform
DoG	Difference of Gaussians
DVP	Descriptive Visual Phrase
DVW	Descriptive Visual Word
EFD	Elliptic Fourier Descriptor
EM	Expectation Maximization
FD	Fractal Dimension
FN	False Negative

FP	False Positive
GLCM	Grey Level Co-occurrence Matrix
GMM	Gaussian Mixture Model
HDIALR	Hidden-concept Driven Iimage Annotation and Label Ranking
HOG	Histogram of Oriented Gradient
IR	Information Retrieval
KFD	Kernel Fisher Discriminant
K-NN	K-Nearest Neighbor
LDA	Latent Dirichlet Allocation
LSA	Latent Semantic Analysis
MAP	Mean Average Precision
MBRM	Multiple Bernoulli Relevance Model
MCMC	Markov Chain Monte Carlo
MD	Minimum Description
MLL	Multi-Labeling Learning
MRF	Markov Random Fields
pLSA	Probabilistic Latent Semantic Analysis
RBF	Radial Basis Function
RBIR	Region-Based Image Retrieval
RBM	Restricted Boltzmann Machine
RBM-LDA	Restricted Boltzmann Machine-Latent Dirichlet Allocation
SBIR	Semantic-Based Image Retrieval
SCD	Scalable Color Descriptor
SIFT	Scale Invariant Feature Transform
SPM	Spatial Pyramid Matching

SSL	Shared-Subspace Learning
SV	Super Vector
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TBIR	Text-Based Image Retrieval
TF-IDF	Frequency-Inverse Document Frequency
TN	True Negative
TP	True Positive
tr-mmLDA	topic-regression multi-modal Latent Dirichlet Allocation
VLAD	Vector of Locally Aggregated Descriptors

CHAPTER 1

INTRODUCTION

“Image is everything” – This is a commercial promotional statement that shows images are becoming the “universal languag Internet. An image captures a mood, an ind in time, therefore image is everything, and very complex entity. The popularity of digital cameras and online communities have given individuals the privilege to capture their worlds in pictures, and suitably share them with others. Undoubtedly, both technology push and application pull have caused the rapid increase in the amount of digital images being produced, archived and transmitted. Confronted by this profusion of images, searching for required information from image documents is a crucial need whereby the more images are available, the more difficult to locate accurate and relevant information similar to written documents.

The practice of archiving written documents can be traced back to the time when Greek and Roman scholars started to compile the various sorts of the data. They found that organizing the data would facilitate the process of retrieving certain passages. A small slip was attached to papyrus scrolls with the work title and the name of the author. The idea of using computers to search for relevant pieces of information was born by Vannevar Bush in the article entitled “As We May Think” in 1945 (Singhal, 2001). Today, this is called “information retrieval” (IR), which is th documents from a collection of document resources. Unlike the data retrieval, the information is not structured in IR systems. They provide a set of methods and techniques for retrieving users’ infor systems usually assign index terms of keywords to the documents, and retrieve ranked items based on their relevancy to the user query keywords.

In written text documents, text information is used for the efficient retrieval, whether documents are organized manually or automatically such as modern digital libraries. However, organizing images manually outperforms machine as unlike text that is semantic, images are carriers of visual objects and scenes, and hence, they are concrete descriptions of which are relatively elusive. Usually, the interpretation of what we see is hard to describe, and even harder to teach a machine. In the past few years, a plenty of researches have been conducted by even more ambitious attempts to make computers understand, index and annotate images efficiently (Ritendra, Joshi, Li, & Wang, 2008). Earlier approaches to utilizing computers in image retrieval used text retrieval techniques on the textual descriptions of images. Nevertheless, images do not have full-text data, and manual annotations are inherently incomplete as translating some images in words is unfeasible.

To organize digital images by their visual content, content-based image retrieval (CBIR) was introduced in the early 1980s (Y. Liu, Zhang, Lu, & Ma, 2007). It includes computer vision technologies to support efficient image searching and browsing on the basis of automatically derived imagery features.

However, relying on the low-level visual features which are readily extracted from the images causes an inconsistency with the high-level image abstractions. This is semantic gap, a well-known challenge in CBIR research community, which is a key obstacle. To alleviate this problem, image retrieval based on semantic learning has recently gained more attention. Semantic-based image retrieval (SBIR) (Jaimes & Chang, 2000; Luo, Savakis, & Singhal, 2005; H.-C. Yang & Lee, 2008) approach tries to discover semantics from the image content. Therefore, the real semantic meaning of images can be detected and searched regardless to the low-level features. In this study, we focus on the enhancing the SBIR to be applicable on the real-world images by investigating underlying issues and aspects of this emerging technology.

1.1 Semantic-based image retrieval

Although there is no simple answer to the question of how search will evolve in the future, but it is absolutely predictable that search would become increasingly graph-based and semantic-based. Google has taken an important step towards the future of search by providing the knowledge graph. It is simply a machine that tries to think like a human when it encounters a massive database of information and facts on the Internet to bring more relevant information to the users. On the other hand, semantic search focuses on the user intention, the contextual relevance and the real meaning of the data to return relevant results. Clearly, the aim of both graph-based and semantic search is to improve the search algorithms to produce relevant, accurate and qualitative results with users‘ Using natural language processing and sophisticated artificial intelligence approaches, semantic search technology is becoming the core of text retrieval systems. However, the current improvement in semantic-based image retrieval is still in the research phase.

The existing methods are usually based on a general framework to describe the image content in different levels. Jaimes & Chang (2000) presented an SBIR scheme consisting of five levels of region level, perceptual region level, object part level, object level, and scene level. Particularly, three layers of abstraction are distinguishable in images; raw data, feature and semantic layers. The raw data shows images in the form of a matrix of pixels. The feature layer describes the key characteristics of the pixel patterns. The semantic layer is about meaning of detected objects in images. A general paradigm of such multi-layer approach is shown in Figure 1.1, where the object characterization plays an important role with two important tasks, i.e., extracting meaningful regions and identification of interesting objects (Luo et al., 2005). However, identification of image objects for object detection is an extremely difficult task, and latent semantics which are not extractable from visual appearance are not considered in this scheme. It also does not take into account the ranking of the results. An alternative general framework is presented in (Spyrou et al., 2014) which uses semantic mapping to show the similar images to the user, based on a text query. It is illustrated in Figure 1.2.

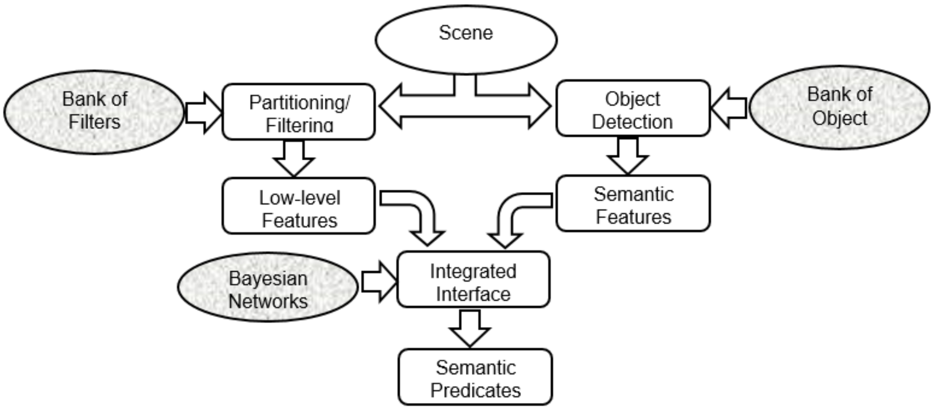

Figure 1.1. A general paradigm for SBIR (Luo et al., 2005)

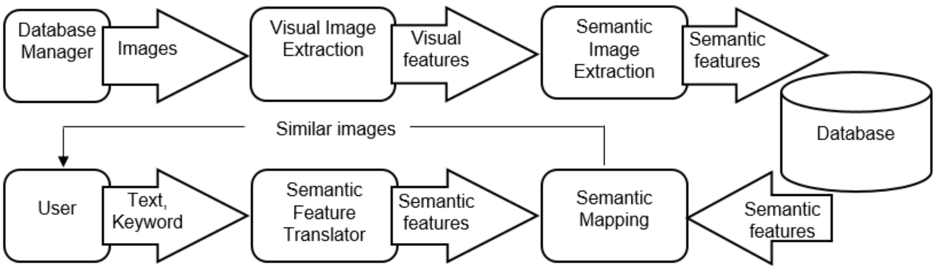

Figure 1.2. An alternative SBIR paradigm (Spyrou et al., 2014)

However, to teach the computer to really know the semantics of an image, it is promising to consider the way humans learn image semantics. Human beings actively search for key changes in a large section of the visual field, which cannot be achieved without visual attention. For rapid visual classification of new natural scenes, very little or no focal attention is required while semantic understanding of images needs features to be automatically registered in parallel across the visual field.

1.2 Motivation

Although SBIR has appeared as a solution for image retrieval with the aim of satisfying user‘s information needs, understanding the semantics within images is a high-level cognitive task, and very tough to automate. Many systems have been proposed and implemented using SBIR technologies, e.g., shared-subspace learning (SSL) (Ji, Tang, Yu, & Ye, 2010), hidden-concept driven image annotation and label ranking (HDIALR) (Bao, Li, & Yan, 2012), and multifaceted indexing (Fauzi & Belkhatir, 2013). Recent studies in SBIR have mainly focused on automatic image annotation (AIA) to associate low level image features with high level semantics through machine learning techniques (Amiri & Jamzad, 2015; Bannour & Hudelot, 2014; Y. Yuan, Wu, Shao, & Zhuang, 2013; S. Zhang, Tian, Hua, Huang, & Gao, 2014). AIA enables effective indexing and retrieval of images in large scale image databases, where manual image labeling is labor-intensive task. Primary AIA approaches grouped images into different classes using image features, and then annotated each image with a class concept. Therefore, each image was labelled with only one concept while natural images often contain several semantic concepts. This is a challenging problem in SBIR where each image can cover multiple semantic concepts. A considerable solution is region-based image retrieval (RBIR) which treats an image as a bag of regions by segmenting the image into primitive regions (Ying Liu, Dengsheng Zhang, 2008; D. Zhang, Islam, & Lu, 2012; D. Zhang, Monirul Islam, & Lu, 2013). The regional features are then matched the concept model for the image annotations. To distinguish visual regions separately, selecting discriminatory features is an essential factor. Most image retrieval systems apply three well-known color, shape and texture features. Among these low level image features, texture has shown to be effective and objective in image retrieval as it can distinguish regions with similar colors and shapes. Therefore, extracting texture features consistent with human perceptual intuitions of the objects inherently enhances the overall performance.

Although employing the local region descriptors for AIA is remarkable, and detection of the meaningless image regions is relatively simple, the performance of the system is limited due to the degrading the semantic integration. This is because by the projection of the 3D real scenes to the 2D images, only one view of the object appearances is captured. Hence, separating the semantic inference from the visual detection to reduce their confusions, and to boost each phase with the related information will be reasonably robust.

To annotate an image, most existing methods generate a visual codebook by grouping the low-level features into a predefined number of clusters, treat the center of each cluster as a visual word, and then annotate an unseen image by finding the closest entry in the codebook with the extracted features of the image (Bolovinou, Pratikakis, & Perantonis, 2013; Gemert, Snoek, Veenman, Smeulders, & Geusebroek, 2010; Lei Wu, Hoi, & Yu, 2010). This is called Bag-of-visual Words (BoW) model (Csurka, Dance, Fan, Willamowski, & Bray, 2004), which is inspired by the traditional bag-of-words model for text processing. However, AIA can be formulated as a multi-label classification problem where an image is associated with multiple labels. Therefore, specializing the classification model to incorporate the high-level image features will be a promising solution as the domain specific knowledge positively affects the performance of the image classification. Given the visual features of the image regions, the probabilities that they belong to different classes can be computed and ranked. The high ranked classes constitute the image annotations.

By the auto-annotation, the image retrieval issue is turned into a text retrieval task, which can efficiently benefit from the robustness and reliability of the traditional text retrieval approaches (Bannour & Hudelot, 2014; Z. Ma, Nie, Yang, Uijlings, & Sebe, 2012; Maier, Kwasnicka, & Stanek, 2012; Ntalianis, Tsapatsoulis, Doulamis, & Matsatsinis, 2012; Y. Yang et al., 2012; Zagoris, Chatzichristofis, Papamarkos, & Boutalis, 2010). However, the image annotations can even be enhanced by incorporating hidden semantic labels. In this situation, discovering the dependencies between individual image keywords and the latent semantic spaces which are not evident can improve the conceptual understanding of the images. Moreover, generating an index structure over these latent semantics, which can support complex and multi-word querying, and returning a ranked list of results will be promising for SBIR.

1.3 Problem statement

In spite of significant advances in SBIR, there are still some challenging open problems. In this thesis, main difficulties of SBIR that need to be tackled are investigated separately in five research issues as follows.

1) One research direction of SBIR is to explore discriminant features. Particularly, texture feature can overcome the limitations of color and shape features and determine the class a region belongs to. Two robust approaches in literature to model texture features are Gabor and curvelet features (Pang, Choi, & Qin, 2013; D. Zhang, Islam, Lu, & Sumana, 2012). However, although both features are close to human visual perception, they capture a large volume of unnecessary information which reduces their distinguishing power in texture classification. Therefore, sufficient information must be extracted from their sub-bands for efficient texture classification. A discriminant texture descriptor also needs to be scale- and rotation-invariant.

2) Though the BoW model is a preferred image representation technique (C.-F. Tsai, 2012; Lei Wu et al., 2010), the semantic ambiguity limits its effectiveness for SBIR. Moreover, generating discriminative and representative visual words for the high-dimensional image data is hard to achieve due to the visual similarity between object classes, and appearance diversity in each class.

3) Existing methods (Y. Liu et al., 2007; C. Zhang, Liu, Tian, Liang, & Huang, 2013; D. Zhang, Islam, & Lu, 2012; X. Zhou, Yu, Zhang, & Huang, 2010) usually assign a set of tags to each image using general kernel-based classification models. Nevertheless, since these models are built to be generally applicable in different applications, they do not satisfy the specific representations of images. Hence, the classification accuracy is degraded. Additionally, different image features pose different representations with differing weights, and thus, simply using a classifier with a single kernel for entire image feature space is not reasonable.

4) Although images can be described by the regional labels extracted from low-level regional features (Y. Wang, Mei, Gong, & Hua, 2009; Ying Liu,

Dengsheng Zhang, 2008; D. Zhang et al., 2013), there are some subjective labels such as names of the places that cannot directly be modeled. In addition, even with the collection of the regional and subjective labels, the image annotation is not descriptive enough.

5) AIA enables indexing and ranking of images by the text retrieval approaches. However, subjectivity of the text descriptions can potentially lead to classic information retrieval problems, namely polysemy and synonymy (Golder & Huberman, 2006). In addition, users tend to send multi-word queries to clarify their information needs. Therefore, developing an efficient indexing scheme over image keywords, which considers semantic meaning of the images, and enables multi-word querying is an open issue in SBIR.

1.4 Objectives

The major aim of this study is to propose new algorithms and methods to cover relevant aspects and issues of image retrieval in SBIR. The research objectives of this thesis can hence be summarized as follows:

● To propose a new texture feature vector with the maximum power of discrimination which is scale- and rotation-invariant, and it can enhance the precision and recall for the retrieval purposes.

● To propose innovative high-level image descriptors which convey semantics in the contextual space. These descriptors need to be discriminative and representative of the image appearances in the high-level visual space such that their application for image annotation results in a high annotation results in terms of the area under the ROC curve (AUC) and annotation accuracy.

● To design an especial kernel-based classification model using high-level feature similarities. This model must exploit the domain knowledge of the image features, and control the data scatter with a high recognition rate.

● To propose a novel automatic image annotation method with high precision and recall, which leverages three types of regional, subjective and latent labels to significantly enrich image description by incorporating semantic labels.

● To propose a novel indexing structure for efficient image retrieval, which enables latent topic extraction and allows multi-word querying. It needs to generate a scalable indexing for ranked image retrieval based on the probability scores. It can outperform the existing methods in terms of mean average precision (MAP) and break-even point (BEP).

1.5 Contributions

The contributions of this thesis are innovative approaches that address the above research issues as follows:

● A new discriminatory texture descriptor that uses polynomial coefficients to encode texture information. It can increase the texture classification rate and produce a high discrimination power in the scale- and rotation-invariant texture classification.

● A new image descriptor called contextual-aware and semantic-consistent (CASC) descriptor. Specifically, this image descriptor can exploit co-occurrence relationships in a contextual space which itself is modeled in a high-level visual space. Therefore, CASC descriptors can be visually representative of the image patches/regions, and contextually discriminative of the semantic classes, whereby their application for the AIA can generate a high annotation accuracy.

● A kernelized classification model that uses a new high-level kernel based on the high-level image feature similarities. It can leverage the specific domain knowledge of the image information in the classification process with a high recognition rate.

● A new image annotation method that benefits from different types of regional, subjective, and latent labels. Incorporating these semantic labels can significantly enhance the precision and recall of the image annotation.

● A novel indexing method called RBM-LDA (Restricted Boltzmann Machine-Latent Dirichlet Allocation) method that can index images based on uncovered latent topics and word dependencies. It enables multi-word querying with high precision and recall for the ranked image retrieval.

1.6 Thesis outline

This thesis is organized as follows. Chapter 2 provides the literature review of related works. Research methodologies are discussed in chapter 3. The contributions of this study are presented as four research works in chapter 4, 5, 6, and 7, entitled as:

● Chapter 4: Texture classification and discrimination for region-based image retrieval,

● Chapter 5: Context-aware and semantic-consistent image feature extraction for automatic image annotation,

● Chapter 6: Semantic label discovery for automatic image annotation, and

● Chapter 7: Latent semantic learning in image retrieval for multi-word text queries.

The first research problem of exploring discriminant features is the main considered issue in chapter 4. A texture classification and discrimination approach is proposed in this chapter. The second research problem of generating discriminative and representative visual descriptors is addressed in chapter 5 by proposing CASC descriptors. The presented research work in chapter 6 deals with the third and fourth research problems, i.e., classification specialization, and comprehensive annotation. The last research problem, i.e., incorporating latent semantic of images in the indexing structure for multi-word queries is investigated in chapter 7 where RBM-LDA is proposed. Finally, chapter 8 concludes the thesis and discusses future work.

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Susstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(11), 2274–2282.

Agarwal, A., & Triggs, B. (2006). Hyperfeatures–multilevel local coding for visual recognition. In *Proc. of Computer Vision–ECCV 2006, Springer Berlin Heidelberg* (pp. 30–43).

Akata, Z., Perronnin, F., Harchaoui, Z., & Schmid, C. (2014). Good practice in large-scale learning for image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(3), 507–20. http://doi.org/10.1109/TPAMI.2013.146

Al-Absi, H. R. H., Samir, B. B., Alhersh, T., & Sulaiman, S. (2013). On the combination of wavelet and curvelet for feature extraction to classify lung cancer on chest radiographs. In *Proc. of 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 3674–7). http://doi.org/10.1109/EMBC.2013.6610340

Altman, N. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Strltistician*, *46*(3), 175–185. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/00031305.1992.10475879

Amiri, S. H., & Jamzad, M. (2015). Automatic image annotation using semi-supervised generative modeling. *Pattern Recognition*, *48*(1), 174–188. http://doi.org/10.1016/j.patcog.2014.07.012

Arivazhagan, S., Ganesan, L., & Subash Kumar, T. G. (2006). Texture classification using curvelet statistical and co-occurrence features. In *Proc. of 18th International Conference on Pattern Recognition (ICPR'06)* (pp. 938–941). Ieee. http://doi.org/10.1109/ICPR.2006.1110

Bannour, H., & Hudelot, C. (2014). Building and using fuzzy multimedia ontologies for semantic image annotation. *Multimedia Tools and Applications*, *72*(3), 2107–2141. http://doi.org/10.1007/s11042-013-1491-z

Bao, B.-K., Li, T., & Yan, S. (2012). Hidden-concept driven multilabel image annotation and label ranking. *IEEE Transactions on Multimedia*, *14*(1), 199–210. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6035980

Barla, A., Odone, F., & Verri, A. (2003). Histogram intersection kernel for image classification. In *Proc. of 2003 International Conference on Image Processing (ICIP)* (Vol. 3, pp. III–513). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1247294

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, *110*(3), 346–359. http://doi.org/10.1016/j.cviu.2007.09.014

Blei, D. M. (2012). Probabilistic Topic Models. *Communications of the ACM*, *55*(4), 77–84. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Introduction+to+Probabilistic+Topic+Models#0

Blei, D. M., & Jordan, M. I. (2003). Modeling annotated data. In *Proc. of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval* (pp. 127–134). New York, New York, USA: ACM Press. http://doi.org/10.1145/860458.860460

Blei, D. M., & Jordan, M. I. (2004). Variational methods for the Dirichlet process. In *Proc. of the twenty-first international conference on Machine learning. ACM* (pp. 12–20). New York, New York, USA: ACM Press. http://doi.org/10.1145/1015330.1015439

Blei, D. M., & Jordan, M. I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, *1*(1), 121–143. http://doi.org/10.1214/06-BA104

Blei, D., & McAuliffe, J. (2008). Supervised topic models. In *Proc. of Neural Information Processing Systems (NIPS)* (pp. 1–8). Retrieved from http://papers.nips.cc/paper/3328-supervised-topic-models

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, *3*, 993–1022. Retrieved from http://dl.acm.org/citation.cfm?id=944937

Bolovinou, a., Pratikakis, I., & Perantonis, S. (2013). Bag of spatio-visual words for context inference in scene classification. *Pattern Recognition*, *46*(3), 1039–1053. http://doi.org/10.1016/j.patcog.2012.07.024

Bosch, A., Zisserman, A., & Munoz, X. (2006). Scene classification via pLSA. In *Proc. of Computer Vision–ECCV 2006, Springer Berlin Heidelberg* (pp. 517–530).

Boser, E., Guyon, I. M., & Vapnik, N. (1992). A training algorithm margin for optimal classifiers. In *Proc. of the fifth annual workshop on Computational learning theory* (pp. 144–152).

Bouachir, W., Kardouchi, M., & Belacel, N. (2009). Improving bag of visual words image retrieval: A fuzzy weighting scheme for efficient indexation. In *Proc. of Fifth International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* (pp. 215–220). Ieee. http://doi.org/10.1109/SITIS.2009.43

Bouguila, N., & Ziou, D. (2011). A countably infinite mixture model for clustering and feature selection. *Knowledge and Information Systems*, *33*(2), 351–370. http://doi.org/10.1007/s10115-011-0467-4

Boutell, M. R., Luo, J., Shen, X., & Brown, C. M. (2004). Learning multi-label scene classification. *Pattern Recognition*, *37*(9), 1757–1771. http://doi.org/10.1016/j.patcog.2004.03.009

Boutemedjet, S., Bouguila, N., & Ziou, D. (2009). A hybrid feature extraction selection approach for high-dimensional non-Gaussian data clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(8), 1429–1443. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4540103

Burghouts, G. J., & Geusebroek, J.-M. (2009). Performance evaluation of local colour invariants. *Computer Vision and Image Understanding*, *113*(1), 48–62. http://doi.org/10.1016/j.cviu.2008.07.003

Camps-Valls, G., Gomez-Chova, L., Muñoz-Marí, J., Vila-Francés, J., & Calpe-Maravilla, J. (2006). Composite kernels for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, *3*(1), 93–97. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4587630

Candes, E., Demanet, L., Donoho, D., & Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, *5*(3), 861–899. Retrieved from http://epubs.siam.org/doi/abs/10.1137/05064182X

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 1–27. http://doi.org/10.1145/1961189.1961199

Chang, S., Sikora, T., & Puri, A. (2001). Overview of the MPEG-7 Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, *11*(6), 688–695.

Chaudhuri, B., & Sarkar, N. (1995). Texture segmentation using fractal dimension. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(1), 72–77. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=368149

Comaniciu, D., & Meer, P. (2002). Mean S feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(5), 603–619.

Comaniciu, D., Meer, P., & Foran, D. J. (1999). Image-guided decision support system for pathology. *Machine Vision and Applications*, *11*(4), 213–224. http://doi.org/10.1007/s001380050104

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed, Vol. 35). Sage publications.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.

Csurka, G., Dance, C. R., Fan, L., Willamowski, J., & Bray, C. (2004). Visual categorization with bags of keypoints. In *Proc. of Workshop on statistical learning in computer vision (ECCV 2004)* (pp. 1–22).

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proc. of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.

de Diego, I. M., Serrano, Á., Conde, C., & Cabello, E. (2010). Face verification with a kernel fusion method. *Pattern Recognition Letters*, *31*(9), 837–844. http://doi.org/10.1016/j.patrec.2009.12.030

Deerwester, S., Dumais, S. T., Furnas, G. W., & Landauer, T. K. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.

Deng, Y., & Manjunath, B. S. (2001). Unsupervised segmentation of colour-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(8), 800–810.

Do, M. N., & Vetterli, M. (2002). Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*, *11*(2), 146–58. http://doi.org/10.1109/83.982822

Do, M. N., & Vetterli, M. (2003). The finite ridgelet transform for image representation. *IEEE Transactions on Image Processing*, *12*(1), 16–28. http://doi.org/10.1109/TIP.2002.806252

Duda, R., Hart, P., & Stork, D. (1995). *Pattern classification and scene analysis*. *John Willey & Sons* (2nd ed). New York. Retrieved from http://www.svms.org/classification/DuHS95.pdf

Duygulu, P., Barnard, K., Freitas, J. de, & Forsyth., D. A. (2002). Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proc. of Computer Vision—ECCV 2002, Springer Berlin Heidelberg* (pp. 97–112).

Escalante, H. J., Hernández, C. a., Gonzalez, J. a., López-López, a., Montes, M., Morales, E. F., … Grubinger, M. (20
annotated IAPR TC-12 benchmark. *Computer Vision and Image Understanding*, *114*(4), 419–428. http://doi.org/10.1016/j.cviu.2009.03.008

Fakheri, M., Amirani, M. C., & Sedghi, T. (2011). Gabor wavelets and GVF functions for feature extraction in efficient content-based colour and texture images retrieval. In *Proc. of 7th Iranian Conference on Machine

Vision and Image Processing (pp. 1–5). Ieee. http://doi.org/10.1109/IranianMVIP.2011.6121598

Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D., & Equitz, W. (1994). Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, *3*(3-4), 231–262. http://doi.org/10.1007/BF00962238

Fan, W., & Bouguila, N. (2013). Visual scenes clustering using variational incremental learning of infinite generalized Dirichlet mixture models. In *Proc. of The International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 24–29). Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Visual+Scenes+Clustering+Using+Variational+Incremental+Learning+of+Infinite+Generalized+Dirichlet+Mixture+Models#0

Fauzi, F., & Belkhatir, M. (2013). Multifaceted conceptual image indexing on the world wide web. *Information Processing & Management*, *49*(2), 420–440. http://doi.org/10.1016/j.ipm.2012.08.001

Fei-Fei, L., Fergus, R., & Perona, P. (2003). A bayesian approach to unsupervised one-shot learning of object categories. In *Proc. of Ninth IEEE International Conference on Computer Vision* (pp. 1134–1141). Ieee. http://doi.org/10.1109/ICCV.2003.1238476

Fei-Fei, L., & Perona, P. (2005). A bayesian hierarchical model for learning natural scene categories. In *Proc. of 2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Vol. 2, pp. 524–531). Ieee. http://doi.org/10.1109/CVPR.2005.16

Feng, S. L., Manmatha, R., & Lavrenko, V. (2004). Multiple bernoulli relevance models for image and video annotation. In *Proc. of 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (p. 1002). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1315274

Feng, Y., & Lapata, M. (2010). Topic models for image annotation and text illustration. In *Proc. of 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Human Language Technologies)* (pp. 831–839). Retrieved from http://portal.acm.org/citation.cfm?id=1858124

Fergus, R., Fei-Fei, L., Perona, P., & Zisserman, A. (2005). Learning object categories from Google's image search. In *Proc. of Tenth IEEE International Conference on Computer Vision (ICCV'05)* (pp. 1816–1823 Vol. 2). Ieee. http://doi.org/10.1109/ICCV.2005.142

Fergus, R., Perona, P., & Zisserman, A. (2003). Object class recognition by unsupervised scale-invariant learning. In *Proc. of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 258–264).

Flickner, M., Sawhney, H., Niblack, W., A
Yanker, P. (1995). Query by image and video content: The QBIC system. *Computer*, *28*(9), 23–32.

Freund, Y., & Haussler, D. (1994). *Unsupervised learning of distributions on binary vectors using two layer networks*. *Computer Research Laboratory [University of California, Santa Cruz]*.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139. http://doi.org/10.1006/jcss.1997.1504

Gaussier, E., & Yvon, F. (2013). *Textual information access: statistical models*. *John Wiley & Sons, Inc.* Wiley-IEEE Press. http://doi.org/10.1002/9781118562796

Gemert, J. C. Van, Snoek, C. G. M., Veenman, C. J., Smeulders, A. W. M., & Geusebroek, J. (2010). Comparing compact codebooks for visual categorization. *Computer Vision and Image Understanding*, *114*(4), 450–462. http://doi.org/10.1016/j.cviu.2009.08.004

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*(2), 198–208. http://doi.org/10.1177/0165551506062337

Gomes, R., Welling, M., & Perona, P. (2008). Incremental learning of nonparametric Bayesian mixture models. In *Proc. of 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–8). Ieee. http://doi.org/10.1109/CVPR.2008.4587370

Gómez, F., & Romero, E. (2011). Rotation invariant texture characterization using a curvelet based descriptor. *Pattern Recognition Letters*, *32*, 2178–2186. http://doi.org/10.1016/j.patrec.2011.09.029

Gönen, M., & Alpaydın, E. (2011). *The* Multipl
Journal of Machine Learning Research, *12*, 2211–2268. Retrieved from http://dl.acm.org/citation.cfm?id=2021071

Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing* (2nd ed.). Prentice Hall.

Grangier, D., & Bengio, S. (2008). A discriminative kernel-based approach to rank images from text queries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(8), 1371–1384. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4359384

Green, P. J., & Richardson, S. (2001). Modelling heterogeneity with and without the Dirichlet process. *Scandinavian Journal of Statistics*, *28*(2), 355–375.

Guillaumin, M., Mensink, T., Verbeek, J., & Schmid, C. (2009). Tagprop:

Discriminative metric learning in nearest neighbor models for image auto-annotation. In *Proc. of 2009 IEEE 12th International Conference on Computer Vision* (pp. 309–316). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5459266

Gupta, A., & Jain, R. (1997). Visual information retrieval. *Communications of the ACM*, *40*(5), 70–79. Retrieved from http://dl.acm.org/citation.cfm?id=253798

Gurram, P., & Kwon, H. (2013). Contextual SVM using Hilbert space embedding for hyperspectral classification. *IEEE Geoscience and Remote Sensing Letters*, *10*(5), 1031–1035. Retrieved from http://adsabs.harvard.edu/abs/2013IGRSL..10.1031G

Han, J., & Ma, K.-K. (2007). Rotation-invariant and scale-invariant Gabor features for texture image retrieval. *Image and Vision Computing*, *25*(9), 1474–1481. http://doi.org/10.1016/j.imavis.2006.12.015

Hannah, L. A., Blei, D. M., & Powell, W. B. (2011). Dirichlet process mixtures of generalized linear models. *The Journal of Machine Learning Research*, *1*, 1–33.

Hasanat, M. H. A., Ramachandram, D., & Mandava, R. (2010). Bayesian belief network learning algorithms for modeling contextual relationships in natural imagery: a comparative study. *Artificial Intelligence Review*, *34*(4), 291–308. http://doi.org/10.1007/s10462-010-9176-8

He, X., King, O., Ma, W., Li, M., Zhang, H., & Member, S. (2003). Learning a semantic space from user‘s relevance fee *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(1), 39–48.

Heisterkamp, D. R. (2002). Building a latent semantic index of an image database from patterns of relevance feedback. In *Proc. of 2002 IEEE International Conference on Pattern Recognition (ICPR)* (Vol. 4, pp. 134–137). IEEE Comput. Soc. http://doi.org/10.1109/ICPR.2002.1047417

Hernández-Gracidas, C. A., Sucar, L. E., & Montes-y-Gómez, M. (2011). Improving image retrieval by using spatial relations. *Multimedia Tools and Applications*, *62*(2), 479–505. http://doi.org/10.1007/s11042-011-0911-1

Hervé, N., & Boujemaa, N. (2007). Image annotation: which approach for realistic databases? In *Proc. of 6th ACM international conference on image and video retrieval* (pp. 170–177). Retrieved from http://dl.acm.org/citation.cfm?id=1282310

Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 50–57). Stockholm, Sweden: Morgan Kaufmann Publishers Inc.

Huang, J., Kumar, S. R., Mitra, M., Zhu, W.-J., & Zabih, R. (1997). Image indexing using color correlograms. In *Proc. of 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 762–768).

Huiskes, M., Thomee, B., & Lew, M. (2010). New trends and ideas in visual concept detection: the MIR flickr retrieval evaluation initiative. In *Proc. of the 11th ACM International Conference on Multimedia Information Retrieval* (pp. 527–536). Retrieved from http://dl.acm.org/citation.cfm?id=1743475

Ishwaran, H., & James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, *96*(453), 161–173. http://doi.org/10.1198/016214501750332758

Islam, M., Zhang, D., & Lu, G. (2009). Rotation invariant curvelet features for texture image retrieval. In *Proc. of IEEE International Conference on Multimedia and Expo (ICME)* (Vol. 1, pp. 562–565). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5202558

Jaimes, A., & Chang, S. (2000). A conceptual framework for indexing visual information at multiple levels. *S&T/SPIE Internet Imaging*, *3964*, 2–15.

Jain, A. K., & Vailaya, A. (1996). Image retrieval using color and shape. *Pattern Recognition*, *29*(8), 1233–1244. http://doi.org/10.1016/0031-3203(95)00160-3

Jegou, H., Perronnin, F., Douze, M., Sanchez, J., Perez, P., & Schmid, C. (2012). Aggregating local image descriptors into compact codes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(9), 1704–1716. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6104058

Jeon, J., Lavrenko, V., & Manmatha, R. (2003). Automatic image annotation and retrieval using cross-media relevance models. In *Proc. of 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval* (pp. 119–126). Retrieved from http://dl.acm.org/citation.cfm?id=860459

Ji, S., Tang, L., Yu, S., & Ye, J. (2010). A shared-subspace learning framework for multi-label classification. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *4*(2), 1–29. http://doi.org/10.1145/1754428.1754431

Kang, F., Jin, R., & Sukthankar, R. (2006). Correlated label propagation with application to multi-label learning. In *Proc. of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (Vol. 2, pp. 1719–1726). Ieee. http://doi.org/10.1109/CVPR.2006.90

Ke, X., Li, S., & Cao, D. (2011). A two-level model for automatic image annotation. *Multimedia Tools and Applications*, *61*(1), 195–212.

http://doi.org/10.1007/s11042-010-0706-9

Ke, X., Li, S., & Chen, G. (2013). Real web community based automatic image annotation. *Computers & Electrical Engineering*, *39*(3), 945–956. http://doi.org/10.1016/j.compeleceng.2012.09.017

Kekre, H., & Sarode, T. (2010). Two-level vector quantization method for codebook generation using Kekre‘ s proportionate erro *International Journal of Image Processing*, (1), 1–11. Retrieved from http://www.researchgate.net/publication/42831780_Two-level_Vector_Quantization_Method_for_Codebook_Generation_using_Kekeres_Proportionate_Error_Algorithm/file/d912f512dce4a1d0b2.pdf

Kelly, D. (2009). Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends® in Information Retrieval*, *3*(1—2), 1–224. http://doi.org/10.1561/1500000012

Keren, D., & Gotsman, C. (1999). Fitting curves and surfaces with constrained implicit polynomials. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(1), 31–41. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=745731

Kherfi, M. L., Ziou, D., & Bernardi, A. (2001). *Web Image Search Engines : A Survey*.

Kumar, M. P., Ton, P. H. S., & Zisserman, A. (2005). Obj Cut. In *Proc. of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 1, pp. 18–25). http://doi.org/10.1109/CVPR.2005.249

Lacoste-Julien, S., Sha, F., & Jordan, M. (2009). DiscLDA: Discriminative learning for dimensionality reduction and classification. In *Proc. of Neural Information Processing Systems (NIPS)* (pp. 1–8). Retrieved from http://papers.nips.cc/paper/3599-disclda-discriminative-learning-for-dimensionality-reduction-and-classification

Lanckriet, G. R. G., Cristianini, N., Bartlett, P., Ghaoui, L. El, & Jordan, M. I. (2004). Learning the kernel matrix with semidefinite programming. *The Journal of Machine Learning Research*, *5*, 27–72.

Larochelle, H., & Bengio, Y. (2008). Classification using discriminative restricted Boltzmann machines. In *Proc. of 25th international conference on Machine learning (ICML '08)* (pp. 536–543). New York, New York, USA: ACM Press. http://doi.org/10.1145/1390156.1390224

Larochelle, H., & Mandel, M. (2012). Learning algorithms for the classification restricted boltzmann machine. *Journal of Machine Learning Research*, *13*, 643–669. Retrieved from http://dl.acm.org/citation.cfm?id=2188407

Lazebnik, S., Schmid, C., & Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proc. of*

2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) (Vol. 2, pp. 2169–2178). Ieee. http://doi.org/10.1109/CVPR.2006.68

Li, J., & Wang, J. Z. (2008). Real-time computerized annotation of pictures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(6), 985–1002.

Li, L., Jiang, S., & Huang, Q. (2012). Learning hierarchical semantic description via mixed-norm regularization for image understanding. *IEEE Transactions on Multimedia*, *14*(5), 1401–1413. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6185680

Li, Y., Zhang, S., & Hu, J. (2010). Combining curvelet transform and wavelet transform for image denoising. In *Proc. of Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence* (Vol. 6216, pp. 317–324). Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-14932-0_40

Liu, C., & Wechsler, H. (2002). Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, *11*(4), 467–76. http://doi.org/10.1109/TIP.2002.999679

Liu, D., Hua, G., Viola, P., & Chen, T. (2008). Integrated feature selection and higher-order spatial feature extraction for object categorization. In *Proc. of 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–8). Ieee. http://doi.org/10.1109/CVPR.2008.4587403

Liu, L., Long, Y., Fieguth, P. W., Lao, S., & Zhao, G. (2014). BRINT: binary rotation invariant and noise tolerant texture classification. *IEEE Transactions on Image Processing*, *23*(7), 3071–84. http://doi.org/10.1109/TIP.2014.2325777

Liu, Y., Zhang, D., Lu, G., & Ma, W.-Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, *40*(1), 262–282. http://doi.org/10.1016/j.patcog.2006.04.045

Liu, Y., & Zhou, X. (2004). Extracting texture features from arbitrary-shaped regions for image retrieval. In *Proc. of 2004 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1891–1894). Ieee. http://doi.org/10.1109/ICME.2004.1394628

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proc. of the Seventh IEEE International Conference on Computer Vision* (pp. 1150–1157). Ieee. http://doi.org/10.1109/ICCV.1999.790410

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. http://doi.org/10.1023/B:VISI.0000029664.99615.94

Luo, J., Savakis, A. E., & Singhal, A. (2005). A Bayesian network-based framework for semantic image understanding. *Pattern Recognition*, *38*, 919–934. http://doi.org/10.1016/j.patcog.2004.11.001

Ma, W.-Y., & Manjunath, B. S. (1999). NeTra: A toolbox for navigating large image databases. *Multimedia Systems*, *7*(3), 184–198. http://doi.org/10.1007/s005300050121

Ma, Z., Nie, F., Yang, Y., Uijlings, J. R. R., & Sebe, N. (2012). Web image annotation via subspace-sparsity collaborated feature selection. *IEEE Transactions on Multimedia*, *14*(4), 1021–1030. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6146459

Mackay, D. J. C. (2003). *Information theory , inference , and learning algorithms*.

Magalhães, J., & Rüger, S. (2006). Logistic regression of generic codebooks for semantic image retrieval. In *Proc. of Image and Video retrieval, Springer Berlin Heidelberg* (pp. 41–50).

Magalhães, J., & Rüger, S. (2010). An information-theoretic framework for semantic-multimedia retrieval. *ACM Transactions on Information Systems (TOIS)*, *28*(4), 1–32. http://doi.org/10.1145/1852102.1852105

Maier, O., Kwasnicka, H., & Stanek, M. (2012). Image auto-annotation with automatic selection of the annotation length. *Journal of Intelligent Information Systems*, *39*(3), 651–685. http://doi.org/10.1007/s10844-012-0207-6

Maji, S., Berg, A., & Malik, J. (2008). Classification using intersection kernel support vector machines is efficient. In *Proc. of 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–8). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4587630

Makadia, A., Pavlovic, V., & Kumar, S. (2010). Baselines for image annotation. *International Journal of Computer Vision*, *90*(1), 88–105. Retrieved from http://link.springer.com/article/10.1007/s11263-010-0338-6

Manjunath, B., & Ma, W. (1996). Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*(8), 837–842. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=531803

Mao, Q., Tsang, I. W.-H., & Gao, S. (2013). Objective-guided image annotation. *IEEE Transactions on Image Processing*, *22*(4), 1585–97. http://doi.org/10.1109/TIP.2012.2233490

Marszałek, M., & Schmid, C. (2006). Spatia of features. In *Proc. of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (Vol. 2, pp. 2118–2125). Ieee. http://doi.org/10.1109/CVPR.2006.288

Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, *22*(10), 761–767. http://doi.org/10.1016/j.imavis.2004.02.006

Materka, A., & Strzelecki, M. (1998). *Texture analysis methods-a review*. *Technical University of Lodz, institute of Electronics* (Vol. COST B11). Retrieved from http://www.eletel.p.lodz.pl/programy/cost/pdf_1.pdf

Mccallum, A., Freitag, D., & Pereira, F. (2000). Maximum entropy Markov models for information extraction and segmentation. In *Proc. of International Conference on Machine Learning (ICML)* (pp. 591– 598).

Mehrotra, R., & Gary, J. E. (1995). Similar-shape retrieval in shape data management. *Computer*, *9*(September), 57–62.

Mei, Y., & Androutsos, D. (2008). Wavelet-based color texture retrieval using the independent component color space. In *Proc. of 15th IEEE International Conference on Image Processing (ICIP)* (pp. 165–168). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4711717

Mezaris, V., Kompatsiaris, I., & Strintzis, M. G. (2003). An ontology approach to object-based image retrieval. In *Proc. of 2003 IEEE International Conference on Image Processing (ICIP)* (pp. 5–11). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1246729

Mohamadzadeh, S., & Farsi, H. (2013). Colour and texture feature-based image retrieval by using Hadamard matrix in discrete wavelet transform. *IET Image Processing*, *7*(3), 212–218. http://doi.org/10.1049/iet-ipr.2012.0203

Mojsilovic, A., Hu, J., & Soljanin, E. (2002). Extraction of perceptually important colors and similarity measurement for image matching, retrieval and analysis. *IEEE Transactions on Image Processing*, *11*(11), 1238–48. http://doi.org/10.1109/TIP.2002.804260

Mokhtarian, F., & Abbasi, S. (2002). Shape similarity retrieval under afine transforms. *Pattern Recognition*, *35*(1), 31–41.

Moosmann, F., Nowak, E., & Jurie, F. (2008). Randomized clustering forests for image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(9), 1632–46. http://doi.org/10.1109/TPAMI.2007.70822

Mori, Y., Takahashi, H., & Oka, R. (1999). Image-to-word transformation based on dividing and vector quantizing images with words. In *Proc. of First International Workshop on Multimedia Intelligent Storage and Retrieval Management*.

Müller, K. R., Mika, S., Rätsch, G., Tsuda, K., & Schölkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE Transactions on*

Neural Networks, *12*(2), 181–201. http://doi.org/10.1109/72.914517

Naphade, M. R. (2004). On supervision and statistical learning for semantic multimedia analysis. *Journal of Visual Communication and Image Representation*, *15*(3), 348–369. http://doi.org/10.1016/j.jvcir.2004.04.010

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, *9*(2), 249–265. http://doi.org/10.1080/10618600.2000.10474879

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. In *Proc. of 28th International Conference on Machine Learning (ICML-11)* (pp. 689–696). Retrieved from http://machinelearning.wustl.edu/mlpapers/paper_files/ICML2011Ngiam_399.pdf

Ngo, C., Pong, T., & Chin, R. (2001). Exploiting image indexing techniques in DCT domain. *Pattern Recognition*, *34*(March 2000), 1841–1851. Retrieved from http://www.sciencedirect.com/science/article/pii/S0031320300001114

Nguyen, C., Phan, X., & Horiguchi, S. (2009). Web search clustering and labeling with hidden topics. *ACM Transactions on Asian Language Information Processing (TALIP)*, *8*(3), 12. http://doi.org/10.1145/1568292.1568295.http

Nguyen, C.-T., Kaothanthong, N., Tokuyama, T., & Phan, X.-H. (2013). A feature-word-topic model for image annotation and retrieval. *ACM Transactions on the Web (TWEB)*, *7*(3), 1–24. http://doi.org/10.1145/2516633.2516634

Nikolopoulos, S., Zafeiriou, S., Patras, I., & Kompatsiaris, I. (2013). High order pLSA for indexing tagged images. *Signal Processing*, *93*(8), 2212–2228. http://doi.org/10.1016/j.sigpro.2012.08.004

Noble, W. S. (2004). Support vector machine applications in computational biology. In *Kernel methods in computational biology* (pp. 71–92). MIT Press, Cambridge.

Nowak, S., Lukashevich, H., Dunker, P., & Rüger, S. (2010). Performance measures for multilabel evaluation: a case study in the area of image classification categories and subject descriptors. In *Proc. of the International Conference on Multimedia Information Retrieval* (pp. 35–44).

Ntalianis, K., Tsapatsoulis, N., Doulamis, A., & Matsatsinis, N. (2012). Automatic annotation of image databases based on implicit crowdsourcing, visual concept modeling and evolution. *Multimedia Tools and Applications*. http://doi.org/10.1007/s11042-012-0995-2

Odone, F., Barla, A., & Verri, A. (2005). Building kernels from binary strings for image matching. *IEEE Transactions on Image Processing*, *14*(2), 169–

180. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1381485

Ojala, T., Mäenpää, T., Pietikainen, M., Viertola, J., Kyllönen, J., & Huovinen, S. (2002). Outex-new framework for empirical evaluation of texture analysis algorithms. In *Proc. of 16th IEEE International Conference on Pattern Recognition (ICPR)* (pp. 701 – 706). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1044854

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 971–987. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1017623

Oliva, A., & Torralba, A. (2001). Modeling
representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145–175.

Pang, W.-M., Choi, K.-S., & Qin, J. (2013). Fast Gabor texture feature extraction with separable filters using GPU. *Journal of Real-Time Image Processing*. http://doi.org/10.1007/s11554-013-0373-y

Pass, G., & Zabih, R. (1996). Histogram refinement for content-based imagwe retrieval. In *Proc. of 1996 IEEE Workshop on Applications of Computer Vision (WACV'96)* (pp. 96–102).

Pentland, a., Picard, R. W., & Sclaroff, S. (1996). Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, *18*(3), 233–254. http://doi.org/10.1007/BF00123143

Perronnin, F., & Dance, C. (2007). Fisher kernels on visual vocabularies for image categorization. In *Proc. of 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*.

Phan, X., Nguyen, C., Le, D., Nguyen, L., Horiguchi, S., & Ha, Q. (2011). A hidden topic-based framework towards building applications with short web documents. *IEEE Transactions on Knowledge and Data Engineering*, *23*(7), 961–976.

Philbin, J., Chum, O., Isard, M., Sivic, J., & Zisserman, A. (2008). Lost in quantization: Improving particular object retrieval in large scale image databases. In *Proc. of 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–8). Ieee. http://doi.org/10.1109/CVPR.2008.4587635

Porter, R., & Canagarajah, N. (1997). Robust rotation-invariant texture classification: wavelet, Gabor filter and GMRF based schemes. In *Proc. of Vision, Image and Signal Processing* (Vol. I, pp. 180–188). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=599893

Putthividhya, D., Attias, H. T., & Nagarajan, S. S. (2010). Topic regression multi-modal Latent Dirichlet Allocation for image annotation. In *Proc. of 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3408–3415). http://doi.org/10.1109/CVPR.2010.5540000

Qin, J., & Yung, N. H. C. (2010). Scene categorization via contextual visual words. *Pattern Recognition*, *43*(5), 1874–1888. http://doi.org/10.1016/j.patcog.2009.11.009

Rasiwasia, N., & Vasconcelos, N. (2012). Holistic context models for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(5), 902–917. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5989834

Rasmussen, C. E. (2000). The infinite Gaussian mixture model. In *Proc. of Neural Information Processing Systems (NIPS)* (pp. 554–560).

Ritendra, D., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR)*, *40*(2), 1–60. http://doi.org/10.1145/1348246.1348248

Robertson, S. (2008). On the history of evaluation in IR. *Journal of Information Science*, *34*(4), 439–456. http://doi.org/10.1177/0165551507086989

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, *77*(1-3), 157–173. http://doi.org/10.1007/s11263-007-0090-8

Saevarsson, B., Sveinsson, J., & Benediktsson, J. (2006). Combined curvelet and wavelet denoising. In *Proc. of the 7th Nordic Signal Processing Symposium (NORSIG)* (Vol. 1, pp. 318–321). Ieee. http://doi.org/10.1109/NORSIG.2006.275244

Saipullah, K. M., & Kim, D.-H. (2011). A robust texture feature extraction using the localized angular phase. *Multimedia Tools and Applications*, *59*(3), 717–747. http://doi.org/10.1007/s11042-011-0766-5

Saracevic, T. (1995). Evaluation of evaluation in information retrieval. In *Proc. of the 18th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 138–146). Retrieved from http://dl.acm.org/citation.cfm?id=215351

Schmah, T., Zemel, R. S., Hinton, G. E., Small, S. L., & Strother, S. (2008). Generative versus discriminative training of RBMs for classification of fMRI images. In *Proc. of Neural Information Processing Systems (NIPS)* (pp. 1409–1416).

Selvaretnam, B., & Belkhatir, M. (2011). Natural language technology and query expansion: issues, state-of-the-art and perspectives. *Journal of Intelligent Information Systems*, *38*(3), 709–740.

http://doi.org/10.1007/s10844-011-0174-3

Semler, L., & Dettori, L. (2006). Curvelet-based texture classification of tissues in computed tomography. In *Proc. of 2006 International Conference on Image Processing* (pp. 2165–2168). Ieee. http://doi.org/10.1109/ICIP.2006.312873

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, *4*(639-650). Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA238689

Shang, Y. A. N., Diao, Y., & Li, C. (2008). Rotation invariant texture classification algorithm based on curvelet transform and svm. In *Proc. of Seventh International Conference on Machine Learning and Cybernetics* (pp. 3032–3036).

Shen, Y., Archambeau, C., Cornford, D., Opper, M., Shawe-Taylor, J., & Barillec, R. (2010). A comparison of variational and Markov chain Monte Carlo methods for inference in partially observed stochastic dynamic systems. *Journal of Signal Processing Systems*, *61*(1), 51–59. http://doi.org/10.1007/s11265-008-0299-y

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(8), 888–905. http://doi.org/10.1109/34.868688

Shotton, J., Winn, J., Rother, C., & Criminisi, A. (2006). Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *Proc. of European Conference on Computer Vision (ECCV), Springer Berlin Heidelberg* (pp. 1–15). Retrieved from http://link.springer.com/chapter/10.1007/11744023_1

Siddiquie, B., Vitaladevuni, S. N., & Davis, L. S. (2009). Combining multiple kernels for efficient image classification. In *Proc. of 2009 Workshop on Applications of Computer Vision (WACV)* (pp. 1–8). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5403040

Singh, R., Hsu, Y.-W., & Moon, N. (2011). Multiple perspective interactive search: a paradigm for exploratory search and information retrieval on the web. *Multimedia Tools and Applications*, *62*(2), 507–543. http://doi.org/10.1007/s11042-011-0910-2

Singhal, A. (2001). Modern information retrieval: A brief overview. *IEEE Data Engineering Bulletin*, 1–9. Retrieved from http://act.buaa.edu.cn/hsun/IR2013/ref/mir.pdf

Sivic, J., Russell, B. C., Efros, A. A., Zisserman, A., & Freeman, W. T. (2005). *Discovering object categories in image collections*. *Computer Science and Artificial Intelligence Laboratory Technical Report*. Retrieved from http://18.7.29.232/handle/1721.1/30525

Sivic, J., & Zisserman, A. (2003). Video Google: a text retrieval approach to object matching in videos. In *Proc. of 9th IEEE International Conference on Computer Vision (ICCV)* (pp. 1470–1477 vol.2). Ieee. http://doi.org/10.1109/ICCV.2003.1238663

Smith, J. R., & Chang, S.-F. (1997). VisualSEEk: a fully automated content-based image query system. In *Proc. of the fourth ACM international conference on Multimedia*.

Sonnenburg, S., Rätsch, G., Schäfer, C., & Schölkopf, B. (2006). Large scale multiple kernel learning. *The Journal of Machine Learning Research*, *7*, 1531–1565.

Spyrou, E., Iakovidis, D., & Mylonas, P. (2014). *Semantic multimedia analysis and processing*. CRC Press. Retrieved from http://books.google.com/books?hl=en&lr=&id=fujMAwAAQBAJ&oi=fnd&pg=PP1&dq=SEMANTIC+MULTIMEDIA+ANALYSIS+and+processing&ots=nHMzgq7-YH&sig=H0bcpnxbacMIcNE6gQNvsRkfkRE

Srivastava, N., & Salakhutdinov, R. (2012). Multimodal learning with deep boltzmann machines. In *Proc. of Neural Information Processing Systems (NIPS)* (pp. 1–9). Retrieved from http://papers.nips.cc/paper/4683-multimodal-learning-with-deep-boltzmann-machines

Stamou, G., Van, J., Pan, J. Z., & Smith, J. R. (2006). Multimedia annotations on the semantic web. *IEEE MultiMedia*, *13*(1), 86–90.

Starck, J.-L., Candès, E. J., & Donoho, D. L. (2002). The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, *11*(6), 670–84. http://doi.org/10.1109/TIP.2002.1014998

Stathopoulos, V. (2012). *Generative probabilistic models for image retrieval*. University of Glasgow. Retrieved from http://theses.gla.ac.uk/id/eprint/3360

Sumana, I. J., Islam, M. M., Zhang, D., & Lu, G. (2008). Content based image retrieval using curvelet transform. In *Proc. of 2008 IEEE 10th Workshop on Multimedia Signal Processing* (pp. 11–16). Ieee. http://doi.org/10.1109/MMSP.2008.4665041

Sumana, I. J., Lu, G., & Zhang, D. (2012). Comparison of curvelet and wavelet texture features for content-based image retrieval. In *Proc. of IEEE International Conference on Multimedia and Expo* (pp. 290–295). Ieee. http://doi.org/10.1109/ICME.2012.90

Swami, P., & Jain, A. (2012). Segmentation based combined wavelet-curvelet approach for image denoising. *International Journal of Information Engineering*, *2*(1), 32–37. Retrieved from http://www.academicpub.org/IJIE/Download.aspx?ID=5

Tamura, H., Mori, S., & Yamawaki, T. (1978). Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics*, *75*(6), 460–473. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4309999

Tang, J., Zha, Z., Tao, D., & Chua, T. (2012). Semantic-gap-oriented active learning for multilabel image annotation. *IEEE Transactions on Image Processing*, *21*(4), 2354–2360. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6111295

Teh, Y., Newman, D., & Welling, M. (2007). A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. In *Proc. of Advances in neural information processing systems* (Vol. 19, pp. 1353–1360). Retrieved from http://books.nips.cc/papers/files/nips19/NIPS2006_0511.pdf

Teh, Y. W. (2010). Dirichlet process. In *Encyclopedia of machine learning. Springer US* (pp. 280–287).

Thomee, B., & Popescu, A. (2012). Overview of the ImageCLEF 2012 flickr photo annotation and retrieval task. In *Peoc. of CLEF (Online Working Notes/Labs/Workshop)* (Vol. 53). Retrieved from http://www.liacs.nl/~bthomee/assets/12clef_p.pdf

Tian, X., Lu, Y., & Yang, L. (2012). Query difficulty prediction for web image search. *IEEE Transactions on Multimedia*, *14*(4), 951–962. http://doi.org/10.1109/TMM.2011.2177647

Tieu, K., & Viola, P. (2000). Boosting image retrieval. In *Proc. of 2000 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Vol. 1, pp. 228–235). IEEE Comput. Soc. http://doi.org/10.1109/CVPR.2000.855824

Tou, J., Tay, Y., & Lau, P. (2008). One-dimensional grey-level co-occurrence matrices for texture classification. In *Proc. of 2008 IEEE International Symposium on Information Technology (ITSim)* (pp. 4–9). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4631992

Town, C., & Sinclair, D. (2000). *Content based image retrieval using semantic visual categories*. *TR2000-14, AT&T Labs Cambridge*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Content-based+image+retrieval+using+semantic+visual+categories#0

Tsai, C.-F. (2012). Bag-of-Words Representation in Image Annotation: A Review. *International Scholarly Research Notices (ISRN Artificial Intelligence)*, *2012*, 1–19. http://doi.org/10.5402/2012/376804

Tsai, C.-F., & Lin, W.-C. (2012). Scenery image retrieval by meta-feature representation. *Online Information Review*, *36*(4), 517–533. http://doi.org/10.1108/14684521211254040

Tsai, D., Jing, Y., Liu, Y., A.Rowley, H., Ioffe, S., & M.Rehg, J. (2011). Large-scale image annotation using visual synset. In *Proc. of 2011 IEEE International Conference on Computer Vision (ICCV)*. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6126295

Tuceryan, M., & Jain, A. (1990). Texture segmentation using Voronoi polygons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *I*(2), 211–216. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=44407

Vapnik, V. N. (1999). *Statistical learning theory*. *Technometrics* (Vol. 41). Wiley. http://doi.org/10.2307/1271368

Vasconcelos, N. (2001). Image indexing with mixture hierarchies. In *Proc. of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (Vol. 1, pp. I–3–I–10). IEEE Comput. Soc. http://doi.org/10.1109/CVPR.2001.990449

Villegas, M., Paredes, R., & Thomee, B. (2013). Overview of the imageclef 2013 scalable concept image annotation subtask. In *Proc. of CLEF 2013 Evaluation Labs and Workshop, Online Working Notes* (pp. 1–20). Valencia, Spain. Retrieved from http://mvillegas.info/pub/Villegas13_CLEF_Annotation-Overview.pdf

Vogel, J., & Schiele, B. (2004). Natural scene retrieval based on a semantic modeling step. In *Proc. of Image and video Retrieval, Springer Berlin Heidelberg* (pp. 207–215). Springer Verlag. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-27814-6_27

Wang, G., Zhang, Y., & Fei-Fei, L. (2006). Using dependent regions for object categorization in a generative framework. In *Proc. of 2006 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Vol. 2, pp. 1597–1604). Ieee. http://doi.org/10.1109/CVPR.2006.324

Wang, J. Z., & Wiederhold, G. (2001). SIMPLIcity: semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(9), 947–963. http://doi.org/10.1109/34.955109

Wang, J., Zhou, J., Xu, H., Mei, T., Hua, X., & Li, S. (2014). Image tag refinement by regularized latent Dirichlet allocation. *Computer Vision and Image Understanding*, (Shipeng Li). Retrieved from http://www.sciencedirect.com/science/article/pii/S107731421400037X

Wang, M., Yang, K., Hua, X.-S., & Zhang, H.-J. (2010). Towards a relevant and diverse search of social images. *IEEE Transactions on Multimedia*, *12*(8), 829–842. http://doi.org/10.1109/TMM.2010.2055045

Wang, Y., Mei, T., Gong, S., & Hua, X.-S. (2009). Combining global, regional and contextual features for automatic image annotation. *Pattern Recognition*, *42*(2), 259–266. http://doi.org/10.1016/j.patcog.2008.05.010

Weber, M., Welling, M., & Perona, P. (2000). Unsupervised learning of models for recognition. In *Proc. of European conference on computer vision* (pp. 18–32).

Wouwer, G. Van de, Scheunders, P., & Dyck, D. Van. (1999). Statistical texture characterization from discrete wavelet representations. *IEEE Transactions on Image Processing*, *8*(4), 1077–1081. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=753747

Wu, L., Hoi, S. C. H., & Yu, N. (2010). Semantics-preserving bag-of-words models and applications. *IEEE Transactions on Image Processing*, *19*(7), 1908–1920. http://doi.org/10.1109/TIP.2010.2045169

Wu, L., Hoi, S., & Yu, N. (2009). Semantics-preserving bag-of-words models for efficient image annotation. In *Proc. of the First ACM workshop on Large-scale multimedia retrieval and mining (LS-MMRM '09)* (pp. 19–26). Retrieved from http://dl.acm.org/citation.cfm?id=1631064

Wu, Y., & Wu, Y. (2009). Shape-based image retrieval using combining global and local shape features. In *Proc. of 2009 2nd International Congress on Image and Signal Processing* (pp. 1–5). Ieee. http://doi.org/10.1109/CISP.2009.5304693

Xu, Y., Gong, X., & Gong, S. (2011). An algorithm for texture-based Image Retrieval. In *Proc. of 2011 International Conference on Multimedia Technology* (pp. 145–147). Ieee. http://doi.org/10.1109/ICMT.2011.6002019

Yang, B. (2010). Dsi: A model for distributed multimedia semantic indexing and content integration. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *6*(1), 1–21. http://doi.org/10.1145/1671954.1671957

Yang, H.-C., & Lee, C.-H. (2008). Image semantics discovery from web pages for semantic-based image retrieval using self-organizing maps. *Expert Systems with Applications*, *34*(1), 266–279. http://doi.org/10.1016/j.eswa.2006.09.016

Yang, N.-C., Chang, W.-H., Kuo, C.-M., & Li, T.-H. (2008). A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval. *Journal of Visual Communication and Image Representation*, *19*(2), 92–105. http://doi.org/10.1016/j.jvcir.2007.05.003

Yang, Y., Wu, F., Nie, F., Shen, H. T., Zhuang, Y., & Hauptmann, A. G. (2012). Web and personal image annotation by mining label correlation with relaxed visual graph embedding. *IEEE Transactions on Image Processing*, *21*(3), 1339–1351. http://doi.org/10.1109/TIP.2011.2169269

Yavlinsky, A., Schofield, E., & Ruger, S. (2005). Automated image annotation using global features and robust nonparametric density estimation. In *Proc. of Image and video retrieval, Springer Berlin Heidelberg* (pp. 507–

517). Retrieved from http://link.springer.com/chapter/10.1007/11526346_54

Yin, Q., Kim, J., & Moon, K. (2007). Rotation invariant texture classification using Gabor wavelets. In *Proc. of Computational Science–ICCS* (pp. 10–17). Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-72588-6_2

Ying Liu, Dengsheng Zhang, G. L. (2008). Region-based image retrieval with high-level semantics using decision tree learning. *Pattern Recognition*, *41*(8), 2554–2570. http://doi.org/10.1016/j.patcog.2007.12.003

Yu, Y., Pedrycz, W., & Miao, D. (2013). Neighborhood rough sets based multi-label classification for automatic image annotation. *International Journal of Approximate Reasoning*, *54*(9), 1373–1387. http://doi.org/10.1016/j.ijar.2013.06.003

Yuan, J., Wu, Y., & Yang, M. (2007). Discovery of collocation patterns: from visual words to visual phrases. In *Proc. of 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)* (pp. 1–8). Ieee. http://doi.org/10.1109/CVPR.2007.383222

Yuan, Y., Wu, F., Shao, J., & Zhuang, Y. (2013). Image annotation by semi-supervised cross-domain learning with group sparsity. *Journal of Visual Communication and Image Representation*, *24*(2), 95–102. http://doi.org/10.1016/j.jvcir.2012.02.007

Zagoris, K., Chatzichristofis, S. a., Papamarkos, N., & Boutalis, Y. S. (2010). Automatic image annotation and retrieval using the joint composite descriptor. In *Proc. of 14th Panhellenic Conference on Informatics*. Ieee. http://doi.org/10.1109/PCI.2010.38

Zha, Z.-J., Hua, X.-S., Mei, T., Wang, J., Qi, G.-J., & Wang, Z. (2008). Joint multi-label multi-instance learning for image classification. In *Proc. of 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–8). Ieee. http://doi.org/10.1109/CVPR.2008.4587384

Zhai, Y., Neuhoff, D. L., & Pappas, T. N. (2013). Local radius index - a new texture similarity feature. In *Proc. of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1434–1438). Ieee. http://doi.org/10.1109/ICASSP.2013.6637888

Zhang, C., Liu, J., Tian, Q., Liang, C., & Huang, Q. (2013). Beyond visual features: A weak semantic image representation using exemplar classifiers for classification. *Neurocomputing*, *120*, 318–324. http://doi.org/10.1016/j.neucom.2012.07.056

Zhang, D., Islam, M. M., & Lu, G. (2012). A review on automatic image annotation techniques. *Pattern Recognition*, *45*(1), 346–362. http://doi.org/10.1016/j.patcog.2011.05.013

Zhang, D., Islam, M. M., Lu, G., & Sumana, I. J. (2012). Rotation invariant curvelet features for region based image retrieval. *International Journal of Computer Vision*, *98*(2), 187–201. http://doi.org/10.1007/s11263-011-0503-6

Zhang, D., & Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, *37*(1), 1–19. http://doi.org/10.1016/j.patcog.2003.07.008

Zhang, D., Monirul Islam, M., & Lu, G. (2013). Structural image retrieval using automatic image annotation and region based inverted file. *Journal of Visual Communication and Image Representation*, *24*(7), 1087–1098. http://doi.org/10.1016/j.jvcir.2013.07.004

Zhang, S., Huang, Q., Lu, Y., Wen, G., & Tian, Q. (2010). Building pair-wise visual word tree for efficent image re-ranking. In *Proc. of 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)* (pp. 794–797). Ieee. http://doi.org/10.1109/ICASSP.2010.5494964

Zhang, S., Tian, Q., Hua, G., Huang, Q., & Gao, W. (2011). Generating descriptive visual words and visual phrases for large-scale image applications. *IEEE Transactions on Image Processing*, *20*(9), 2664–77. http://doi.org/10.1109/TIP.2011.2128333

Zhang, S., Tian, Q., Hua, G., Huang, Q., & Gao, W. (2014). ObjectPatchNet: Towards scalable and semantic image annotation and retrieval. *Computer Vision and Image Understanding*, *118*, 16–29. http://doi.org/10.1016/j.cviu.2013.03.008

Zhang, X., Li, Z., & Chao, W. (2011). Improving image tags by exploiting web search results. *Multimedia Tools and Applications*, *62*(3), 601–631. http://doi.org/10.1007/s11042-011-0863-5

Zheng, Y., Zhao, M., Neo, S.-Y., Chua, T.-S., & Tian, Q. (2008). Visual synset: towards a higher-level visual representation. In *Proc. of 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–8). Ieee. http://doi.org/10.1109/CVPR.2008.4587611

Zheng, Y.-T., Neo, S.-Y., Chua, T.-S., & Tian, Q. (2009). Toward a higher-level visual representation for object-based image retrieval. *The Visual Computer*, *25*(1), 13–23. http://doi.org/10.1007/s00371-008-0294-0

Zhou, L., Zhang, C., Wan, W., Birch, J., & Chen, W.-B. (2009). An image clustering and retrieval framework using feedback-based integrated region matching. In *Proc. of 2009 International Conference on Machine Learning and Applications* (pp. 596–601). Ieee. http://doi.org/10.1109/ICMLA.2009.131

Zhou, X., Yu, K., Zhang, T., & Huang, T. S. (2010). Image classification using super-vector coding of local image descriptors. In *Proc. of Computer*

Vision–ECCV 2010. Springer Berlin Heidelberg (pp. 141–154).

Zhou, Z., & Zhang, M. (2006). Multi-instance multi-label learning with application to scene classification. In *Proc. of Advances in neural information processing systems* (pp. 1609–1616). Retrieved from http://machinelearning.wustl.edu/mlpapers/paper_files/NIPS2006_348.pdf

Zhu, J., Ahmed, A., & Xing, E. (2009). MedLDA: maximum margin supervised topic models for regression and classification. In *Proc. of 26th Annual International Conference on Machine Learning* (pp. 1257–1264). Retrieved from http://papers.nips.cc/paper/3328-supervised-topic-models

Zujovic, J., Pappas, T. N., & Neuhoff, D. L. (2013). Structural texture similarity metrics for image analysis and retrieval. *IEEE Transactions on Image Processing*, *22*(7), 2545–58. http://doi.org/10.1109/TIP.2013.2251645