UNIVERSITI PUTRA MALAYSIA

EFFECTIVE SALIENCE-BASED FUSION MODEL FOR IMAGE RETRIEVAL

LEILA MANSOURIAN

FSKTM 2016 14

EFFECTIVE SALIENCE-BASED FUSION MODEL FOR IMAGE RETRIEVAL

By

LEILA MANSOURIAN

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

August 2016

DEDICATIONS

I dedicate my dissertation work to my family and many friends. A special feeling of gratitude to my loving parents and my in low parents, Maryam Salehi, Jamshid Mansourian, Mahin Esmailpoor and Ramezan Ali Esmailpoor whose words of encouragement and push for tenacity ring in my ears. I also dedicate this dissertation to my great husband, Babak Esmaielpour, and my kindhearted son, Parsa, because without their love, encouragement and smiles, I would not have finished this thesis. Finally, I am indebted to my lovely sister in law Meymanat Daneshvari, who helped me in English writing and gave me motivations to continue my Ph.D.

Abstract of a thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirements for the degree of Doctor of Philosophy

EFFECTIVE SALIENCE-BASED FUSION MODEL FOR IMAGE RETRIEVAL

By

LEILA MANSOURIAN

August 2016

Chairman: Muhamad Taufik Abdullah, PhD
Faculty : Computer Science and Information Technology

Recently Bag of Visual Words (BoVW) has shown promising results for image annotation and retrieval tasks. In the traditional BoVW model, all visual words are collected and treated the same, regardless of whether or not they are from an important part or the background of a picture. Traditional Scale Invariant Feature Transform (SIFT) features have no spatial information; therefore, the recognition of difficult objects requires more attention. The first objective of this thesis was to develop a new BoVW model, the Salient Based Bag of Visual Word (SBBoVW) model, to recognize difficult objects that previous methods were unable to accurately identify. This new model collects visual words based on their importance and combines several Pyramidal Histogram of visual Words (PHOW) feature vectors from the salient, rectangular part of a picture, as well as from the whole picture, to overcome the above-mentioned problem. After implementation, it was found that this method of feature extraction affects the accuracy of the results, which were more accurate than results obtained using seven other state-of-the-art models. However, the SBBoVW model focused only on gray-scale pictures.

Previous research found that integrating color, significantly improved the overall performance of both feature detection and extraction because color is an important characteristic of human vision. Based on the literature, most of the image classification strategies have been developed for gray-based SIFT descriptors. Since color content is ignored, misclassification may occur. The Dominant Color Descriptor (DCD) is the best color descriptor for region color and the focus of improvements because it is a low-dimensional or less expensive descriptor representing colors in images. The DCD uses one to eight colors for each picture, and one to four colors for each region. However, some background colors are not used in the object of an image. Therefore, the second objective of this research was to establish a new Salient Dominant Color Descriptor (SDCD) to estimate the number of colors in a salient region using an easily implemented algorithm. Based on the results, it was found that if the maximum Euclidean color distance (d_{max}) was set to 20, as suggested by other researchers, more accurate results were obtained.

The DCD is both low-dimensional and less expensive for representing image colors compared to the previous BoVW model that concentrated on the Color Scale Invariant Feature Transform (CSIFT), combinations of color SIFTs extracted from different color spaces, and opponent-color SIFTs extracted from opponent color spaces to add color information to a SIFT. Therefore, the final objective of this research was to develop a late fusion model, the SDCD BoVW and SBBoVW model. This model fuses the SDCD BoVW, and SB-BoVW models using late fusion from histograms and is a comprehensive model for color object recognition. After implementation, the final proposed model provided more accurate results than the other three state-of-the-art models mentioned here and 19 additional color feature extraction methods.

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

MODEL GABUNGAN BERASASKAN KEPENTINGAN YANG EFEKTIF UNTUK DAPATAN SEMULA IMEJ

Oleh

LEILA MANSOURIAN

Ogos 2016

Pengerusi: Muhamad Taufik Abdullah, PhD

Fakulti : Sains Komputer dan Teknologi Maklumat

Kebelakangan ini, model *Bag of Visual Word* (BoVW) telah mempamerkan keputusan yang memberangsangkan dalam tugas anotasi dan dapatan imej. Menerusi model BoVW yang tradisional, semua perkataan visual dikumpul dan dianggap sama rata antara satu sama lain, tidak kira ianya dari bahagian penting gambar ataupun dari bahagian latar belakang gambar. Ciri-ciri *Scale Invariant Feature Transform* (SIFT) yang tradisional tidak mempunyai maklumat berkaitan ruang, oleh itu, lebih perhatian diperlukan untuk mengenalpasti objek yang sukar. Objektif pertama tesis ini adalah untuk membangunkan model *Salient Based Bag Visual Word* (SBBoVW) yang baharu untuk mengenalpasti objek sukar yang tidak dapat dikenalpasti dengan tepat oleh kaedahkaedah yang lepas. Model baharu ini mengumpul perkataan visual berdasarkan tahap kepentingan dan menggabungkan beberapa ciri vektor *Pyramidal Histogram of visual Words* (PHOW) daripada bahagian penting, bahagian segi empat tepat gambar serta bahagian keseluruhan gambar bagi mengatasi masalah yang dibincangkan di atas. Selepas perlaksanaan, didapati bahawa kaedah pengekstrakan ciri ini mempengaruhi ketepatan keputusan, yang mana keputusan ini adalah lebih tepat berbanding keputusan yang diperoleh daripada tujuh model *state-of-art* yang lain. Walau bagaimanapun, model SBBoVW ini hanya memberi tumpuan kep-ada gambar berskala kelabu sahaja. Kajian lepas mendapati penggabungan warna dapat memperbaiki prestasi keseluruhan keduadua ciri pengesanan dan pengekstrakan dengan ketara kerana warna merupakan sifat penting dalam penglihatan manusia.

Berdasarkan kajian yang lepas, kebanyakan strategi pengelasan imej telah dibangunkan untuk penghurai SIFT berasaskan kelabu. Oleh sebab kandungan warna diabaikan, kesilapan dalam pengelasan mungkin terjadi. *Dominant Color Descriptor* (DCD) merupakan penghurai warna yang terbaik untuk warna serantau, dan tumpuan diberikan kepada penambahbaikannya kerana ia merupakan penghurai berdimensi rendah atau berkos rendah yang mew-akili warna-warna di dalam imej. DCD menggunakan satu hingga lapan warna untuk setiap gambar, dan satu hingga empat warna untuk setiap rantau. Walau bagaimanapun, sesetengah warna latar belakang tidak digunakan dalam objek dalam sesuatu imej. Oleh itu, objektif kedua kajian ini adalah untuk mewujudkan *Salient Dominant Color Descriptor*(SDCD) yang

baharu bagi menganggarkan bilangan warna di dalam satu rantau yang penting dengan menggunakan satu algoritma yang mudah dilaksanakan. Berdasarkan keputusan, didapati bahawa sekiranya jarak maksimum warna Euclidean ditetapkan pada 20 seperti yang dicadangkan oleh penyelidik-penyelidik yang lain, maka keputusan yang lebih tepat akan diperolehi.

DCD mempunyai ciri yang berdimensi rendah serta berkos rendah dalam mewakili warna imej berbanding dengan model BoVW yang lepas yang memberi tumpu-an kepada *Color Scale Invariant Feature Transform* (CSIFT), gabungan warna SIFTs yang diekstrak daripada ruang warna yang berbeza, dan warna berlawanan SIFTs yang diekstrak daripada ruang warna yang berlawanan untuk menambah maklumat warna kepada SIFT. Maka, objektif terakhir kajian ini adalah untuk membangunkan satu model gabungan akhir di antara model SDCD, BoVW dan SBBoVW. Model ini menggabungkan model-model SDCD, BoVW dan SBBoVW dengan menggunakan gabungan akhir daripada histogram dan ia merupakan model yang komprehensif dalam pengenalan warna objek. Selepas perlaksanaan, model terakhir yang dicadangkan memberikan keputusan yang lebih tepat berbanding dengan tiga model *state-of-art* yang disebutkan dan sembilan belas kaedah pengekstrakan ciri warna tambahan.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Associate Prof. Dr. Muhamad Taufik Abdullah, whose expertise, understanding, and patience added considerably to my graduate experience. I appreciate his vast knowledge. I would like to thank the other members of my committee, Associate Prof. Dr. Lili Nurliyana Abdullah, Dr. Azreen Azman, and Dr. Mas Rina Mustaffa for the assistance they provided at all levels of the research project.

Hopefully, the completion of this work will mark the starting point of many quality research works in the future, Insya-Allah.

I certify that a Thesis Examination Committee has met on 11 August 2016 to conduct the final examination of Leila Mansourian on her thesis entitled "Effective Salience-Based Fusion Model for Image Retrieval" 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:

Hamidah binti Ibrahim, PhD
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Shyamala a/p C. Doraisamy, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

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

Alan F. Smeaton, PhD
Professor
Dublin City University
Ireland
(External Examiner)

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

Date: 28 September 2016

This thesis was submitted to the Senate of the Universiti Putra Malaysia and has been accepted as fulfillment of the requirements for the degree of Doctor of Philosophy.

The members of the supervisory committee were as follows:

Muhamad Taufik Abdullah, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairperson)

Lili Nurliyana Abdullah, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Azreen Azman, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Mas Rina 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.: Leila Mansourian, GS 33658

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:	________________	Signature:	________________
Name of Chairman of Supervisory Committee:	________________-	Name of Member of Supervisory Committee:	________________
Signature:	________________	Signature:	________________
Name of Member of Supervisory Committee:	________________	Name of Member of Supervisory Committee:	________________

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIA	Automatic Image Annotation
ANN	Artificial Neural Network
ASIFT	Affine Scale Invariant Feature Transform
BoW	Bag of Words
BoVW	Bag of Visual Words
BPLRs	Boundary-Preserving Local Regions
CBIR	Content Based Image Retrieval
CCV	Color Coherence Vector
CIE	the International Commission on Illumination
CLD	Color Layout Descriptor
CSD	Color Structure Descriptor
CMYK	Cyan, Magenta, Yellow, and Black
DBSCAN	Density Based Spatial Clustering of Applications with Noise
DCD	Dominant Color Descriptor
DCT	Discrete Cosine Transform
DoG	Difference-of-Gaussian
DRFI	Discriminative Regional Feature Integration
DSIFT	Dense Scale Invariant Feature Transform
DT	Decision Tree
EM	Expectation-Maximization
FAIR-SURF	Fully Affine Invariant version of Speeded Up Robust Features
FA-SIFT	Fast SIFT
FD	Fractal Dimension
FFT	Fast Fourier Transform
FT	Fourier Transform
GIST	Global feature descriptor
GLCM	Gray Level Co-occurrence Matrix
GLOH	Gradient Location and Orientation Histogram
GoF	Group of Frames
GoP	Group of Pictures
GSIFT	Global context SIFT
HMMD	Hue-Min-Max-Differences
HOG	Histogram of Oriented Gradients
HSV	Hue, Saturation, Value
IL-NBNN	Inclined Local Naive Bayes Nearest Neighbor
LAB	Lightness and A and B for the color-opponent dimensions
LBP	Local Binary Pattern
LCC	Local Coordinate Coding
LLC	Locality-constrained Linear Coding
LoG	Laplacian-of-Gaussian
LPC	Locality Preserving Clustering
MIML	Multi-Instance Multi-Label

MPEG-7	Multimedia Content Description Interface
MRF	Markov Random Field
MSDSIFT	Multi-scale Dense Scale Invariant Feature Transform
MSER	Maximally Stable External Region
NBNN	Naive Bayes Nearest Neighbor
NCut	Normalized Cut
PCA	Principal Component Analysis
PCA-SIFT	Principal Component Analysis Scale Invariant Feature Transform
PGF	Peer-Group Filtering
PHOG	Pyramid Histogram of Oriented Gradients
PHOW	Pyramidal Histogram of visual Words
RBIR	Region-Based Image Retrieval
RGB	Red, Green, Blue
ROI	Region of Interest
SAR	Simultaneous Autoregressive
SBBoVW	Salient Based Bag of Visual Words
SC	Sparse coding
ScSPM	Sparse coding Spatial Pyramid Matching
SCD	Scalable Color Descriptor
SDCD & SBBoVW	Salient Dominant Color Description and Salient Based Bag of Visual Words
SIFT	Scale Invariant Feature Transform
SPM	Spatial Pyramid Matching
SURF	Speeded Up Robust Features
SVM	Support Vector Machines
SVM Chi2	Support Vector Machines Chi-square
VQ	Vector Quantization
WDCD	Weighted Dominant Color Descriptor
WT	Wavelet Transform

CHAPTER 1

INTRODUCTION

This chapter serves as the starting point for the entire thesis. It provides the background of Content-Based Image Retrieval (CBIR), the Scale-Invariant Feature Transform (SIFT), and the Bag of Visual Words (BoVW). Then, the motivation of this research in image retrieval is discussed. The details of the research's problem statement, research objectives, research scope, and research contributions are given in the following sections. The final section presents an outline of the thesis structure.

1.1 Background

Content-Based Image Retrieval (CBIR) was proposed by Qi and Snyder in the late 1990s. This method extracts low-level features (color, texture, and shape) from images and categorizes images based on the differences in these features. Although low-level features, such as texture, color, spatial relationship, and shape, are extracted automatically when using computer vision techniques, CBIR often fails to describe high-level semantic concepts (Zhou and Huang, 2000).

CBIR systems are limited when dealing with large image databases (Liu et al., 2007); however, low-level features require some preprocessing. In 1999, Lowe presented a robust feature, Scale-Invariant Feature Transform (SIFT) (Lowe, 1999), which accurately scales, rotates, translates, illuminates, and partially invariant to affine distortion. SIFT features must be quantized using the well-known Bag of Visual Words (BoVW) technique, originally presented by Csurka et al. (2004), to generate a visual word vocabulary (or codebook). The BoVW method was first proposed for document classification and originally named as the Bag of Words (BoW); it represents a document as a bag of words and features extracted based on the frequency of occurrence of each word. Recently, the BoW model has been applied in computer vision (Fei-Fei and Perona, 2005), where it has been renamed to BoVW and used with the visual word vectors of images to extract information from each visual word in images. For BoVW extraction, blobs and features (e.g., SIFT) are first extracted, a visual vocabulary is built using the clustering method (e.g., K-means), and representations of images are compiled from BoVW histograms. In the final stage, images are classified using methods such as the Support Vector Machine (SVM). Mikolajczyk and Schmid (2005) compared several feature descriptors and found that, in many situations SIFT-like descriptors outperform other descriptors. Therefore, this thesis focused on using SIFT features. Later, Bay et al. (2006) proposed Speeded Up Robust Features (SURF), which is quicker than the SIFT, Liu et al. (2008) suggested a faster algorithm for computation of dense sets of SIFT descriptors, and Dalal and

Triggs (2005) used the histogram of oriented gradient (HOG) descriptor for pedestrian detection.

1.2 Motivation and Importance of the Study

The problem with the traditional SIFT feature extraction strategy is that it disregards all information about the spatial layout of features. To overcome this limitation, Pyramidal Histogram of Visual Words (PHOW) is currently used for SIFT features and was proposed by Lazebnik et al. (2006). It uses a dense SIFT at different scales to build a pyramid of descriptors. Pyramidal Histogram of Oriented Gradient (PHOG) is the edge version of PHOW, which means it gathers features of edge detected images. In early experiments, the author used BoVW for SIFT features with different SVM methods such as LIBSVM (Chang and Lin, 2011). However, the traditional BoVW model cannot collect visual words based on their locations in a picture. Therefore, all the visual words are collected and treated the same even if they are from both the important areas or background regions (i.e., the classifier relies on visual words that fall in the background and merely describes the context of the object (Oquab et al., 2015)). Traditional SIFT features contain no spatial information, making it difficult to obtain precise object recognition. To address these problems, a Salient Based Bag of Visual Words (SBBoVW) model for difficult object recognition and object locating was proposed. This model collects visual words from whole and salient parts of an image using spatial PHOW histograms.

Based on the literature, most image classification strategies were developed for gray-based SIFT descriptors (Yang et al., 2009; Wang et al., 2010; Shabou and LeBorgne, 2012; Liu et al., 2011; Yang et al., 2010, 2011), despite the fact that color information is also very important. Misclassification occurs because color content is ignored during image classification. Vigo et al. (2010) found that integrating color significantly improved the overall performance of both feature detection and extraction. Also, by adding color information to illumination changes, the matching rate becomes more accurate (Chen et al., 2015; Krylov and Sorokin, 2011).

To add color information to SIFT features, different types of color SIFT (CSIFT) descriptors were proposed and developed by researchers to utilize the color information inside the SIFT descriptors (Chen et al., 2015). Bosch et al. (2007) added color information to SIFT by extracting features from all of the channels in the Hue, Saturation, Value (HSV) color model, called HSV-SIFT. Chen et al. (2015) investigated CSIFTs with different color spaces, including Red, Green, Blue (RGB), HSV, an M-by-3 matrix that contains the luminance (Y) and chrominance (Cb and Cr) color values as columns (YCbCr), Opponent, rg, and color invariant spaces and found that YCbCr-SIFT descriptors achieved the most stable and accurate image classification performance among the CSIFT descriptors.

The MPEG-7 standard proposes different methods to obtain color descriptors. Zhang et al. (2012) found that among the various MPEG-7 color features, the Dominant Color Descriptor (DCD) was a good descriptor for representing colors in regions with low dimensionality or regions that were less expensive to compute. Additionally, the Color Coherence Vector (CCV), color correlogram, and Scalable Color Descriptor (SCD) are useful for whole image representation. Recently, Talib et al. (2013) proposed a new weighted dominant color descriptor (weight for each Dominant Color (DC)), to reduce the bad effect of image background on the accuracy retreival results. They also proposed a new similarity measurement to measure the similarity based on the DCs distances. However, the model cannot focus on the foreground and removes the background colors completely.

Because of the advantages of the DCD for color region extraction of low-dimensio-nal features, the author focused on the using of this color feature to solve current problem related to the DCD. Most of the previous BoVW models concentrated on CSIFTs (Abdel-Hakim et al., 2006), combinations of color SIFTs (Rassem and Khoo, 2011), opponent colors, and color histograms to add color information to SIFT and PHOW SIFT features. Therefore, a new Salient Dominant Color Descriptor (SDCD) BoVW and SBBoVW fusion model was proposed in the current study to improve the final results.

1.3 Problem Statement

This research addressed the problem of difficult object recognition using a fusion model. The author found that traditional BoVW models collect visual words similarly, regardless of where they are located in the image and that traditional SIFT features contained no spatial information. Therefore, the recognition of difficult objects requires more attention. Based on the literature, most image classification strategies were developed for gray-based SIFT descriptors (Yang et al., 2009; Wang et al., 2010; Shabou and LeBorgne, 2012; Liu et al., 2011; Yang et al., 2010, 2011) and because color content is ignored, misclassification occurs. The DCD is a low-dimensional and cost effective descriptor for representing colors from all regions of an image (Zhang et al., 2012); however, previous BoVW models concentrated on CSIFTs (Abdel-Hakim et al., 2006), combinations of color SIFTs (Rassem and Khoo, 2011), and opponent colors to add color information to SIFTs.

1.4 Research Objectives

The main aim of this research is to propose a new late fusion model for image retrieval by fusing dominant colors and invariant features. The objectives of this research are as follows:

1. To propose a new approach, which is a Salient Based Bag of Visual Word model (SBBoVW) in order to recognize difficult objects which have had low accuracy in previous methods, and to introduce a new algorithm for finding object place based on the salient map automatically.

2. To introduce a new algorithm, Salient Dominant Color Descriptor (SDCD), to extract dominant colors of the salient object in the picture and find the suitable maximum Euclidean color distance for the proposed color descriptor.

3. To present a new model for fusing the SDCD and SBBoVW models to create a single comprehensive model.

1.5 Research Scope

The potential uses of salient maps for recognizing difficult objects and estimating object locations in pictures were investigated. The Multi-scale Dense SIFT (MSDSIFT) PHOW features for invariant feature extraction, Spatial Pyramid Matching (SPM) for adding spatial information, Elkan K-means for fast visual word dictionary construction, Chi2 Support Vector Machine (SVM-Chi2) were used as classifiers. A new dominant color descriptor based on the saliency map was proposed to generate the DCs of the salient region. Luminescence satUration hue angle Value (LUV) was the color space used to determine the maximum Euclidean color distance (d_{max}). Color similarity was measured using the formula proposed in Yang et al. (2008).

To assist the evaluation, previous BoVW models were implemented, and the same train and test pictures were selected based on their location in the folder. For the images that could not be run, the same experimental setup was followed. The system divided the dataset to train and test the images. With the help of a confusion matrix, in which each column of the matrix represents the instances in a predicted class and each row represents the instances in an actual class (or vice-versa), visualization of the performance of each classification strategy, including measuring the precision, recall, accuracy, and classification rate, was possible.

1.6 Research Contributions

Three BoVW models for object recognition were developed, including the SBBoVW model to collect visual words based on their locations in a picture and to recognize difficult objects, the SDCD to extract the DCs from salient objects in a picture, and the combined SDCD and SBBoVW model to enhance the final results with salient feature selection and late feature fusion methods. Each of the proposed models was evaluated and compared against some of the benchmark methods using large image dataset queries and various accuracy measurements (including precision, recall, accuracy, and the classification rate). This research contributes to establish the BoVW models for object recognition and Automatic Image Annotation (AIA) methods. There are many applications in both commercial and scientific fields that routinely use image data. These proposed methods allow users to find objects and automatically annotate pictures based on the DCD color and SIFT features. The contributions of the research are as follows:

1. A novel BoVW model, the SBBoVW model, was introduced to collect visual words based on their importance and whether they were from the salient object or background region of an image. Additionally, a new algorithm for automatically finding object locations based on salient maps was proposed.
2. A novel SDCD was introduced to extract colors from salient objects in an image to estimate a suitable color number based on the Euclidean distance.
3. A novel BoVW model, the combined SDCD and SBBoVW model for superior color object recognition, was introduced to create a comprehensive model for color object recognition.

1.7 Outline of the Thesis Structure

There are three research structure styles used at the University Putra Malaysia (UPM) based on the Graduate School of Studies (GSO) guidelines for thesis preparation (2009). The second style was chosen for this thesis, which was divided into four parts: an introduction, a literature review, the research methodology, and a conclusion. Each research chapter represents a separate study and includes introduction, methodology, results and discussion sections. The sections of this research complement the technical elements that form the project under discussion. The overall organization of the thesis is as follows.

Chapter one provides an introduction to the research work and discusses the study research background, problems that arise from each of the respective object recognition and AIA methods, this research was derived the benefits of combining these features to create an integrated SDCD and SBBoVW model, the research objectives, scope of the research, and the contributions of this research.

Chapter two provides a comprehensive literature review of AIA components and stages. The components that contributed to the current research are discussed and compared to determine the advantages and disadvantages of each.

Chapter three explains the methodology used for this research. There are five phases involved in the methodology: the research problem identification phase, experimental research planning phase, conduction of the experiments phase, data analysis and discussion phase, and report writing phase. The experimental framework design is also explained in this chapter. Each of the experiments conducted for this research are explained thoroughly.

Chapter four explains the design, experimental results, and a discussion for the first experiment: evaluating six kinds of classification strategies in the bag of SIFT feature method for animal recognition, and the second experiment evaluated SIFT feature strategies (including SIFT, Dense SIFT, and Multi-Scale Dense SIFT) in the special animal domain. These experiments used in the proposed SBBoVW model, which is also described in detail. The implementation results and a comparison of all three experiments are discussed in this chapter. Additionally, the model design, final results, and discussion, including a detailed break down of the proposed algorithm for locating objects, are explained thoroughly.

Chapter five provides a comprehensive explanation of the design, experimental results, and discussion related to the proposed SDCD algorithm for extracting colors from salient objects in pictures. In this experiment, different maximum Euclidean color distances were tested to determine which distance works the best.

Chapter six provides a comprehensive explanation of the design, experimental results, and discussion related to the novel, combined SDCD and SBBoVW model for recognising color objects. In this chapter, the model for SDCD and MSDSIFT PHOW feature fusion is explained thoroughly. Then, the final results and a comparison of three state-of-the-art models and 19 different color feature extraction methods are provided.

Chapter seven summarizes the strengths and limitations of each of the proposed methods, namely the *SBBoVW* model, *SDCD* algorithm, and *SDCD and SBBoVW* model. Suggestions for future research of the mentioned methods are also provided in this chapter.

BIBLIOGRAPHY

Abdel-Hakim, A. E., Farag, A. et al. 2006. CSIFT: A SIFT descriptor with color invariant characteristics. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference*, 1978–1983. IEEE.

Alham, N. K., Li, M. and Liu, Y. 2012. Parallelizing multiclass support vector machines for scalable image annotation. *Neural Computing and Applications* 24 (2): 367–381.

Alham, N. K., Li, M., Liu, Y., Hammoud, S. and Ponraj, M. 2011. A distributed SVM for scalable image annotation. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference*, 2655–2658. IEEE.

Alqasrawi, Y., Neagu, D. and Cowling, P. I. 2011. Fusing integrated visual vocabularies-based bag of visual words and weighted colour moments on spatial pyramid layout for natural scene image classification. *Signal, Image and Video Processing* 7 (4): 759–775.

Bahri, A. and Zouaki, H. 2015. Content-Based Bird Retrieval using Shape Context , Color Moments and Bag of Features. *International Journal of Computer Science Issues IJCSI* 12 (1): 101–105.

Bannour, H. and Hudelot, C. 2014. Building and using fuzzy multimedia ontologies for semantic image annotation. *Multimedia Tools and Applications* 72 (3): 2107–2141.

Barata, C., Marques, J. S. and Rozeira, J. 2013. Evaluation of color based keypoints and features for the classification of melanomas using the bag-of-features model. In *International Symposium on Visual Computing*, 40–49. Springer.

Bay, H., Tuytelaars, T. and Van Gool, L. 2006. Surf: Speeded up robust features. In *European Conference on Computer Vision*, 404–417. Springer.

Bilen, H., Namboodiri, V. P. and Van Gool, L. J. 2013. Object and Action Classification with Latent Window Parameters. *International Journal of Computer Vision* 106 (3): 237–251.

Binder, A., Samek, W., Müller, K.-R. and Kawanabe, M. 2013. Enhanced representation and multi-task learning for image annotation. *Computer Vision and Image Understanding* 117 (5): 466–478.

Borji, A. 2015. What is a salient object? a dataset and a baseline model for salient object detection. *IEEE Transactions on Image Processing* 24 (2): 742–756.

Borji, A., Cheng, M.-M., Jiang, H. and Li, J. 2014. Salient object detection: A survey. *arXiv preprint arXiv:1411.5878* .

Borji, A., Cheng, M.-M., Jiang, H. and Li, J. 2015. Salient object detection: A benchmark. *IEEE Transactions on Image Processing* 24 (12): 5706–5722.

Bosch, A., Zisserman, A. and Munoz, X. 2007. Image classification using random forests and ferns. In *2007 IEEE 11th International Conference on Computer Vision*, 1–8. IEEE.

Bosch, A., Zisserman, A. and Muoz, X. 2008. Scene classification using a hybrid generative/discriminative approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions* 30 (4): 712–727.

Budikova, P., Botorek, J., Batko, M., Zezula, P. et al. 2014. DISA at ImageCLEF 2014: The Search-based Solution for Scalable Image Annotation. In *CLEF (Working Notes)*, 360–371. Citeseer.

Carson, C., Belongie, S., Greenspan, H. and Malik, J. 2002. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (8): 1026–1038.

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

Chen, J., Li, Q., Peng, Q. and Wong, K. H. 2015. CSIFT based locality-constrained linear coding for image classification. *Pattern Analysis & Applications* 18 (2): 441–450.

Chiang, C.-C. 2013. Interactive tool for image annotation using a semi-supervised and hierarchical approach. *Computer Standards & Interfaces* 35 (1): 50–58.

Corinna Cortes, V. V. 1995. Support-Vector Networks. *Machine Learning* 20 (3): 273 – 297.

Csurka, G., Dance, C. R., Fan, L., Willamowski, J., Bray, C. and Maupertuis, D. 2004. Visual Categorization with Bags of Keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, 1–2. Prague.

Dalal, N. and Triggs, B. 2005. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 886–893. IEEE.

Deng, Y., Kenney, C., Moore, M. S. and Manjunath, B. 1999. Peer group filtering and perceptual color image quantization. In *Circuits and Systems, 1999. ISCAS'99. Proceedings of the 1999 IEEE International Symposium*, 21–24. IEEE.

Dey, V., Zhang, Y. and Zhong, M. 2010. *A Review on Image Segmentation Techniques with Remote Sensing Perspective*. , vol. XXXVIII.

Deza, M. M. and Deza, E. 2009, In Encyclopedia of Distances, In *Encyclopedia of Distances*, 1–583, Springer, 1–583.

El Rube, I., Sharks, M., Salem, A. R. et al. 2009. Image registration based on multi-scale SIFT for remote sensing images. In *Signal Processing and Communication Systems, 2009. ICSPCS 2009. 3rd International Conference*, 1–5. IEEE.

Fakhari, A. and Moghadam, A. M. E. 2013. Combination of classification and regression in decision tree for multi-labeling image annotation and retrieval. *Applied Soft Computing* 13 (2): 1292–1302.

Fei-Fei, L., Fergus, R. and Perona, P. 2007. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding* 106 (1): 59–70.

Fei-Fei, L. and Perona, P. 2005. A bayesian hierarchical model for learning natural scene categories. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 524–531. IEEE.

Geusebroek, J.-M., Van den Boomgaard, R., Smeulders, A. and Geerts, H. 2001. Color invariance. *Pattern Analysis and Machine Intelligence, IEEE Transactions* 23 (12): 1338–1350.

Grabner, M., Grabner, H. and Bischof, H. 2006. Fast approximated SIFT. In *Asian conference on computer vision*, 918–927. Springer.

Griffin, G., Holub, A. and Perona, P. 2007. Caltech-256 object category dataset .

Hong, R., Wang, M., Gao, Y., Tao, D. and Member, S. 2014. Image Annotation By Multiple-Instance Learning With Discriminative Feature Mapping and Selection. *IEEE transactions on cybernetics* 44 (5): 669–680.

Hu, J. and Lam, K.-M. 2013. An efficient two-stage framework for image annotation. *Pattern Recognition* 46 (3): 936–947.

Hung, C. and Tsai, C.-F. 2008. Automatically Annotating Images with Keywords: A Review of Image Annotation Systems. *Recent Patents on Computer Sciencee* 1: 55–68.

Ionescu, R. T., Popescu, M. and Grozea, C. 2013. Local learning to improve bag of visual words model for facial expression recognition. In *Workshop on Challenges in Representation Learning, ICML*.

Islam, M. M., Zhang, D. and Lu, G. 2008. Automatic Categorization of Image Regions Using Dominant Color Based Vector Quantization. In *Digital Image Computing: Techniques and Applications (DICTA), 2008*, 191–198. IEEE.

Islam, M. M., Zhang, D. and Lu, G. 2009. Region based color image retrieval using curvelet transform. In *Asian Conference on Computer Vision*, 448–457. Springer.

Jalali, S., Tan, C., Ong, S.-h., Seekings, P. J. and Taylor, E. A. 2013. Visual Recognition using a Combination of Shape and Color Features. In *(CogSci), The Annual Meeting of the Cognitive Science Society*, 2638–2643.

Jeannin, S., Cieplinski, L., Ohm, J. R. and Kim, M. 2001. Mpeg-7 visual part of experimentation model version 9.0. *ISO/IEC JTC1/SC29/WG11 N* 3914.

Jégou, H., Douze, M. and Schmid, C. 2010. Improving bag-of-features for large scale image search. *International Journal of Computer Vision* 87 (3): 316–336.

Ji, Z., Theiler, J., Chartrand, R., Kenyon, G. and Brumby, S. P. 2013. Decoupling sparse coding of SIFT descriptors for large-scale visual recognition. In *SPIE Defense, Security, and Sensing*, 87500K–87500K. International Society for Optics and Photonics.

Jiang, H., Wang, J., Yuan, Z., Liu, T., Zheng, N. and Li, S. 2011. Automatic salient object segmentation based on context and shape prior. In *BMVC*.

Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N. and Li, S. 2013. Salient object detection: A discriminative regional feature integration approach. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2083–2090. IEEE.

Juan, L. and Gwun, O. 2009. A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)* 3 (4): 143–152.

Kazmi, W. and Andersen, H. J. 2015. A comparison of interest point and region detectors on structured, range and texture images. *Journal of Visual Communication and Image Representation* 3: 156–169.

Ke, Y. and Sukthankar, R. 2004. PCA-SIFT: a more distinctive representation for local image descriptors. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, II–506. IEEE.

Khan, F. S., vande Weijer, J. and Vanrell, M. 2011. Modulating Shape Features by Color Attention for Object Recognition. *International Journal of Computer Vision* 98 (1): 49–64.

Khan, R., Barat, C., Muselet, D. and Ducottet, C. 2015. Spatial histograms of soft pairwise similar patches to improve the bag-of-visual-words model. *Computer Vision and Image Understanding* 132: 102–112.

Kim, J. and Grauman, K. 2011. Boundary preserving dense local regions. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference*, 1553–1560. IEEE.

Krylov, A. S. and Sorokin, D. V. 2011. Gauss-Laguerre keypoints descriptors for color images. In *Visual Communications and Image Processing (VCIP), 2011 IEEE*, 1–4. IEEE.

Lampert, C. H., Blaschko, M. B. and Hofmann, T. 2008. Beyond sliding windows: Object localization by efficient subwindow search. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference*, 1–8. IEEE.

Lankinen, J., Kangas, V. and Kamarainen, J.-K. 2012. A comparison of local feature detectors and descriptors for visual object categorization by intra-class repeatability and matching. In *21st International Conference on Pattern Recognition (ICPR 2012)*, 780–783. IEEE.

Lazebnik, S., Schmid, C. and Ponce, J. 2006. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*, 2169–2178. IEEE.

Lee, C.-H., Yang, H.-C. and Wang, S.-H. 2011. An image annotation approach using location references to enhance geographic knowledge discovery. *Expert Systems with Applications* 38 (11): 13792–13802.

Lei, Y., Wong, W., Liu, W. and Bennamoun, M. 2011, In Computer Vision–ACCV 2010, In *Computer Vision–ACCV 2010*, 115–126, Springer, 115–126.

Li, Z., Liu, J., Yang, Y., Zhou, X. and Lu, H. 2014. Clustering-guided sparse structural learning for unsupervised feature selection. *IEEE Transactions on Knowledge and Data Engineering* 26 (9): 2138–2150.

Lin, W.-C., Tsai, C.-F., Chen, Z.-Y. and Ke, S.-W. 2016. Keypoint selection for efficient bag-of-words feature generation and effective image classification. *Information Sciences* 329: 33–51.

Liu, C., Yuen, J., Torralba, A., Sivic, J. and Freeman, W. T. 2008. Sift flow: Dense correspondence across different scenes. In *European Conference on Computer Vision*, 28–42. Springer.

Liu, L., Wang, L. and Liu, X. 2011. In defense of soft-assignment coding. In *IEEE International Conference on Computer Vision (ICCV 2011)*, 2486–2493. IEEE.

Liu, Y., Zhang, D., Lu, G. and Ma, W.-Y. 2007. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition* 40 (1): 262–282.

Lowe, D. 1999. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1150–1157. IEEE.

Lowe, D. G. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60 (2): 91–110.

Luo, H.-L., Wei, H. and Lai, L. L. 2011. Creating efficient visual codebook ensembles for object categorization. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions* 41 (2): 238–253.

Maeda, T., Yamasaki, T. and Aizawa, K. 2014. Multi-stage object classification featuring confidence analysis of classifier and inclined local naive bayes nearest neighbor. In *2014 IEEE International Conference on Image Processing (ICIP)*, 5177–5181. IEEE.

McCann, S. and Lowe, D. 2012. Local Naive Bayes Nearest Neighbor for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference*, 3650–3656.

Mikolajczyk, K., Leibe, B. and Schiele, B. 2005. Local features for object class recognition. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference*, 1792–1799. IEEE.

Mikolajczyk, K. and Schmid, C. 2004. Scale & affine invariant interest point detectors. *International Journal of Computer Vision* 60 (1): 63–86.

Mikolajczyk, K. and Schmid, C. 2005. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions* 27 (10): 1615–1630.

Moh'd A Mesleh, A. 2007. Chi square feature extraction based SVMs Arabic language text categorization system. *Journal of Computer Science* 3 (6): 430–435.

Morel, J.-M. and Yu, G. 2009. ASIFT: A New Framework for Fully Affine Invariant Image Comparison. *SIAM Journal on Imaging Sciences* 2 (2): 438–469.

Mortensen, E., Deng, H. and Shapiro, L. 2005. A SIFT descriptor with global context. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference*, 184–190 vol. 1.

Murphy, K., Torralba, A., Eaton, D. and Freeman, W. 2006, In Toward Category-Level Object Recognition, In *Toward Category-Level Object Recognition*, 382–400, Springer, 382–400.

Murthy, V. N., Can, E. F. and Manmatha, R. 2014. A Hybrid Model for Automatic Image Annotation. In *Proceedings of International Conference on Multimedia Retrieval*, 369–370. ACM.

O'Hara, S. and Draper, B. A. 2011. Introduction to the bag of features paradigm for image classification and retrieval. *arXiv preprint arXiv:1101.3354* .

Oquab, M., Bottou, L., Laptev, I. and Sivic, J. 2015. Is object localization for free?–Weakly-supervised learning with convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 685–694.

Pang, Y., Li, W., Yuan, Y. and Pan, J. 2012. Fully affine invariant {SURF} for image matching. *Neurocomputing* 85: 6–10.

Ping Tian, D. 2013. A review on image feature extraction and representation techniques. *International Journal of Multimedia and Ubiquitous Engineering* 8 (4): 385–396.

Plataniotis, K. and Venetsanopoulos, A. N. 2013. *Color image processing and applications*. Springer Science & Business Media.

Qi, H. and Snyder, W. E. 1999. Content-based image retrieval in picture archiving and communications systems. *Journal of Digital Imaging* 12 (1): 81–83.

Qu, Y., Wu, S., Liu, H., Xie, Y. and Wang, H. 2014. Evaluation of local features and classifiers in BOW model for image classification. *Multimedia Tools and Applications* 70 (2): 605–624.

Rassem, T. H. and Khoo, B. E. 2011. Object class recognition using combination of color SIFT descriptors. In *Imaging Systems and Techniques (IST), 2011 IEEE International Conference*, 290–295. IEEE.

Russell, B. C., Torralba, A., Murphy, K. P. and Freeman, W. T. 2007. LabelMe: A Database and Web-Based Tool for Image Annotation. *International Journal of Computer Vision* 77 (1-3): 157–173.

Saito, P. T. M., de Rezende, P. J., Falcão, A. X., Suzuki, C. T. N. and Gomes, J. F. 2013. A data reduction and organization approach for efficient image annotation. In *Proceedings of the 28th annual ACM symposium on applied computing*, 53–57. ACM, New York, New York, USA.

Saleem, S., Bais, A. and Sablatnig, R. 2012. A performance evaluation of SIFT and SURF for multispectral image matching. In *International Conference Image Analysis and Recognition*, 166–173. Springer.

San Biagio, M., Bazzani, L., Cristani, M. and Murino, V. 2014. Weighted bag of visual words for object recognition. In *2014 IEEE International Conference on Image Processing (ICIP)*, 2734–2738. IEEE.

Shabou, A. and LeBorgne, H. 2012. Locality-constrained and spatially regularized coding for scene categorization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012)*, 3618–3625. IEEE.

Shi, J. and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (8): 888–905.

Snoek, C. G., Worring, M. and Smeulders, A. W. 2005. Early versus late fusion in semantic video analysis. In *Proceedings of the 13th annual ACM international conference on Multimedia*, 399–402. ACM.

Su, Y. and Jurie, F. 2012. Improving image classification using semantic attributes. *International Journal of Computer Vision* 100 (1): 59–77.

Talib, A., Mahmuddin, M., Husni, H. and George, L. E. 2013. A weighted dominant color descriptor for content-based image retrieval. *Journal of Visual Communication and Image Representation* 24 (3): 345–360.

Tan, P.-N., Steinbach, M. and Kumar, V. 2005, Chap 8: Cluster Analysis: Basic Concepts and Algorithms.

Tang, J., Yan, S., Zhao, C., Chua, T.-S. and Jain, R. 2013. Label-specific training set construction from web resource for image annotation. *Signal Processing* 93 (8): 2199–2204.

Tian, D. 2014. Semi-supervised Learning for Automatic Image Annotation Based on Bayesian Framework. *International Journal of Control and Automation* 7 (6): 213–222.

Tian, D. 2015. Support Vector Machine for Automatic Image Annotation. *International Journal of Hybrid Information Technology* 8 (11): 435–446.

Tousch, A. M., Herbin, S. and Audibert, J. Y. 2012. Semantic hierarchies for image annotation: A survey. *Pattern Recognition* 45 (1): 333–345.

Tsai, C.-F. 2012. Bag-of-Words Representation in Image Annotation: A Review. *ISRN Artificial Intelligence* 2012: 1–19.

Tsai, C.-F. and Lin, W.-C. 2009. A Comparative Study of Global and Local Feature Representations in Image Database Categorization. In *2009 Fifth International Joint Conference on INC, IMS and IDC*, 1563–1566.

Van De Sande, K., Gevers, T. and Snoek, C. 2010. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (9): 1582–1596.

Van de Weijer, J., Gevers, T. and Bagdanov, A. D. 2006. Boosting color saliency in image feature detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions* 28 (1): 150–156.

Van De Weijer, J. and Khan, F. S. 2013, In Computational Color Imaging, In *Computational Color Imaging*, 25–34, Springer, 25–34.

Vedaldi, A. and Fulkerson, B. 2010. VLFeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM International Conference on Multimedia*, 1469–1472. ACM.

Verma, Y. and Jawahar, C. 2013. Exploring SVM for Image Annotation in Presence of Confusing Labels. In *BMVC*, 1–11. British Machine Vision Association.

Vigo, D. A. R., Khan, F. S., van de Weijer, J. and Gevers, T. 2010. The Impact of Color on Bag-of-Words Based Object Recognition. In *2010 20th International Conference on Pattern Recognition*, 1549–1553. IEEE.

Wang, C. and Huang, K. 2015. How to use Bag-of-Words model better for image classification. *Image and Vision Computing* 38: 65–74.

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T. and Gong, Y. 2010. Locality-constrained linear coding for image classification. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3360–3367. IEEE.

Wang, L., Xue, J., Zheng, N. and Hua, G. 2011. Automatic salient object extraction with contextual cue. In *2011 International Conference on Computer Vision*, 105–112. IEEE.

Wang, P., Wang, J., Zeng, G., Feng, J., Zha, H. and Li, S. 2012. Salient object detection for searched web images via global saliency. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3194–3201. IEEE.

Westerveld, T., De Vries, A. P., van Ballegooij, A., de Jong, F. and Hiemstra, D. 2003. A probabilistic multimedia retrieval model and its evaluation. *EURASIP Journal on Applied Signal Processing* 2003: 186–198.

Yan, Q., Xu, L., Shi, J. and Jia, J. 2013. Hierarchical saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1155–1162.

Yang, J., Yu, K., Gong, Y. and Huang, T. 2009. Linear spatial pyramid matching using sparse coding for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009.*, 1794–1801. IEEE.

Yang, J., Yu, K. and Huang, T. 2010. Supervised translation-invariant sparse coding. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, 3517–3524. IEEE.

Yang, M., Zhang, L., Feng, X. and Zhang, D. 2011. Fisher discrimination dictionary learning for sparse representation. In *Computer Vision (ICCV), 2011 IEEE International Conference*, 543–550. IEEE.

Yang, N.-C., Kuo, C.-M., Chang, W.-H., Lee, T.-H. et al. 2008. A Fast Method for Dominant Color Descriptor with New Similarity Measure. *Journal of Visual Communication and Image Representation* 19 (2): 92–105.

Yu, K., Zhang, T. and Gong, Y. 2009. Nonlinear learning using local coordinate coding. In *Advances in Neural Information Processing Systems*, 2223–2231.

Yuan, Y., Wu, F., Shao, J. and 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.

Zhang, D., Islam, M. M. and Lu, G. 2012. A review on automatic image annotation techniques. *Pattern Recognition* 45 (1): 346–362.

Zhang, D., Monirul Islam, M. and 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.

Zhang, D. Z. D. 2004. Improving image retrieval performance by using both color and texture features. In *Third International Conference on Image and Graphics (ICIG'04)*, 172–175.

Zhang, H., Berg, A. C., Maire, M. and Malik, J. 2006. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2126–2136. IEEE.

Zhong, S.-h., Liu, Y., Liu, Y. and Chung, F.-l. 2012. Region level annotation by fuzzy based contextual cueing label propagation. *Multimedia Tools and Applications* 70 (2): 625–645.

Zhou, X. S. and Huang, T. S. 2000. CBIR: from low-level features to high-level semantics. In *Electronic Imaging*, 426–431. International Society for Optics and Photonics.