COMPUTATIONAL FRAMEWORKS FOR EFFICIENCY ENHANCEMENT OF CONTENT BASED IMAGE RETRIEVAL SYSTEMS

Thesis Submitted to

Cochin University of Science and Technology

in partial fulfilment of the

requirements for the award of the Degree of

Doctor of Philosophy

Under

Faculty of Technology

By

VIMINA E R
Reg. No: 4046

Under the Guidance of

Dr. K POULOSE JACOB

DEPARTMENT OF COMPUTER SCIENCE
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
Kochi – 682022

January 2017

Computational Frameworks for Efficiency Enhancement of Content Based Image Retrieval Systems

Ph.D Thesis

Author

Vimina E R
Department of Computer Science
Rajagiri College of Social Sciences
Cochin - 683 104, Kerala, India
vimina@rajagiri.edu

Supervisor

Dr. K, Poulose Jacob
Pro–Vice–Chancellor
Professor in Computer Science
Cochin University of Science and Technology
Cochin - 682 022, Kerala, India
kpj@cusat.ac.in

January 2017

Dedicated to my
Parents
Husband
&
Kids

Certificate

This is to certify that the thesis entitled ***"Computational Frameworks for Efficiency Enhancement of Content Based Image Retrieval Systems"*** *is a bona fide record of the research work carried out by* ***Ms. Vimina E R*** *under my supervision in the Department of Computer Science, Cochin University of Science and Technology, Kochi 22. The results presented in this thesis or parts of it have not been presented for the award of any other degree.*

09-01-2017

Dr. K Poulose Jacob
(Supervising Guide)
Pro–Vice–Chancellor
Professor in Computer Science
Cochin University of Science and Technology

Certificate

This is to certify that all the relevant corrections and modifications suggested by the audience during the pre-synopsis seminar and recommended by the Doctoral Committee of the candidate have been incorporated in the thesis entitled ***"Computational Frameworks for Efficiency Enhancement of Content Based Image Retrieval Systems".***

09-01-2017

Dr. K Poulose Jacob
(Supervising Guide)
Pro–Vice–Chancellor
Professor in Computer Science
Cochin University of Science and Technology

Declaration

*I hereby declare that the thesis entitled "**Computational Frameworks for Efficiency Enhancement of Content Based Image Retrieval Systems**" is the authentic record of research work carried out by me, for my Doctoral Degree under the supervision and guidance of Dr. K Poulose Jacob, Pro-Vice-Chancellor, Cochin University of Science and Technology, and that no part thereof has previously formed the basis for the award of any degree or diploma or any other similar titles or recognition.*

Kochi **Vimina E R**
09-01-2017

Acknowledgements

First and foremost, I thank God Almighty for the wisdom and grace bestowed upon me throughout my life and for reasons too numerous to mention.

I wish to express my sincere and heartfelt gratitude to my research guide Prof. Dr. K Poulose Jacob, Pro-Vice-Chancellor, Cochin University of Science and Technology for his valuable guidance, keen observations, suggestions and encouragement throughout the course of this research work. He gave me the freedom to explore a variety of topics and whenever I struggled, his generous support and suggestions always came just at the right time.

I am greatly indebted to the Management of Rajagiri College of Social Sciences, Kochi, for permitting me to do this research work and for providing all the necessary facilities. My sincere thanks to Dr. Joseph I Injodey, Executive Director, Rajagiri College of Social Sciences and Dr. Binoy Joseph, Principal, Rajagiri College of Social Sciences for the unfailing support and encouragement extended to me during this research work.

I am grateful to Dr. Sumam Mary Idicula, Professor and Head, Department of Computer Science, Cochin University of Science and Technology for her support and for providing me all the facilities in the department for carrying out the research.

I wish to express my gratitude to Dr. Supriya M H, Professor and Head, Department of Electronics, Cochin University of Science and Technology for her monitoring, valuable suggestions and for finding time to discuss my work even during her busy schedule.

My sincere thanks to Dr. G Santhosh Kumar, Mr. K B Muralidharam and all the faculty members of the Department of Computer Science, Cochin University of Science and Technology for the help and support extended to me.

I am thankful to each and every one of the technical and office staff of the Department of Computer Science, Cochin University of Science and Technology for all the help rendered to me.

Special thanks to the lab-mates and friends of the Department of Computer Science, Cochin University of Science and Technology for providing a supportive and friendly environment during my tenure there as a research scholar.

I thank all the present and former faculty members of Department of Computer Science, Rajagiri College of Social Sciences, for their co-operation, cordial relationship and valuable help.

It is beyond words to express my gratitude to my parents Mr. E M Raveendran and Mrs. Girija Raveendran and to my siblings Nina and Deepak for the unconditional love, compassion and help given to me throughout my life. I thank my father-in-law Mr. K K Aravindakshan and mother-in-law Prof. Thankamani Aravindakshan for their wholehearted support extended to me all these years.

Words are not enough to thank my husband Sreejith, who has always been a shoulder to lean on during difficult times. I appreciate and I am thankful for the endless love and care he has given me since the day we met. Finally, I thank my dear little ones, Rahul and Rohan, for loving me, giving me joy and a fulfilling life and for being patient with me during the crystallization of this thesis.

Vimina E R

Abstract

Content-based image retrieval (CBIR) is focused on efficient retrieval of relevant images from image databases based on automatically derived imagery features. However, images with high feature similarities to the query image may be very different from the query in terms of semantics. This discrepancy between low-level content features and high-level semantic concepts is known as "semantic gap", an open challenging problem in every CBIR systems. Various techniques ranging from single query based systems to multiple query and Bag of Visual Words (BoVW) approaches have been employed to address this issue. Single query CBIR systems use global features, local features or both global and local features extracted from the image to retrieve content related images from the database. As the global features cannot sufficiently capture the important properties of individual objects, region-based approaches are developed that utilize features extracted from identified regions of the images to retrieve similar images. Limitations of such systems are the difficulty in identifying the significance of different regions and the requirement of efficient region matching algorithms to effectively find the similarity between images. In addition, they are computationally costly, time consuming and are not suitable for large-scale retrieval. Alternative is to use BoVW framework, which is, reported to have high scalability and exhibit good performance in object recognition and classification. However, their retrieval rate is limited in natural image datasets due to the presence of rich colour and texture information, which are not prime features, considered in the case of objects. Some other methods employ multiple queries for effective retrieval. They exploit the fact that the relationship of visual content

in the multiple images contained in the query image set can represent the user's requirement more precisely than a single query; leading to enhanced retrieval performance. Major challenges in such systems are the aggregation of features extracted from many images and the computation of similarity between the query image set and the target images.

In the light of this, the purpose of this research work is to design and develop methodologies for enhancing the retrieval efficiency of the aforementioned approaches considering the scalability and response time. Focusing on the region identification and matching issues in the region based retrieval systems, a salient sub-block based framework along with region matching technique employing minimum distance algorithm is proposed for faster and enhanced retrieval. The method uses fixed-block segmentation approach for dividing the image into regions and utilizes their respective edge information for determining the salience. The similarity between the most salient sub-blocks of the query and target images in the dataset are then computed by employing the minimum distance algorithm. Repudiation of the less salient sub-blocks in similarity computation accelerates the retrieval process without compromising the efficiency of retrieval when compared with systems employing all the sub-blocks.

The present work also focuses on the possibility of integrating colour and edge information with the interest-point based invariant descriptors in the creation of visual bags for improving the retrieval performance in natural image databases. For this, a multi-fusion approach is adopted, in which, the edge-colour features of the images are combined through early fusion for creating the

vocabulary of visual words. The image histogram built with the resultant vocabulary is then combined with histogram constructed with vocabulary of invariant features through late fusion to characterize the image. The incorporation of additional information helps in providing a better representation of the images and aids in improving the retrieval performance.

Additionally, for multi- query CBIR systems, a new feature replacement algorithm is proposed for similarity computation that can contribute better retrieval results without query refinement or feature reweighting. The algorithm determines the similarity between query and target images in the database by computing the cumulative sum of the displacements of the query-set images' centroid caused by replacing each element in the query image set with the candidate images in the database; thereby effectively accumulating the dissimilarity between the images in query-set and the target dataset images.

Contents

List of Tables

List of Figures

Abbreviations

AQE	Average Query Expansion
BOLD	Binary Online Learned Descriptor
BoVW	Bag of Visual Words
BRISK	Binary Robust Invariant Scalable Keypoints
CBIR	Content Based Image Retrieval
CCV	Colour Coherence Vector
CEDD	Colour and Edge Directivity Descriptor
CLD	Colour Layout Descriptor
CNN	Convolutional Neural Networks
CSD	Colour Structure Descriptor
DCD	Dominant Colour Descriptor
DoG	Difference of Gaussian
DQE	Discriminative Query Expansion
EHD	Edge Histogram Descriptor
EMD	Earth Movers Distance
FCTH	Fuzzy Color and Texture Histogram
GLCM	Gray Level Co-occurrence Matrix
GPU	Graphics Processing Unit
HOG	Histogram Oriented Gradients
HSV	Hue Saturation Value
HTD	Homogeneous Texture Descriptor
IRM	Integrated Region Matching
LBP	Local Binary Patterns

LECoP	Local Extrema Co-occurrence Patterns
LoG	Laplacian of Gaussian
MAP	Mean Average Precision
MCMCM	Modified Color Motif Co-occurrence Matrix
MSHP	Most Significant Highest Priority
ORB	Oriented FAST and Rotated BRIEF
QBE	Query By Example
QBIC	Query By Image Content
RBIR	Region Based Image Retrieval
RLBP	Robust LBP
SCD	Scalable Colour Descriptor
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Features
SVM	Support Vector Machines
TBD	Texture Browsing Descriptor

Chapter

1 INTRODUCTION

Contents

1.1 Overview

The volume of image databases are growing at an exponential rate with the steady growth of computing power, declining cost of storage devices and increasing access to Internet. Hence, to effectively store, manage, and retrieve information according to various needs, it is imperative to advance automated image learning techniques. In the traditional method of text-based image retrieval, the image search is mostly based on textual description of the image found on the web pages containing the image and the file names of the image. The problem here is that the accuracy of the search result highly depends on the textual description associated with the image. In addition, un-annotated image collections cannot be searched. Language dependency, user subjectivity etc. are other issues. An alternate method is to retrieve images based on the content of the image. In Content Based Image Retrieval (CBIR) systems, the visual contents of the image such as colour, texture, shape or any other information

that can be automatically extracted from the image itself are extracted and is used as a criterion to retrieve content related images from the database. The retrieved images are then ranked according to the relevance between the query image and images in the database in proportion to a similarity measure calculated from the features. Hence, the accuracy of a CBIR system greatly depends on the low-level features extracted from the image and the effectiveness with which they represent its high-level semantics, which is often termed as the semantic gap. The method used for similarity computation and the type of images in the dataset also plays a major role. Though numerous methods have been proposed and studied in the past, CBIR is a complex and challenging problem and is still far from solved because of the diverse algorithms required all over the retrieval process varying from feature extraction, similarity computation, retrieval strategies, etc. and post retrieval methods such as relevance feedback and re-ranking used for enhancing the accuracy of retrieval. Additionally, compared to the retrieval in narrow domain systems such as finger print recognition, medical imagery retrieval, satellite imagery retrieval etc., retrieval in broad domain systems such as photo collections, natural image datasets and Internet are extremely challenging due to the diversity of the images in the database and the scarcity of the information that is available about the database. A comparison of both the systems is depicted in Table 1.1 (Smeulders et al., 2000). Hence, this research work mainly focuses on the general broad domain CBIR and tries to find solution or alternative to some key issues.

Table 1.1 Comparison of narrow and broad domain CBIR approaches (Smeulders et al., 2000)

	Narrow Domain	**Broad Domain**
Aimed application	Specific	Generic
Type of application	Professional	Public
Variance of content	Low	High
Sources of knowledge	Specific	Generic
Semantics	Homogeneous	Heterogeneous
Size	Small/ medium	Small/ Medium/Large
Ground truth	Likely	Unlikely
Content description	Objective	Subjective
Scene and sensor	Possibly controlled	Unknown
Evaluation	Quantitative	Qualitative
Tools	Model driven, specific invariants	Perceptual, cultural, general invariants
System architecture	Tailored database driven	Modular interaction driven
A source of inspiration	Object recognition	Information retrieval

1.2 Motivation

From a computational perspective, a typical CBIR system views the query image and images in the database (target images) as a collection of features, and the retrieval is performed based on the relevance between the query image and any target images according to their similarity. In this sense, these features, or signatures of images, characterize the content of images. According to the scope of representation, features fall roughly into

two categories: global features and local features. Global features describe an image as a whole and the extracted features include texture histogram, colour histogram, colour layout etc. of the whole image, and features selected from multidimensional discriminant analysis of a collection of images (Chen, Li & Wang, 2006). In the latter category are features extracted from sub-images/ sub-blocks, segmented regions, and interest points.

The global feature based retrieval methods are fast and computationally efficient but retrieval performance is limited, as they cannot sufficiently capture the important properties of individual objects. This can be circumvented by employing region-based approaches that utilize features extracted from local regions of the image to retrieve similar images. Pixel-wise segmentation methods or block based segmentation methods are usually used for identifying the regions. Major challenges in such systems are the identification of significant regions in an image and the requirement of efficient region matching algorithms to effectively find the similarity between images. Computational cost and response time are other concerns. In order to further improve the performance of retrieval, methods such as relevance feedback, re-ranking etc. are often incorporated with the CBIR systems. The idea is to extract more information from the initial retrieved results and to improve the performance of subsequent retrievals.

Though the region based approaches work well in small- medium databases, their scalability is limited leading to the development of Bag of Visual Words (BoVW) model based retrieval systems, which are reported to

have good scalability. In the BoVW model, a large set of local descriptors extracted from many images is converted to a final global representation of the image. This is performed by a succession of two steps: coding and pooling. Coding consists of hard assigning each local descriptor to the closest visual word, while pooling averages the local descriptor projections. The final BoVW vector can thus be regarded as a histogram counting the occurrences of each visual word in the image (Law, Thome & Cord, 2014), i.e., every image in the database can be represented as a histogram built over the visual words. The BoVW based approaches are widely accepted for object recognition and classification purposes. However, their retrieval rate is reported to be limited in natural image datasets, as natural images vary widely in colour and texture constitution, which are not often considered important in object recognition.

Other methods employ multi-query approaches to further ameliorate the performance of CBIR systems, considering the fact that information extracted from more than one image can provide a better representation of the user requirement. Moreover, most of the image collections and photo sharing websites contain multiple images of the same object or scenery. In a general multi-query system, users input multiple query images including both positive and negative samples using which a discriminative classification model is learned, to rank all images in the dataset. Support vector machines, nearest neighbour classifier etc. are usually used for this purpose. One of the limitations here is the requirement of both positive and negative samples. Also, the performance of the system depends on the

number and quality of samples used to learn the model. Some systems use only positive samples and employs methods such as feature reweighting, joint query averaging, etc. to refine the query. Major challenges in such systems are the aggregation of features extracted from multiple images and the computation of similarity between the query image set and the candidate images.

1.3 Objectives

Considering the above-mentioned issues, the main objective of this research work is to develop methodologies to improve the retrieval efficiency of broad domain CBIR systems. The existing techniques in single query, multiple query and BoVW approaches are studied and attempts are made to find solution to some of the identified problems leading to enhanced retrieval performance.

The research work mainly focuses on the following objectives:

1. To survey various single query CBIR systems employing region based and sub-block based approaches.
2. To propose a novel salient sub-block method to identify the significance of the sub-blocks in an image.
3. To propose an effective region matching technique to reduce the computational time.
4. To explore the various image features considered in BoVW model based image retrieval systems and to propose a combined feature approach for enhancing retrieval, especially in natural image datasets.

5. To survey various multi-query CBIR approaches and propose an algorithm for similarity computation.

1.4 Contributions of the Research Work

This research work focuses on the design, implementation and evaluation of different CBIR frameworks, with the aim to increase the retrieval accuracy while reducing the overall computational time. The main contributions are listed below:

Salient sub-block framework and a new region-matching algorithm for single query CBIR

In this work, various region based and sub-block based approaches for CBIR are explored and a salient sub-block method is proposed. It is an extension of the sub-block based retrieval method in which the image is divided into different blocks of simple geometric shapes and features extracted from these sub-blocks are used to represent the image and for similarity computation. Unlike the existing methods in which all the sub-blocks are involved in similarity computation, in the proposed method only salient sub-blocks are used for this purpose. The salience score of a sub-block is computed based on the presence of object in it, which is coarsely determined by finding the white pixel density in the sub-block and applying morphological operations. Sub-blocks with less salience score are abstained from participating in the similarity computation process, leading to

considerable reduction in computation time. An image-matching algorithm is also proposed which can be used for both sub-block based retrieval and region based retrieval. Experimental results prove that the proposed matching algorithm improves the response time as well as the retrieval performance measured in terms of precision and recall. The main contributions of this work are:

- A method to determine the salience of sub-blocks in an image for CBIR systems employing fixed block segmentation approach, based on which, less salient sub-blocks can be circumvented from participating in the region matching process.
- A method for matching image regions in Region Based Image Retrieval systems, to retrieve relevant images in response to the given query.

A combined feature approach for Bag of Visual Words model based retrieval.

In this work, the possibility of integrating multiple image cues with the invariant interest point descriptors in Bag of Visual Words framework is investigated. BoVW model has been very successfully employed in the fields of object recognition and classification in recent years due to its scalability and high precision. In this method, interest point descriptors extracted from large set of images are clustered to form visual words and each image in the database is encoded with these visual words to be characterized in the form of

a histogram. However, their performance is found to be incompetent compared to region-based systems in natural image datasets, as they are rich in colour and texture information, which are not often considered generally in systems employing BoVW model. The proposed work tries to incorporate these features along with the invariant descriptors aiming to achieve better retrieval performance. Here, a visual vocabulary is constructed using a combined edge-colour descriptor extracted from local image patches in addition to the vocabulary constructed using interest point descriptors. Feature histograms are built using these vocabularies which are further fused together to characterize the image. Various combinations of feature integration are studied and the results are presented. The main focus of this work include:

- A combined edge-colour descriptor to describe local regions of an image.
- A feature fusion methodology exploiting both early and late fusion approaches to characterize images for CBIR applications.

Feature replacement based similarity computation for multi-query CBIR

Here, a new algorithm to compute the similarity between the images in the query image set and the target images is proposed for multiple- query CBIR systems. Multiple queries are employed in image retrieval systems, as the information contained in more than one query can represent the user's need more precisely than a single query, leading to enhanced retrieval

performance. Generally, the features extracted from these queries are used either for reforming the query to a better representation or to learn a model using supervised learning algorithms to enhance retrieval. The former method includes joint query averaging, feature reweighting, query point movement etc., which generally involves query refinement, while the latter uses support vector machines and other machine learning algorithms. Despite the different methods used, ultimately the retrieval is performed based on the similarity score of the query set with the target images in the database. The proposed feature replacement algorithm utilizes the features extracted from the images in the query set only for similarity computation without any query refinement. Here, the similarity is computed by considering the displacements of the centroid of the query set caused by the replacement of each element in it with an element from the target image set. The cumulative sum of these displacements add more discriminative property in deciding the similarity between the query image set and the target images as individual images' dissimilarity is taken into consideration. The proposed algorithm can also be used effectively with CBIR systems employing relevance feedback, as significant improvement in performance can be achieved with minimal user feedback. The main contribution of this work is:

- A feature replacement algorithm to compute the similarity between the query image set and target images in the dataset in a multi-query image retrieval system.

1.5 Outline of the Thesis

Chapter 1 provides an introduction about the content based image retrieval systems and its need in the present world. Significance of the present study, objectives and contributions of this research work are also summarized.

Chapter 2 describes CBIR in detail including various steps involved in retrieval process varying from feature extraction, similarity computation etc. to retrieval and methods to enhance the efficiency of retrieval. A review of various image retrieval techniques is also presented.

Chapter 3 explains the design, implementation and evaluation of the proposed salient sub-block based framework. An evaluation of different region matching techniques including the proposed minimum distance algorithm is also presented.

Chapter 4 describes a combined edge and colour feature descriptor for BoVW model based retrieval. Late fusion of the histogram built over this descriptor with the traditional invariant feature based histogram is discussed and performance evaluation is presented.

Chapter 5 provides an overview of various methods of retrieval enhancement using multiple queries. A new feature replacement algorithm for similarity computation in multiple query environments is proposed and performance evaluation presented.

Chapters 6 summarizes the research work and its limitations, highlights the contributions and draws some conclusions. Some new directions for future research work are also discussed in this chapter.

…….ଔ……

Chapter 2

LITERATURE REVIEW

Contents

This chapter presents a brief overview of a typical CBIR framework including image content description, system learning, benchmark datasets, similarity matching and performance evaluation. Different CBIR systems such as region based, block based and local feature based approaches are discussed outlining the merits and demerits of each. Description of various techniques used for enhancing the retrieval performance are also discussed.

2.1 Introduction

Large repositories of images have become a commonplace reality due to the rapid advances in digital imaging technology, declining cost of storage and the ubiquitous use of pictorial information in every field of life. However, maintaining such repositories is meaningless in the absence of technologies that can enable a user to extract or retrieve information of interest as and when required. The CBIR systems have been developed with this primary objective; availing the required content related information to the user in response to a query, by automatically analysing the image content.

2.2 CBIR Architecture

As aforementioned, an image retrieval system retrieves content related images from the database in response to a user query. In the context of CBIR, content refers to the visual features such as colour, texture, shape or any other automatically extracted information that describes the image. The similarity between the visual features of the query and the images in the database is then computed and the top ranked images are returned as retrieval results. The performance of the system depends on many factors ranging from selection of query, feature extraction etc. to the metric used to compute the similarity between imagery features. Relevance feedback and other query refinement techniques can be also incorporated to improve the retrieval performance. The following sections provide an overview of the components of a CBIR system (Figure 2.1) along with its various implementations.

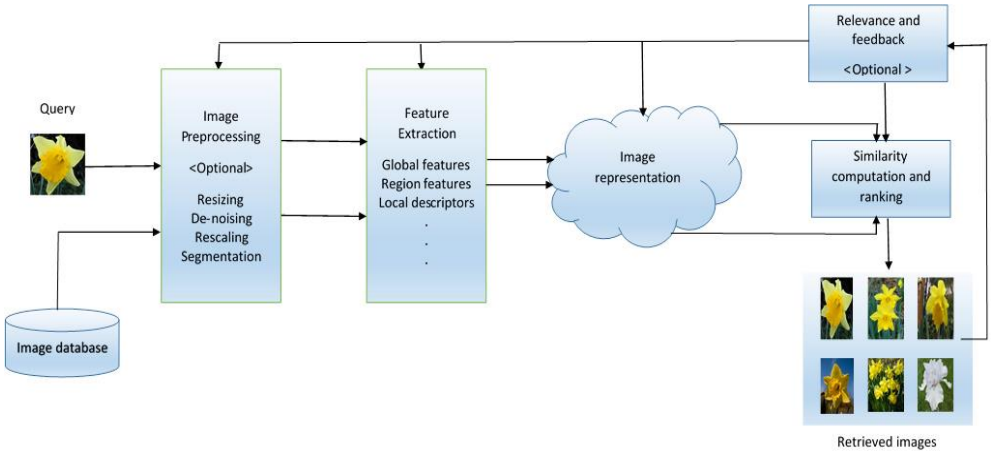


Figure 2.1 The architecture of a general CBIR system

2.2.1 User Query

In an image retrieval system, the user expresses his requirement in the form of a query. Based on the kind of information used during the retrieval process, the query can be in the form of text (Query – By- text) or sample image itself (Query- By- Example). The Query- By- Text or Annotation Based Image Retrieval approaches are of cross medium type, as the queries issued by the user are in the form of text and the search targets are images. On the other hand, the Query- By- Example are of mono-medium type because both the user's query and the search targets are images (Huang, Gao & Chan, 2010). The query by example (QBE) is the generally accepted paradigm in most of the CBIR systems. Here the user inputs image or sketches as query and features extracted from this are used for retrieving similar images from the database. Some of the recent retrieval systems encourage the use of multiple queries also to enhance the performance of retrieval, as the information extracted from more than one image can provide better characterization of the user's requirement than a single image (Tahaghoghi, Thom & Williams, 2001). Multiple example images are provided by the user at initial query time or attained through relevance feedback.

2.2.2 Visual Features

Upon receiving a query image, a CBIR system views the query and other images in the database as a collection of visual features. The feature extraction phase plays a key role in retrieval as performance of the system

greatly depends on the ability of these extracted visual features in describing the image content. Numerous features can be extracted from an image; among which colour, texture and shape features are the well- studied and extensively used ones for CBIR applications. In addition, local features such as Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) have also gained attention because of their invariance property and performance in object retrieval.

2.2.2.1 Colour

Colour is one of the most effective, simplest and widely used low-level visual features employed in CBIR. Human visual perception can easily discriminate different colours compared to other features. It is also a robust feature, as it does not depend on the state of image such as the direction, size, and angle. According to various application requirements, colour features can be defined over different colour spaces such as RGB, CMY, XYZ, HSV or HSL or HSB, YCrCb, CIE-L*u*v*, CIE-L*a*b* etc. The RGB colour space is an additive colour space with three primary colours red, green and blue using which various secondary colours can be generated. Despite its simplicity in representation, the RGB space is less close to human visual perception, because of which the HSV and L*a*b colour spaces are more popular in CBIR systems than RGB. The HSV model has three constituents namely hue, saturation and value. Hue refers to the purity of colour and is described by a number that specifies the position of the corresponding pure colour on the colour wheel as a fraction

between 0 and 1. The saturation (S) of a colour describes how white the colour is. For example, pure red is fully saturated, with a saturation of 1; tints of red have saturations less than 1; and white has a saturation of 0. The value (V) of a colour, also called its lightness, describes how dark the colour is. Value of 0 is black, with increasing lightness moving away from black.

Lab colour space is a colour-opponent space with dimension L for lightness, and *a*, *b* for the colour-opponent dimensions, based on nonlinearly compressed (e.g. CIE XYZ colour space) coordinates. It is device independent; i.e., the colours are defined independent of their nature of creation and the device they are displayed on. The three coordinates of CIE-Lab represent the lightness of the colour (L* = 0 yields black and L* = 100 indicates diffuse white; specular white may be higher), its position between red/magenta and green (a*, negative values indicate green while positive values indicate magenta) and its position between yellow and blue (b*, negative values indicate blue and positive values indicate yellow).

Once the colour space is chosen, colour features of the image are extracted. Commonly used colour descriptors include colour histogram, colour coherence vector, colour moments, colour- correlogram, colour auto correlogram and colour co-occurrence matrix (Swain & Ballard, 1991; Stricker & Orengo, 1995; Pass & Zabih, 1996; Huang, et al., 1997). Colour histogram provides the distribution of colours in an image. It focuses only on the proportion of the number of different types of colours, and avoids spatial location of the colours. The colour coherence vector (CCV) incorporates

spatial information by measuring the spatial coherence of the pixels with a given colour. For example, if the red pixels in an image are members of large red regions, this colour will have high coherence, while, if the red pixels are widely scattered it will have low coherence. Colour- correlogram and colour auto correlogram also incorporate spatial information.

The MPEG-7 family of descriptors includes various colour descriptors such as Dominant Colour Descriptor (DCD), Scalable Colour Descriptor (SCD), Colour Layout Descriptor (CLD) and Colour Structure Descriptor (CSD). The DCD represents the colour information of the whole image or regions in an image by a small number of representative colours (Manjunath et al., 2001). The SCD is derived from a colour histogram defined in the uniformly quantized HSV colour space. A total of 256 coefficients is used to represent the descriptor. It is invariant to rotation and transformation and presents good tolerance to change of lighting conditions and hue variations. The CLD represents the spatial distribution of the colour in images. In order to incorporate the spatial relationship, each image patch is divided into 8×8 discrete blocks and dominant colour in each block is detected. It is a very compact descriptor and is suitable for fast browsing and search applications. It can be applied to still images as well as to video segments. CSD is also computed based on colour histogram, which captures the local colour distribution in an image using a structuring element. It counts the number of times a particular colour is contained within the structuring element as the structuring element scans the image.

An alternative way to describe colour is by means of colour names. Colour names are linguistic labels humans usually use to describe the colours in the world such as 'red', 'black', 'magenta' etc. In (Van de Weijer et al., 2009), eleven colour names of English language are learnt from Google images resulting in partitioning the colour space to eleven regions. An eleven dimensions local colour descriptor is then deduced by counting the occurrence of each colour name over a local neighbourhood.

2.2.2.2 Texture

Another significant visual feature is texture. Texture can be considered as repeating patterns of local variation of pixel intensities. Unlike colour, texture occurs in a region of an image than at a point. Image features like contrast, coarseness, directionality, regularity, entropy etc. can be measured with various texture descriptors. A number of techniques have been used for measuring the texture features such as fractals (Kaplan, Murenzi & Namuduri, 1998), wavelets, co-occurrence matrix, Gray Level Co-occurrence Matrix (GLCM) (Haralick & Shanmugan, 1973), Local Binary Patterns (LBP), Tamura features etc. In addition, the MPEG-7 multimedia content description interface includes three texture descriptors namely Texture Browsing Descriptor(TBD) that characterise perceptual directionality, regularity, and coarseness of a texture; Homogeneous Texture Descriptor (HTD) to quantitatively characterise homogeneous texture regions for similarity retrieval using local spatial statistics of the texture obtained by scale and orientation-selective Gabor filtering, and Edge

Histogram Descriptor (EHD) to characterise non-homogeneous texture regions (Manjunath et al., 2001; Sikora, 2001). In other methods like spectral texture methods, images are transformed into frequency domain using certain spatial filter bank. Texture features are then extracted from the transformed spectra using statistics. Due to the large neighbourhood support of the filters, spectral methods are very robust to noise. Common spectral methods include Fourier transform (Hervé & Boujemaa, 2007), Wold texture (Long, Zhang & Feng, 2003; Liu & Picard, 1996), discrete cosine transform (DCT) (Ngo, Pong & Chin, 2001; Lu, Li & Burkhardt, 2006), wavelet transform (Datta et al., 2008; Wang, Li & Wiederhold, 2001; Do & Vetterli 2003; Bhagavathy & Chhabra, 2007; Suematsu et al., 2002) and Gabor filters (Manjunath & Ma, 1996).

2.2.2.3 Shape

Shape is another low-level feature, which plays an important role in describing image contents, especially for object retrieval. Shape representation techniques can be mainly divided into two categories, boundary based and region based (Mehtre, Kankanhalli & Lee, 1997; Wang, Yu, & Yang, 2011). Boundary based methods use contour or border of the object ignoring the interior of the object under consideration, while the region-based methods take into account both internal details and boundary details (Shahabi & Safar, 2007). Commonly used boundary based methods employ Fourier descriptors (based on objects' shape radii) (Sajjanhar, Lu & Wright, 1997) grid-based methods based on chain codes (Sajjanhar & Lu,

1997), Delaunay triangulation method (Tao & Grosky, 1998) (based on corner points), MBC-based methods (based on minimum bounding circles and angle sequences) etc., for shape feature extraction. Region-based methods represent shape based on the interior descriptions of the object's "body" within the closed boundary. These methods commonly employ moment descriptors such as Hu moments (Hu, 1962), Zernike moments (Khotanzad & Hong, 1990), Legendre moments (Mukundan & Ramakrishnan, 1998; Mukundan, Ong & Lee, 2001) etc. to represent shape. Other methods include polygonal approximation, deformable templates, B-splines, curvature scale space (CSS), aspect ratio, circularity, and consecutive boundary segments (Liu et al., 2007).

Shape features are often used effectively for specific applications involving man- made objects. However, as these features are susceptible to image transformations like translation, scaling and rotation, they are not as popular as colour and texture features in retrieval applications and are often used in conjunction with other image features.

2.2.2.4 Composite Feature Descriptors.

In addition to the aforementioned basic features, some systems employ composite descriptors for image representation. The Colour and Edge Directivity Descriptor (CEDD) integrates colour information and edge distribution in images and characterize the images in the form of histograms (Chatzichristofis & Boutalis, 2008a). The Fuzzy Colour and Texture Histogram descriptor (FCTH), a variant of CEDD, also integrates the colour

and texture information (Chatzichristofis & Boutalis, 2008b). Descriptors like Local Extrema Co-occurrence Patterns (LECoP) (Verma, Raman & Murala, 2015), Modified Colour Motif Co-occurrence Matrix (MCMCM) (Subrahmanyam et al., 2013) etc. also combines colour and texture features for characterizing images.

2.2.2.5 Local Image Descriptors

In the context of retrieval, the images may be described either by global descriptors or by local descriptors. In the former case, a single descriptor captures the entire information of the image usually by averaging the image features. They often fail to represent the high-level image semantics as the global features cannot discriminate the objects and the background depicted in the images. Local features are computed from local image regions and have several advantages over global features. They are distinctive, robust to rotation, scale, occlusion, and do not require segmentation. They are widely used in various applications such as object detection, scene recognition, image matching, image registration etc. (Zhao et. al, 2007; Arandjelović & Zisserman, 2012; Simonyan & Zisserman, 2014; Everingham et al., 2015)

Most of the recent local feature extraction techniques focus mainly on keypoints, the salient image regions, which contain the rich local information in an image, for local feature extraction. The keypoints can be automatically detected using various detectors- Laplacian of Gaussian (LoG), Difference of Gaussian (DoG), Harris Laplace, Hessian Laplace,

Harris Affine, Hessian Affine etc. being the popular ones. In LoG, the scale-space representation is built by successive smoothing of high-resolution image with Gaussian based kernels of different sizes. This is followed by the detection of a feature point if a local 3D extremum is present and if its absolute value is higher than a threshold. The LoG detector is circularly symmetric and it detects blob-like structures. In DoG, the input image is successively smoothed with a Gaussian kernel and sampled. The DoG representation is obtained by subtracting two successive smoothed images. Thus, all the DoG levels are constructed by combined smoothing and sub-sampling. The DoG is an approximate but more efficient version of LoG. The Harris Laplace detector responds to corner-like regions. It uses a scale-adopted Harris function to localize points in scale-space, and then selects the points for which the Laplacian of Gaussian attains a maximum over a scale. Keypoints of Hessian Laplace are points, which reach the local maxima of Hessian determinant in space and fall into the local maxima of Laplacian of Gaussian in a scale. Harris Affine, which is derived from Harris-Laplace, estimates the affine neighbourhood by the affine adaptation based on the second moment matrix, while Hessian Affine is achieved after the affine adaptation procedure based on Hessian Laplace (Liu, 2013). Another robust feature detector not based on keypoint is MSER (maximally stable extremal regions) (Matas et al., 2004), which is based on the concept of regions which remain "stable" over large ranges of binarization thresholds.

Once a keypoint is detected, a small patch around the point, which is also expected to have some invariance, is used to compute the descriptor. Here, invariance means that the descriptors should be robust against various image variations such as affine distortions, scale changes, illumination changes or compression artefacts (Roth & Winter, 2008). Popular descriptors include Scale Invariant Feature Transform (SIFT) (Lowe, 2004), PCA-SIFT (Ke & Sukthankar, 2004), SURF (Bay et al., 2008), ORB (Rublee et al., 2011), BRISK (Leutenegger, Chli & Siegwart, 2011) and BOLD (Tombari, Franchi & Di Stefano, 2013), of which SIFT and SURF are the most widely used ones.

Scale Invariant Feature Transform (SIFT)

SIFT is an algorithm to detect and describe regions of interest within an image which is both scale and rotation invariant. Here, keypoints/ interest points in an image are initially extracted by the SIFT detector and their descriptors are computed by the SIFT descriptor. Alternatively, the SIFT detector and SIFT descriptor can also be used independently, i.e., the SIFT detector can be used to compute the keypoints without descriptors and SIFT descriptor can be used to compute the descriptors from custom keypoints. The key points are determined by finding extrema of difference of Gaussian images, which are robust across multiple scales. Once the keypoints are detected, the circular region around the key-point is divided into 4 x 4 non-overlapping patches and the histogram gradient orientations

within these patches are calculated. Histogram smoothing is done in order to avoid sudden changes of orientation and the bin size is reduced to eight bins in order to limit the descriptor's size. This results into a 4 x 4 x 8 = 128 dimensional feature vector for each keypoint.

Speeded Up Robust Features (SURF)

SURF (Bay et al., 2008) is another robust local feature descriptor which is partly inspired by SIFT descriptor. Like SIFT, SURF is also a combination of detector and descriptor. It is faster than SIFT and is claimed by its authors to be more robust against different image transformations than SIFT. SURF uses an integer approximation of the determinant of Hessian blob detector, which can be computed with three integer operations using a pre-computed integral image to detect the interest points. Its feature descriptor is based on the sum of the Haar wavelet response around the point of interest. This can also be computed with the aid of the integral image. The SURF descriptor is a 64 dimensional concise feature vector which makes it desirable over SIFT for many applications.

Table 2.1 depicts a summary of the various image features.

Table 2.1 Summary of image features

Features	Representation/ Descriptor	Remarks
Colour	Histogram Correlogram Auto-correlogram Moments Co-occurrence matrix Colour coherence vector CLD, SCD, DCD,, CSD	• Invariant to image transformations • Limited perceptual similarity • Limited spatial information in most of the representations
Texture	Fractals GLCM LBP EHD, TBD, HTD Tamura Gabor filters Wavelets	• Describes image structure, coarseness, granularity, regularity, homogeneity etc. • Computational complexity • Sensitive to noise
Shape	Invariant moments (Zernic, Hu, Legendre etc.) Fourier descriptors B-splines Polygonal approximations Curvature scale space	• Binary representation of image objects • Sensitive to image transformations
Composite descriptors	CEDD FCTH LECoP MCMCM	• Composite colour and texture descriptors
Local descriptors	SIFT SURF BRISK ORB HOG BOLD	• Mostly invariant to image transformations • Complex to compute

2.2.3 Similarity Computation

Once the images are characterized with features, the next step in retrieval is finding the similarity between the query and the target images. Similarity measure plays an important role in CBIR as the relevance of the retrieved images greatly depends on the feature similarity (dissimilarity) between the query and the images in the dataset. A simple way to measure the image similarity is to use distance measures. However, in Region Based Image Retrieval (RBIR) systems, where multiple regions are present, similarity computation is performed in two levels. First is in the region level and second in the image level. In the first level, the similarity between two image regions based on their extracted features is computed while in the second level the overall similarity between two images, which might contain different number of regions, is computed. Several distance measures and image matching algorithms are employed for this purpose.

2.2.3.1 Distance Measures

A number of distance measures are used in CBIR systems for similarity computation, of which, Minkowski metric based distance measures are the popular ones used in image retrieval applications. The Minkowski distance of order p between two points $X= \{x_1, x_2, \ldots, x_n\}$ and $Y= \{y_1, y_2, ..., y_n\} \in R^n$ is defined as:

$$L_p = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p} \qquad (2.1)$$

Minkowski distance is typically used with p being 1 or 2. The latter is the Euclidean distance (eq. 2.2), while the former is known as the Manhattan distance or City block distance (eq. 2.3).

$$Euclidean\ distance\ (L_2) = \left(\sum_{i=1}^{n} |x_i - y_i|^2\right)^{1/2} \quad (2.2)$$

$$Cityblock\ distance\ (L_1) = \sum_{i=1}^{n} |x_i - y_i| \quad (2.3)$$

A variant of Euclidean distance, commonly used to specify weightage when multiple features are present is weighted Euclidean distance, defined as:

$$D_{L2} = \left(\sum_{i=1}^{n} w_i |x_i - y_i|^2\right)^{1/2} \quad (2.4)$$

where, w_i is the weight of the j^{th} component of X and Y. Canberra distance is another popular distance measure, defined as:

$$D_{Canberra} = \sum_{i=1}^{n} \frac{|x_i - y_i|}{x_i + y_i} \quad (2.5)$$

2.2.3.2 Image Matching Methods for RBIR Systems

In RBIR systems, a single image may contain multiple regions. Hence, to find the similarity between images, region matching algorithms such as Earth Movers Distance (EMD) (Rubner, Tomasi & Guibas, 2000), Integrated Region Matching (Li, Wang, & Wiederhold, 2000; Wang, Li & Wiederhold, 2001) algorithm (IRM), Integrated Image Matching algorithm (Hiremath & Pujari, 2008) etc. are used. The IRM algorithm and integrated image-matching algorithm are commonly employed in block based / region based image retrieval systems and are described below.

Integrated Region Matching (IRM) Algorithm

The IRM, proposed by Wang et al. is one of the most widely used methods to find the distance between two images with multiple regions that can be of unequal number. It measures the overall similarity between images by integrating properties of all the regions in the images. The algorithm is described below.

Assume that Image 1 and 2 with *m* and *n* regions be represented by region sets R_1= {r_1, r_2, . . ., r_m} and R_2= {r'_1, r'_2, …, r'_n}, where r_i or r'_i is the feature descriptor of region i. Let the distance between region r_i and r'_j be denoted by $d(r_i, r'_j)$ or $d_{i,j}$ in short. To compute the similarity measure between region sets R_1 and R_2, $D_{R1,R2}$, all regions in the two images are matched first. The matching between every r_i and r'_j is assigned a significance credit $s_{i,j}$, $s_{i,j}$> 0. The significance credit indicates the importance of the matching for determining similarity between images and it is computed from the significance assigned to each region in R_1 and R_2. Assuming that the significance of r_i in image1 is p_i and r_j in Image2 is p_j, it is required that

$$\sum_{j=1}^{n} s_{ij} = p_i, i = 1, ..., m \qquad (2.6)$$

$$\sum_{j=1}^{m} s_{ij} = p'_i, i = 1, ..., n \qquad (2.7)$$

For normalization, $\sum p_i = \sum p_j = 1$, where i=1, …, m and j=1, …, n, ensuring that all the regions play a role for measuring similarity. The distance between the two region sets R1, R2 is then computed as:

$$D_{R1,R2} = \sum s_{ij} \cdot d_{ij} \quad (2.8)$$

The algorithm is summarized as follows:

1. Set L = {}, denote M={(i, j) : i=1, …, m; j=1, …, n}.
2. Choose the minimum $d_{i,j}$ for (i,j) ϵ M – L. Label the corresponding (i, j) as (i', j').
3. $\min(p_{i'}, p'_{j'}) \rightarrow s_{i',j'}$.
4. if $p_{i'} < p'_{j'}$, set $s_{i',j} = 0$, $j \neq j'$; otherwise, set $s_{i,j'} = 0$, $i \neq i'$.
5. $p_{i'} - \min(p_{i'}, p'_{j'}) \rightarrow p_{i'}$.
6. $p'_{j'} - \min(p_{i'}, p'_{j'}) \rightarrow p'_{j'}$.
7. L + {(i', j')} → L.
8. if $\sum p_i > 0$ and $\sum p'_j > 0$, go to step 2, Otherwise stop.

The IRM algorithm considers all regions in the image for similarity computation and assign significance to each region to compensate the errors that might arise from incorrect region segmentation. This allows a single region of image1 to be matched with multiple regions of image 2.

Integrated Image Matching Algorithm

In integrated image matching algorithm (Hiremath & Pujari, 2008), the image is divided into different regions/ sub-blocks and each sub-block of the query image is matched with any sub-block of the target image. It is required that both the query and target image should have equal number of

sub-blocks. The algorithm ensures that one sub-block participates in the matching process only once. For this purpose, a bipartite graph of sub-blocks for the query image and the target image is built with the edges indicating the distances between the corresponding blocks. A minimum cost matching is done for this graph using the adjacency matrix of the bipartite graph. In this, the distance matrix is computed as an adjacency matrix. The minimum distance d_{ij} of this matrix is found between block *i* of query and *j* of target. The distance is recorded and the row corresponding to block *i* and column corresponding to block *j*, are blocked preventing block *i* of query image and block *j* of target image from further participating in the matching process. The distances, between *i* and other sub-blocks of target image and, the distances between *j* and other sub-blocks of query image, are ignored. This process is repeated until every sub-block finds a match. The integrated minimum cost match distance between images is now defined as:

$$D_{qt} = \sum_{i=1,n} \sum_{j=1,n} d_{ij} \qquad (2.9)$$

Where, d_{ij} is the best-match distance between sub-block *i* of query image *q* and sub-block *j* of target image t and D_{qt} is the distance between images q and t.

2.3 Performance Evaluation Metrics

The most commonly used performance evaluation measures in information retrieval are precision and recall. Precision (P) measures the accuracy of the retrieval and Recall (R) measures the robustness. They are defined as:

$$Precision = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ images\ retrieved} \quad (2.10)$$

$$Recall = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ relevant\ images\ in\ the\ dataset} \quad (2.11)$$

Precision and recall values are usually represented in a precision-recall-graph, R→P(R) summarizing (R, P(R)) pairs for varying numbers of retrieved images (Deselaers, Keysers, & Ney, 2008).

Precision and recall are single-value metrics based on the whole list of images returned by the system. For systems that return a ranked sequence of images for a query (q), average precision measure is often computed which consider the order in which the returned images are presented, i.e., average precision, AP, for a single query 'q' is the mean over the precision scores after each retrieved relevant item, computed as:

$$Average\ precision, AP(q) = \frac{\sum_{n=1}^{N_R}(P(q)R_n)}{N_R} \quad (2.12)$$

where, R_n is an indicator function equal to 1 if the item at rank *n* is a relevant image, zero otherwise. In other words, it is the recall after the n^{th} relevant image is retrieved. N_R is the total number of relevant documents retrieved for the query. It should be noted that the average is over all relevant images and the relevant images not retrieved get a precision score of zero.

The mean average precision (MAP) is the mean of the average precision scores over all queries:

$$Mean\ Average\ Precision\ (MAP) = \frac{\sum_{q=1}^{|Q|} AP\ (q)}{|Q|} \quad (2.13)$$

where, Q is the set of image queries.

2.4 Benchmark Datasets for CBIR

There are many datasets available in the Web for performance evaluation of various retrieval and classification applications, out of which the Wang's dataset, COIL 100, ZuBuDu, Corel 5000 dataset etc. are the commonly used ones for CBIR evaluation purposes.

The Wang's dataset is a subset of Corel stock photo database. It consists of 1000 images of 10 different categories namely African people and village, Buildings, Beaches, Bus, Dinosaur, Elephant, flowers, Horses, Mountains and Food. Each category contains 100 images. Figure 2.2 shows the sample images from the Wang's dataset.

Figure 2.2 Sample images from Wang's Dataset

Corel 5K dataset is also a subset of the Corel stock photo database. It consists of 5000 images of 50 different categories having 100 images each.

The Coil-100 (Columbia Object Image Library) object database consists of 7200 images; 72 views of 100 objects acquired by rotating the object under study about the vertical axis. The sample images from this dataset are depicted in Figure 2.3.

Since this work mainly deals with the retrieval accuracy and response time, the Wang's dataset, Corel 5K dataset and Coil 100 dataset are used for carrying out experiments for evaluating various methods proposed in this thesis and for comparison purpose.

Figure 2.3 Sample images from Coil-100 Dataset

2.5 CBIR Approaches

2.5.1 Global Feature Based Image Retrieval

Early CBIR systems used global feature extraction methods to obtain the image descriptors. Here, features are extracted from the entire image rather than from confined regions in the image. For example, the QBIC (Flickner et al., 1995) system extracts features such as colour texture and shape features, which are obtained globally by extracting information on the means of color histograms for color features; global texture information on coarseness, contrast, and direction; and shape features about the curvature, moments invariants, circularity, and eccentricity. Other retrieval systems like the Photobook (Pentland, Picard & Sclaroff, 1996), Virage (Bach et al., 1996) etc., also use global features to represent image semantics. As these features (i.e., features that are extracted from the entire image), often fail to describe the semantic content of the image, the CBIR systems employing global features usually have low retrieval precision (Aulia, 2005). To avoid these problems and to add 'semantic knowledge' to the retrieval systems, region-based approaches, multiple feature fusion techniques, probabilistically inferring the context and the techniques of providing relevant feedback to the system are often employed.

2.5.2 Region Based Image Retrieval (RBIR)

Region based retrieval systems have been introduced to overcome the shortcomings of global feature based systems. They work on the fact that high-level semantic understanding of the image can be better reflected

by features extracted from local regions in the image rather than global features. In general, the RBIR systems partition/ segment an image into a number of regions and extract local features from each region. Later image matching algorithms are used to determine the similarity between the regions of the query and the candidate images in the database. The local region extraction in RBIR approaches can be broadly classified into two, namely, fixed block segmentation and pixel-wise segmentation. Pixel-wise segmentation schemes usually evaluate the features of the neighborhood pixels (colour, texture etc.) surrounding each pixel in the image to extract perceptually meaningful homogeneous regions in the image (Figure 2.4). Though these methods are aimed at obtaining the exact boundary of the regions, automatic image segmentation is still a challenging task and hence the risk of breaking significant objects present in the image to different parts. Moreover, the computational load is also heavier.

Figure 2.4 Example of pixel-wise image segmentation

Other systems employ fixed block segmentation approaches, where the image is divided into predefined number of blocks (Figure 2.5). Though object is not a concern for this method, computational cost is low.

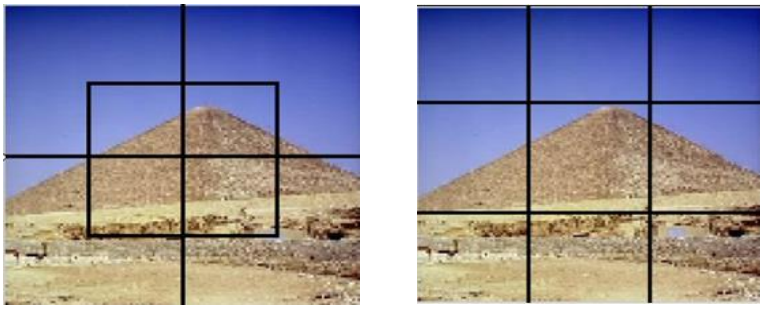


Figure 2.5 Example of fixed- block image segmentation

2.5.2.1 Pixel-wise Segmentation Approaches

The early RBIR systems like Blobworld (Carson et al., 1999) employ pixel-wise segmentation approach considering the colour, texture and position features to decompose the image into homogeneous regions called blobs. In this system, as query, the user needs to select a category of images for search, specify a blob and its importance; in response to which the system retrieves a number of images having similar blobs. In SIMPLIcity (Wang, Li & Wiederhold, 2001), the wavelet features are extracted from an image and regions are identified by k-means clustering. The similarity between images is computed by matching their respective regions using Integrated Region Matching (IRM) algorithm considering all regions in the image for similarity computation. The regions in the images are assigned

different significance values according to a chosen criterion based on which a single region is matched with more than one regions of the candidate image. The advantage of using such soft matching is the improved robustness against poor segmentation.

(Liu, Zhang & Lu, 2008) proposes a retrieval system that supports both query by keyword and query by region of interest. Here, the image is segmented into different regions and colour, texture features are extracted from them. The dominant colour of a region, computed from the quantized HSV colour histogram of the region, is taken as the colour feature and texture features are extracted using Gabor features. From these features, high-level image semantics are obtained using a decision tree-based learning algorithm named DT-ST. During retrieval, a set of images whose semantic concept matches the query is returned.

In (Su, Chen & Lien, 2010), an RBIR system with pre-clustering relevance feedback is described. The image regions are characterized with colour and texture features. Each region is assigned a weight based on the distance between the visual attention center of the image and the pixels of the respective region. Once the weights are computed, the similarity between images is determined using the IRM algorithm.

A texture feature based retrieval method is proposed in (Zhang et al., 2012). Here, the image regions are identified using JSEG segmentation algorithm. The texture features are extracted using rotation invariant curvelet transform and Gabor wavelet transform from these regions. As spectral transforms need to be applied on squared shapes for efficiency and

accuracy, the irregular shaped regions are converted to square regions by extracting the largest internal square from the region and using the pattern in the internal square to fill the blank area of the bounding box. The bounding box is then mirror padded into a square region. For retrieval, every region in the query image is used. The texture descriptor computed over rotation invariant curvelet transform is found to have superior performance than Gabor texture features for retrieval.

In (Fang et al., 2012), an image is represented with both its global and local colour information. Salient regions in the image are represented with their dominant colours, identified using the Linear Block Algorithm (LBA) (Yang et al., 2008) and colour percentage set. Image matching is performed using Quick-match algorithm, which excludes regions that are of low possibility to be in the top-matches thereby reducing the retrieval time. A relevance feedback phase is also incorporated to improve the performance of retrieval.

(Zand et al., 2015) describes a texture feature based RBIR system. The system extracts regions from the images using JSEG segmentation algorithm and the irregular regions are converted to regular shapes using an improved mirror padding method. The Gabor wavelet and curvelet filters are used for feature extraction. The sub-bands information in the texture feature vectors are then encoded using polynomial coefficients, which reduces the feature space while increasing the classification rate and discrimination power.

(Manipoonchelvi & Muneeswaran, 2015) identifies significant region in an image using a visual attention-based mechanism, considering visual metrics such as proximity, size, color contrast and nearness to image's boundaries, to locate viewer's attention. The extracted region is then represented with color layout descriptors and curvelet descriptors. The likeness between the query region and database image region is ranked according to a similarity measure (Euclidean distance) computed from the feature vectors. In another work (Manipoonchelvi & Muneeswaran, 2014), a multi- region based image retrieval system is described, in which the regions are identified based on the saliency map, location, size and homogeneity cues. The colour and texture features are obtained from these regions and represented using histogram and curvelet transform, respectively. The colour feature is represented with histogram, computed in the HSV colour space quantized into 16 bins of hue, 4 bins of saturation and 4 bins of value. The distance between the query image and the target image is computed as the minimum distance among all possible query and target region pairs using histogram intersection distance.

In (Kanimozhi & Latha, 2015), a relevance feedback method for region based image retrieval is proposed using support vector machines and firefly algorithm. Here, the images are segmented into regions using normalized cut algorithm (Shi & Malik, 2000) which is based on graph partitioning. The colour, texture features are extracted from the regions. The similarity between regions of the query and the candidate images are computed using Earth Mover's distance. The retrieved images are then presented to the user for relevance judgments and positive and negative

results are obtained as feedback. This information is then used to train a binary SVM and selected number of top scored images is then fed to a firefly algorithm for stochastic optimization. The firefly algorithm identifies the optimal feature subset and explores the solution space efficiently leading to improved retrieval performance.

2.5.2.2 Fixed Block Segmentation Approaches

Many works (Hiremath & Pujari, 2008; Hsiao et al., 2010; Fakheri et al., 2013; Takala, Ahonen & Pietikäinen, 2005) employ fixed block segmentation method to represent the image regions. In (Takala, Ahonen & Pietikäinen, 2005), two block-based texture methods are proposed employing Local Binary Pattern (LBP) as texture descriptor. The first method divides the query and database images into equally sized blocks, from which LBP histograms are extracted. Then the block histograms are compared using a relative L1 dissimilarity measure based on the Minkowski distances. The second approach uses the image division on database images and calculates a single feature histogram for the query. It sums up the database histograms according to the size of the query image and finds the best match by exploiting a sliding search window. The first method is evaluated against color correlogram and edge histogram based algorithms. In the second, user interaction dependent approach is used to provide example queries.

In (Hiremath & Pujari, 2008), the images are partitioned into non-overlapping tiles of equal size. Texture and color features are extracted from these tiles at two different resolutions in two-grid framework. Features drawn from conditional co-occurrence histograms computed by

using the image and its complement in RGB color space, serve as color and texture descriptors. Gradient vector flow fields are used to extract shape of objects and invariant moments are used to describe the shape features. For computing the similarity between image tiles, an integrated matching scheme based on most significant highest priority (MSHP) principle and adjacency matrix of a bipartite graph constructed between image tiles is used. As shape features are globally extracted, the distance between the respective feature vectors of the two images are computed using Canberra distance. The overall distance between two images is the cumulative sum of their colour-texture features and the shape features.

Hsiao et al. (Hsiao et al., 2010) proposes an interactive block based CBIR system in which the image is divided into five equally sized rectangular regions; four corner regions and one central region. Each image is represented by its local and global features, characterized by the low-frequency DCT coefficients in the YUV color space. For matching the images, two policies are provided namely local match and global match. In the local match, the user formulates a query by selecting the interested region in the image. Candidate images are then analyzed by inspecting each region in turn, to find images with the best matching region with the query region. For query images, which do not have clear objects, the user can opt for global matching.

In (Fakheri et al., 2013), the images are partitioned into non-overlapping tiles of different sizes having different weightages. The central block is the largest block and has maximum weightage while the corner

blocks are of smaller dimension and hence assigned lower weightage. The energy and standard deviation of Hartley transform coefficients of each tile, which serve as the local descriptors of texture, are extracted as sub-block features. The invariant moments of edge image are used to record the shape features. An image is characterized with the combined shape features and the sub-block features. The most similar highest priority (MSHP) (Hiremath & Pujari, 2008) principle is used for sub-block feature matching while Canberra distance is used to compute the similarity between global shape features. The retrieved image is the image, which has less MSHP and Canberra distance from the query image.

Most of the RBIR systems generally exhibit good retrieval performance. However, their response time is high due to the complexity in region matching. Hence, RBIR systems are suitable for retrieval in small-medium datasets and are not recommended for large datasets.

2.5.3 Local Feature Based Image Retrieval- Bag of Visual Words

The Bag of Visual Words (BoVW) framework is highly exploited in recent years for scalable image retrieval and is considered to be one of the most successful approaches to scene and object recognition (Yang et al., 2007; Philbin et al., 2007; Yang & Newsam, 2010; Xu et al., 2010; Wang et al., 2011). It is inspired from the bag-of words model, a popular feature representation method for document representation in information retrieval and has been first introduced in computer vision applications, image search, by Sivic and Zisserman in 2003 (Sivic & Zisserman, 2003; Sivic & Zisserman, 2009). In BoVW model, the images are characterized in

the form of global features built from local region features. In other words, images are represented as an order-less collections of local features. Figure 2.6 depicts a general BoVW framework.

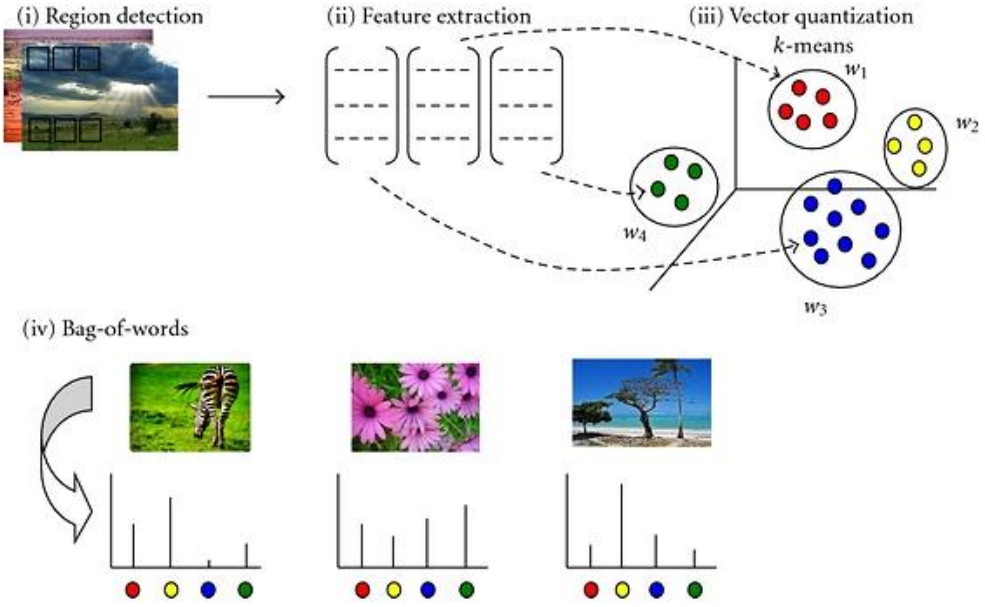


Figure 2.6 Bag of Visual Words framework (Tsai, 2012)

A typical BoVW representation includes the following steps:

i. Automatic detection of local regions/ patches or points of interest in images.

ii. Computation of local descriptors over the local regions or points of interest.

iii. Quantization of descriptors into words to form visual vocabulary/ codebook.

iv. Representation of the image in the form of histogram built over the visual vocabulary.

Detection of Local Regions and Feature Computation

Local region features are extracted from a set of training images. The local patches can be extracted in two ways: (i) on a dense grid and at various scales (to provide scale invariance), or (ii) from regions obtained from a detector, designed to provide sparse regions focusing on "interesting" parts of the image and ensuring scale invariance at the detection stage. The dense method is not feasible for large scale retrieval, as information is extracted and stored on a per-patch level and extracted patches will be too numerous for large datasets. Hence, sparse patch or points of interest extraction method is used. The patches are further represented with local descriptors such as SIFT (Lowe, 2004), SURF (Bay et al., 2006), BRIEF (Calonder et al., 2010), HOG (Dalal & Triggs, 2005) etc., of which SIFT and SURF are popularly used for image retrieval and object recognition purposes. Various methods of local region extraction and feature representation are mentioned in section 2.2.2.5 of Chapter 2.

Vocabulary Generation and Histogram Representation

To create a visual vocabulary, a large set of local features extracted from a training image corpus are clustered, i.e., given a training dataset containing N images, initially a set of visual features $D = f_1, f_2, \ldots, f_n$ are extracted from the images. These features are further quantized using clustering algorithms such as k-means, approximate k-means etc., to form a

finite number of visual words represented by W = w_1, w_2, ..., w_k, where $w_{1...k}$ represents the cluster centers.

Once the vocabulary of visual words is created, each image in the dataset is encoded to form a histogram of *k* bins, where each bin represents the frequency of occurrence of the *k* visual words in the image under consideration. The retrieval is performed by matching the histograms.

2.5.3.1 Extensions of BOVW

Despite the success of bag-of-visual-words approach, their major drawbacks include the lack of spatial information and the absence of other image features like colour, texture etc. This is because most of the local invariant descriptors are computed based on the intensity information of the images. The following sections describe various methodologies adopted to tackle these shortcomings.

2.5.3.1.1 Incorporation of Spatial Information in BoVW Framework

A number of methods have been proposed to incorporate spatial information to improve the retrieval performance of BoVW model. Lazebnik, Schmid and Ponce (Lazebnik, Schmid & Ponce, 2006) have proposed a spatial pyramid approach in which the images are characterized by concatenating the local histograms of image sub-regions from different levels of a pyramid. Here, the image is partitioned into a sequence of spatial grids at resolutions l= 0, ..., L, such that the grid at level l has 2^lx 2^lcells along each dimension. A BoVW histogram is then computed separately for each cell in the multi-resolution grid. If X and Y represent two sets of

vectors in the D-dimensional feature space, the similarity between two sets at scale l is the sum of the similarity among all corresponding sub-regions, given by:

$$K_l(X,Y) = \sum_{i=1}^{2^{2l}} I\left(X_i^l, Y_i^l\right), \quad (2.14)$$

where, X_l^i is the set of feature descriptors in the i^{th} sub-region at scale l of the image vector set X. The intersection kernel I between X_l^i and Y_l^i is formulated as:

$$I(X_i^l Y_i^l) = \sum_{i=1}^{v} min(H_{X_i^l}(j), H_{Y_i^l}(j)) \quad (2.15)$$

where, V is the total number of visual words and $H\alpha(j)$ is the number of occurrences of the j^{th} visual word, which is obtained by quantizing feature descriptors in the set α. Finally, the Spatial Pyramid Matching Kernel (SPMK) is computed as the sum of weighted similarity over the scale sequence:

$$K(XY) = \frac{1}{2^L} K_0(X,Y) + \sum_{i=1}^{L} \frac{1}{2^{L-l+1}} K_l(X,Y) \quad (2.16)$$

The weight $\frac{1}{2^{L-l+1}}$ associated with scale l is inversely proportional to the sub-region width at that scale. This weight is used to penalize the matching because it is easier to find the matches in the larger regions.

Zhang and Mayo (Zhang & Mayo, 2010) have attempted to improve the spatial information with three techniques, namely pairs frequency

histogram, shapes frequency histogram and binned log-polar representation of image features.

In (Yang & Newsam, 2011), a spatial pyramid co-occurrence method is proposed which considers the photometric and geometric aspects of an image. The co-occurrences of visual words are computed with respect to spatial predicates over a hierarchical spatial partitioning of an image. The representation captures both the absolute and relative spatial arrangements of visual words and can characterize a wide variety of spatial relationships through the choice of the underlying spatial predicates.

2.5.3.1.2 Aggregation of Multiple Image Features in BoVW Framework

Another method used to improve the retrieval effectiveness of BoVW framework is to incorporate multiple image features along with the invariant feature descriptors. The feature fusion is usually implemented in two ways, namely, early fusion and late fusion. The early fusion approach fuses multiple image features such as colour, shape etc. at the feature level so that a joint feature vocabulary is produced. In late fusion, the vocabularies are created independently for different image cues. These vocabularies are further used for building the histograms for the respective image features, which are then concatenated to represent the image.

Early fusion approach is adopted in retrieval systems like (Velmurugan & Baboo, 2011), (Fan et al., 2009), (Vigo et al., 2010), (van de Weijer & Schmid, 2006) and (Bosch, Zisserman & Muñoz, 2008). Photometrically invariant histograms are combined with SIFT for image

classification in (van de Weijer & Schmid, 2006) while in (Bosch, Zisserman & Muñoz, 2008), SIFT descriptor is computed in the HSV colour space and the features are concatenated into one combined color-shape descriptor. In (Abdel-Hakim & Farag, 2006), a coloured local invariant descriptor, CSIFT, is proposed that considers the photometric features of the image by building the SIFT descriptor in a colour invariant space. Other methods (Velmurugan & Baboo, 2011; Fan et al., 2009; Vigo et al., 2010; Chu & Smeulders, 2010; Fu et al., 2012) combine colour information with SURF descriptor with the intension of achieving better image representation.

In general, most of the object recognition techniques employ early fusion as it provides a more discriminative visual vocabulary. Nevertheless, their performance may deteriorate for image classes, which vary significantly over one of the visual cues. Late fusion performs well in such cases, as images are represented with histograms built with vocabularies constructed for individual image cues.

The late fusion approach is employed in (Yuan et al., 2011) and (Yu et al., 2013), where a combination of LBP texture feature and SIFT descriptor is used to represent the local image features. Separate histograms are constructed for the features based on the respective distinct vocabularies. The histograms are further fused to form a joint histogram to characterize the image.

Wengert, Douze and Jegou (Wengert, Douze & Jegou, 2011) have proposed a Bag of Colours method to improve BoVW, in which the colour

signature is extracted either for the whole image, producing a global descriptor, or from patches (extracted by a region detector), producing local color descriptors. For this, a colour dictionary is learnt from a large set of images in the CIE-Lab colour space, based on which histograms are computed for each image. In the latter case the color signature is applied directly on local patches and used in conjunction with SIFT descriptors to provide a rich combination of both color and texture.

In (Khan, Van de Weijer, & Vanrel, 2012), a method for recognizing object categories using multiple cues are described. The shape and colour cues are separately processed and combined by modulating the shape features through category specific colour attention. Colour is used to compute bottom-up and top-down attention maps. These color attention maps are subsequently used to modulate the weights of the shape features. In regions with higher attention, shape features are given more weight than in regions with low attention. The paper also provides an analysis of early fusion and late fusion approaches in various datasets.

Zhang et al. (Zhang et al., 2012) have considered various features, both local and holistic such as VOC, GIST and HSV, for initial retrieval and the obtained results are further combined together through graph fusion to improve the precision. VOC is a variant of vocabulary tree based retrieval, in which, up to 2,500 SIFT features are detected for each image using the VLFeat library. Holistic features are represented using GIST and HSV. For each image 960-dimensional GIST descriptor and 2000-dimensional HSV

color histogram (using $20 \times 10 \times 10$ bins for H, S, V components) are computed which are then compressed to 256 bits by applying a PCA hashing method. Retrieval is based on exhaustive search using the Hamming distance. Upon performing retrieval using various features, the obtained results are further utilized to improve the efficacy through graph fusion or graph re-ranking.

2.6 Methods to Enhance the Efficiency of Retrieval

Generally, a CBIR system retrieves content related images from the database in response to the query. However, the performance of retrieval can be limited in certain cases, as the features extracted from a single image may not be sufficient to effectively represent the image semantics. Methods such as query expansion, relevance feedback etc., involving multiple images have been employed as a solution to this.

In the relevance feedback approach, the user makes relevance judgments on the initial retrieval results, and the feedback images are used in the subsequent retrievals for query refinement to improve the performance of the system. Query refinement can be implemented through short-term learning or long-term learning. In short-term learning, only the feedbacks for the current search session are used in the learning algorithm, and image features are the primary source of data. The main challenge in this approach is to find the best combination of image features that presents the user's query. Such optimum set of feature can include features that capture similarities between positive images, or features that discriminate

positive examples from negative ones (Patil & Kokare, 2011). Short term learning approaches utilizes support vector machines (SVM), feature weighing, Bayesian learning, discriminant analysis etc. for query refinement. In long-term learning the relationship between the current and past retrieval sessions are analyzed to learn and refine the query (Lakshmi, Nema & Rakshit, 2015; Su et al., 2011). (Lakshmi, Nema & Rakshit, 2015) proposes a statistical method using axis reweighing scheme for relevance feedback to reduce the memory overhead. In (Su et al., 2011), a user navigation pattern based relevance feedback method is used that combines three query refinement techniques viz. query point movement, query expansion and query reweighting. One of the commonly used and most well known relevance feedback techniques is Rocchio's method, which was initially developed to improve document retrieval performance, defined as:

$$\vec{q}_m = \alpha \vec{q}_0 + \beta \frac{1}{|D_r|} \sum_{d_j \in D_r} \vec{d}_j - \gamma \frac{1}{|D_{nr}|} \sum_{d_j \in D_{nr}} \vec{d}_j \qquad (2.17)$$

where, $\overrightarrow{q_m}$ is the modified query, $q_0$is the original query, Dr is the set of relevant images, D_{nr} is the set of non–relevant images and α, β and γ are weights. Typical values for the weights are α = 1, β = 0.8 and γ = 0.2. It is implied that the query is refined by adding a weighted average feature vector factor of positive feedback to the initial query and subtracting the same in case of negative feedback. That is, using this formula, the original query feature vector is moved towards the relevant vectors and away from

the irrelevant ones. A survey of the recent relevance feedback methods can be found in (Li & Allinson, 2013).

Though relevance feedback methods have been employed widely to improve the retrieval performance, it may require many rounds of feedback in order to achieve satisfactory results. Hence, the learning task of relevance feedback can be a very time-consuming procedure. In addition, specifying the relevance of images is considered as a tedious and boring step for the users. Additionally, it is difficult to improve the performance if the initial retrieval itself fails. Hence, some systems employ automated query expansion method in which a number of highly ranked images from the original query are reissued as new query, providing additional relevance information about the query. It is a form of blind relevance feedback and may fail if the reissued set contains false positives.

Some other systems use only positive samples, obtained from the user either as a query set or through relevance feedback, and employs methods such as feature reweighting (Liu & Peng, 2014; Banerjee, Kundu & Maji, 2009), joint query averaging (Arandjelović & Zisserman, 2012), average query expansion (AQE) method (Chum et al., 2007) etc. to refine the query. Major challenges in such systems are the aggregation of features extracted from multiple images and the computation of similarity between the query image set and the candidate images.

Systems employing Discriminative Query Expansion (DQE) methods (Giouvanakis & Kotropoulos 2014; Torralba, Ferguz & Weiss,

2008; Chen et al., 2012; Arandjelović & Zisserman 2012b; Chatfield et al., 2015; Perdoch, Chum & Matas, 2009) uses data-dependent weight vectors, learnt from both positive and negative samples to rank the dataset images. In (Arandjelovi´c & Zisserman, 2012b; Chatfield et al., 2015) five retrieval methods employing BoVW framework on multiple query images are proposed, namely Joint-average, joint-SVM, MQ-Max, MQ-avj and Exemplar SVM. In the Joint -average method, BoVW vectors of the images in the query set (multiple images) are used to query the database and final results are obtained by ranking images based on the tf-idf score. In the joint-SVM method, a linear SVM is used to discriminatively learn a weight vector for visual words online. The query set BoVWs are used as positive training data, and BoVWs of a random set of 200 database images formed the negative training data. The weight vector is then used to efficiently rank all images in the database. In MQ-Max (Maximum of multiple queries) method, retrieval is performed for each image in the query set independently and retrieved ranked lists are combined by scoring each image by the maximum of the individual scores obtained from each query. Average of multiple queries (MQ-Avg) method is similar to MQ-Max method but the ranked lists are combined by scoring each image by the average of the individual scores obtained from each query. Exemplar SVM (MQ-ESVM) is originally used for classification purpose and trains a separate linear SVM for each positive example. The score for each image is computed as the maximal score obtained from the SVMs. In (Fernando & Tuytelaars, 2013), a query expansion method based on pattern mining, namely pattern based

query expansion (PQE) is described, which utilizes information from multiple queries to find the object of interest for retrieval. Here, consistent keypoints across multiple images (images in the query set and the top ranked retrieved results obtained by issuing images in the query set as individual query) are used to identify the object of interest.

In addition to the aforementioned methods, recently deep learning techniques are extensively employed in CBIR tasks, mainly for learning feature representations from image data (Singh, 2015; Wan et al., 2014; Gordo et al., 2016; Lin et al., 2015). Deep learning refers to a class of machine learning techniques, where many layers of information processing stages in hierarchical architectures are exploited for pattern classification and for feature or representation learning. One of the widely used deep learning architectures is Convolutional Neural Networks (CNN). A CNN is a feed-forward architecture with three main types of layers namely 2-D convolution layers, 2-D sub-sampling layers and 1-D output layers. A convolution layer consists of several adjustable 2-D filters. The output of each filter is called a feature map, because it indicates the presence of a feature at a given pixel location. A sub-sampling layer follows each convolution layer, and reduces the size of each input feature map, via mean pooling or max pooling. The 1-D layers map the extracted 2-D features to the final network output (Wei, Phung & Bouzerdoum, 2016). Designing and training CNN is a tedious and computationally intensive task. However, they are successfully implemented for various tasks like image classification (Krizhevsky, Sutskever & Hinton,

2012), scene categorization (Zhou et al., 2014), image retrieval (Gordo et al., 2016) etc. because of efficiency.

2.7 Summary

This chapter has provided an overview of a general CBIR framework describing its various phases such as query formulation, region identification, feature extraction, similarity computation, retrieval and various methods used for boosting retrieval. Different approaches such as global feature based approaches, region based approaches, local feature based approaches etc. are discussed laying groundwork for the various methods discussed in the following chapters. Region based retrieval approaches have an edge over global feature based approaches and are effectively used for retrieval in small and medium scale datasets. Most of these methods use variants of colour and texture features for characterizing the image / image regions. However, the high retrieval time arising from the region matching process make it undesirable for large-scale image retrieval. Hence, retrieval systems based on local features such as BoVW are developed to tackle this problem. They have high scalability and are successfully used for object recognition and classification purposes. Various local descriptors and combination of features used in BoVW framework are also discussed in this chapter. Description of different methods employed in CBIR systems to enhance the performance concludes the review.

Chapter

3 SALIENT SUB-BLOCK FRAMEWORK FOR SINGLE QUERY CBIR

Contents

This chapter describes a new salient sub-block method and a minimum distance algorithm for enhancing and speeding up retrieval in block based RBIR systems. Unlike the general block based approaches, which considers all the sub-blocks for image representation, the proposed method employs only selected sub-blocks based on their salience. The salience of the sub-blocks is determined by segmenting the image into fixed partitions of different configurations, finding the edge density in each partition and applying morphological operations. The colour and texture features of the salient sub-blocks are then computed from the histograms of the quantized HSV colour space and Gray Level Co- occurrence Matrix (GLCM) respectively. In addition, global image features are represented with Edge Histogram Descriptor (EHD) and HSV colour histogram. Minimum distance algorithm is introduced to match the identified sub-blocks of the query and the target images. Experimental results show that the proposed method provides better retrieval results than some of the existing methods.

3.1 Introduction

Region based approaches have been employed in CBIR systems due to their enhanced ability to retrieve query relevant images compared to global feature based retrieval schemes. A general RBIR system partitions the image into a number of regions, through fixed block segmentation or pixel wise segmentation, and extracts local features from each region. Later image matching algorithms are used to determine the similarity between the regions of the query and the candidate images in the database. Hence, the performance of retrieval highly depends on the quality of segmented regions and their feature representation. Pixel based segmentation methods, though are aimed at obtaining the exact boundary of the various object regions, may not be efficient, as automatic image segmentation is still a challenging task and are far from solved. Small areas of incorrect segmentation might make the representation very different from that of the real object, as the average features of all pixels in a segment are often used as the features of that segment. It affects the shape features also. Additionally, computational load of pixel wise segmentation method is heavier. In block-based segmentation, the image is sub divided into sub- blocks of fixed size, either of equal dimension or of unequal dimension. Though the object regions will not be segmented properly, the computational cost of segmentation is low in this approach. Also, some image retrieval systems employing block based segmentation methods perform well in practice. Hence, this work follows block-based approach for segmenting the images.

Once the regions/ blocks are identified, the similarity between images is determined by region matching. This again poses a problem and is a challenging task in RBIR systems, as a single image may consist of multiple regions and the method used for region matching affect the speed and efficacy of retrieval.

In the light of the above-mentioned issues, this work focuses on decreasing the computational overhead by devising methods to reduce the number of regions/ sub-blocks in a block based segmentation approach and by introducing a simple method for region matching, without compromising the efficiency of retrieval.

3.2 Proposed Method

The proposed method employs block based segmentation approach and characterize the images using both local and global colour-texture features. The global features are extracted from the entire image while the local features are extracted from the salient sub-blocks. This involves three steps: identification of salient sub-blocks from the segmented image, their feature extraction followed by block matching to compute the similarity. The images are then ranked according to their similarity score with the query image.

3.2.1 Salient Sub-block Selection

Here, the images are initially partitioned to grids of three different configurations; 3x3 grid, horizontal partitions and vertical partitions

(Figure 3.1). In addition, a centre block, having 1/4th the dimension of the original image is also extracted. The centre block is included, as most of the images are captured with the object as the centre of attraction.

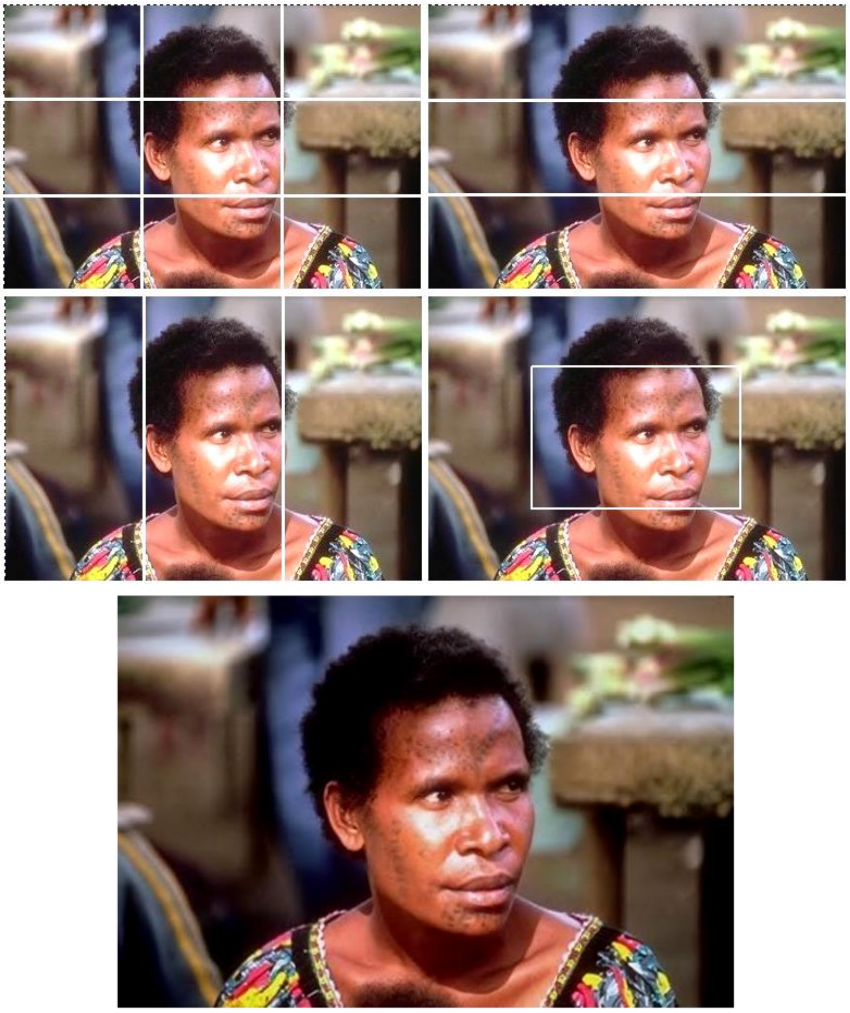


Figure 3.1 Different image configurations for feature extraction

To identify the salient blocks, initially, the grayscale image is computed and edge map is detected using Sobel edge filter with a threshold value of τ (τ<1 so that the edges are boosted). The gaps in the edge map are then bridged by dilating it with 'line' structuring element, which consists of three 'on' pixels in a row, in the 0^o, 45^o, 90^o and 135^o directions. The holes in the resultant image are then filled to get the approximate location of the object regions. The objects in the image will be identified correctly if the background is uniform.

A sub-block is selected for further processing, feature extraction and is identified as a salient sub-block if τ'% of the sub-block is part of the object region, i.e., if the number of white pixels in that sub-block is τ'% of the sub-block with maximum white pixel density, it is identified as a salient sub-block. Here, τ' is taken as 40%. For example, for the 3x3 partitioned image in Figure 3.2 (d), regions 2, 3, 5, 7, 8 and 9 are the salient sub-blocks. Salience of the horizontal and vertical blocks are also determined similarly.

The problems that may arise in this method are

i. There are images for which the edges cannot be detected; e.g. uniformly coloured images, plain images etc.

ii. The edge density/ white pixel density is confined to negligibly small regions of the image.

Figure 3.2 Salient sub-block identification. (a) Original image (b) Edge map after Sobel edge filtering (c) Edge map after edge thresholding / boosting (d) Regions of 3x3 grid (e) Horizontal regions (f) Vertical regions

In the former case no sub-block will be identified as salient as the edge density in all the sub-blocks will be zero. In the latter case a single sub-block or a very few number of blocks will be identified as salient which may affect the retrieval performance. In-order to overcome these issues, the minimum number of salient sub-blocks are fixed as five for 3x3 partition and one for both horizontal and vertical partitions. For images of case (i), sub-blocks 2, 4, 5, 6, 8 are selected for the 3x3 partition, and sub-block 2 is selected for both vertical and horizontal partitions. For case (ii), the sub-blocks are arranged in the descending order of white pixel density and the required number of blocks are selected from the top ranked ones.

The sub-block selection algorithm is depicted below.

Algorithm 3.1: Salient sub-block identification algorithm

Input: Query image, threshold factor τ'.

Output: Salient sub-blocks.

Steps

1. Find the edge map of the image (using any edge filter).
2. Apply morphological dilation with 'line' structuring element in the 0^o, 45^o, 90^o and 135^o directions and fill the holes in the edge map.
3. Partition the image into non-overlapping tiles of different configurations (3x3, horizontal, vertical)
4. For each configuration, arrange the sub-blocks in descending order of their white pixel density.

5. For each configuration, identify the blocks with maximum white pixel density. Let $W_{d_{3x3}}$, $W_{d_{3x1}}$ and $W_{d_{1x3}}$ be the respective white pixel densities corresponding to the 3x3, horizontal and vertical configurations.
6. If Wd is zero for all the configurations, i.e., white pixel density of an image= 0, choose tiles with numbers 2, 4, 5, 6 and 8 of the 3x3 grid configuration and the center tile in the vertical and horizontal configurations.
7. Otherwise, identify the sub-blocks that have white pixel density > = τ'% of W_d for each configuration and mark as salient.
8. If number of identified salient sub-blocks is less than the predefined number for each configuration, take the first 5 sub-blocks having higher white pixel density for 3x3 configuration and the highest ranked sub-block for both horizontal and vertical configurations.

Once the salient sub-blocks are identified, the image features are extracted from these sub- blocks.

3.2.2 Feature Extraction

Feature extraction is the process of creating a representation for the original data. It is most critical, as the features that characterize the image directly influence the efficacy of the retrieval. Of the many features, colour and texture are the extensively used ones for image retrieval tasks. In the proposed approach, the local colour and texture features are extracted from

the identified sub-blocks and central block in addition to the global features (HSV colour histogram and EHD texture features) of the whole image. To test the robustness of the system, the retrieval is also performed using a combined colour and texture descriptor- CEDD. The following sections describe the various features extracted from the images.

3.2.2.1 Colour

The HSV colour space, quantized into 18 bins of hue, 3 bins of saturation and 3 bins of value, is used for extracting the colour features and is represented as histogram. The colour histogram *h* for a given image is defined as a vector

$$h = \{h[1], h[2], \dots, h[i], \dots, h[N]\} \quad (3.1)$$

where, *i* represent the colour in colour histogram, h[i] represent the number of pixels of colour *i* in the image and N, the number of bins of the histogram. The normalized colour histogram is obtained as:

$$h' = \frac{h}{p} \quad (3.2)$$

where, p is the total number of pixels in the image.

The histogram of each of the three channels are extracted and is normalized in the range of [0, 1]. For every image in the dataset, both global and local colour features are extracted.

3.2.2.2 Texture

Texture measures describe the visual patterns in images and how they are spatially defined. In the proposed method, Gray Level Co-occurrence Matrix (GLCM) and Edge Histogram Descriptor (EHD) are used for representing image texture.

Gray Level Co-occurrence Matrix (GLCM)

The Gray Level Co-occurrence Matrix (GLCM) is a statistical method of examining texture that considers the spatial relationship of pixels (Haralick & Shanmugam, 1973). It is a matrix showing how often a pixel with the intensity (gray-level) value *i* occurs in a specific spatial relationship to a pixel with the value *j*. It is defined by p(i, j| d, Θ), which expresses the probability of the couple of pixels at *Θ* direction and *d* interval. Once the GLCM is created, various features can be computed from it. The proposed method computes texture features by considering d=1 and Θ = 0^o, 45^o, 90^o and 135^o. Features like contrast, energy, correlation, entropy and homogeneity are extracted for each salient block from the GLCM. Contrast is the local grey level variation in the grey level co-occurrence matrix and provides a measure of the intensity contrast between a pixel and its neighbour over the entire image (Gadkari, 2004). Contrast is zero for a constant image. Energy or Uniformity or Angular second moment measures the textural uniformity or pixel pair repetitions. It detects disorders in textures. Energy has a normalized range and can have a maximum value equal to 1. High-energy values occur when the gray level distribution has a

constant or periodic form. Entropy measures randomness/ complexity/ disorder of an image. The entropy is large when the image is not texturally uniform and many GLCM elements will have very small values. Complex textures tend to have high entropy. Entropy is strongly, but inversely correlated to energy. Homogeneity or Inverse Difference Moment measures image homogeneity and assumes larger values for smaller gray tone differences in pair elements. It is more sensitive to the presence of near diagonal elements in the GLCM and has maximum value when all elements in the image are same. GLCM contrast and homogeneity are strongly, but inversely, correlated in terms of equivalent distribution in the pixel pairs population. The correlation feature is a measure of gray tone linear dependencies in the image. It reflects how correlated is a pixel is to its neighbour over the entire image.

The various features are computed using the following formulae:

$$Contrast = \sum_{i,j} |i-j|^2 \, p(i,j) \tag{3.3}$$

$$Energy = \sum_{i,j} p(i,j)^2 \tag{3.4}$$

$$Entropy = -\sum_i \sum_j p(i,j) \, log_2 \, p(i,j) \tag{3.5}$$

$$Homogeneity = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \tag{3.6}$$

$$Correlation = \sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j} \tag{3.7}$$

Edge Histogram Descriptor (EHD)

The global texture features are represented using EHD (Park, Park & Won, 2000; Manjunath, Salembier & Sikora, 2002; Won, Park, & Park,

2002). EHD represents the local distribution of edges in each local area called a sub-image which is defined by dividing the image space into 4×4 non-overlapping blocks. Thus, the image partition always yields 16 equal-sized sub-images regardless of the size of the original image. Edges in the sub-images are categorized into 5 types: vertical, horizontal, 45^{o} diagonal, 135^{o} diagonal, and non-directional edges (Figure 3.3).

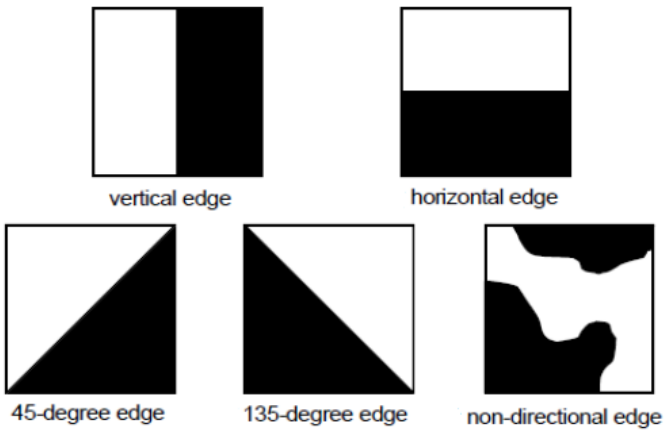


Figure 3.3 Different types of edges for EHD computation (Won, Park & Park, 2002)

The histogram for each sub-image represents the relative frequency of occurrence of these 5 types of edges in the corresponding sub-image. Since there are 16 sub-images for an image, a total of 5×16=80 histogram bins are required. Each sub-image is further divided into non-overlapping square image blocks with particular size, which depends on the image resolution. Each of the image blocks is then classified into one of the five aforementioned edge categories or as a non-edge block. A simple method to do this classification is to treat each image-block as a 2×2 super-pixel

image-block and apply appropriate oriented edge detectors to compute the corresponding edge strengths. The edge detector with maximum edge strength is then identified. If this edge strength is above a given threshold, then the corresponding edge orientation is associated with the image-block. If the maximum of the edge strengths is below the given threshold, then that block is not classified as an edge block (Won, 2004).

3.2.2.3 Colour and Edge Directivity Descriptor

The Colour and Edge Directivity Descriptor (CEDD) (Chatzichristofis & Boutalis, 2008; Kimura et al., 2011) is a 144 dimensional composite image descriptor that incorporates both colour and texture information of an image in a histogram. For computing the descriptor, initially the image is divided into a predefined number of blocks. The colour feature is represented in a 24 bin histogram in the HSV colour space by applying a set of fuzzy rules. The texture is computed by considering the YIQ colour space and applies the EHD descriptor to construct a histogram of 5 bins as explained above. The size of the CEDD is limited to 54 bytes per image, making it suitable for image retrieval tasks. The computational power required for CEDD extraction is also low, compared to the MPEG-7 descriptors.

3.3 Image Matching- Minimum Distance Method

Once the features are extracted from salient sub-blocks to represent the images, the similarity is computed by employing region-matching techniques. It plays significant role in the retrieval process as the speed and performance of the system is greatly affected by the method employed.

Here, a minimum distance method is proposed for matching various regions of images under consideration. The algorithm is described as follows:

Assume that query image I_Q has *m* salient sub-blocks represented by $I_Q = \{R^i_Q \mid i = 1, 2, \ldots, m\}$ and the target I_T has *n* salient sub-blocks represented by $I_T = \{R^j_T \mid j = 1, 2, \ldots, n\}$, where, R^i_Q and R^j_T are the i^{th} and j^{th} sub-blocks of the query and target images respectively (Figure 3.4). Every sub-block R^i_Q of I_Q is compared with all the sub-block R^j_T of I_T. The distance between R^i_Q and R^j_T, $d(F^i_Q, F^j_T)$, denoted as $d_{i,j}$, is computed as the Euclidean distance between the sub-block features $F^i_Q = \{hsv^i_Q, glcm^i_Q\}$ and $F^j_T = \{hsv^j_T, glcm^j_T\}$ of the query and the target images; hsv^i_Q, $glcm^i_Q$ and hsv^j_T, $glcm^j_T$ being their HSV colour and GLCM texture features respectively.

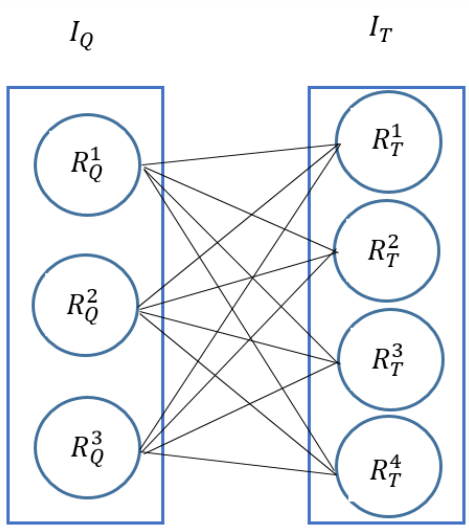


Figure 3.4 Sub-block matching for distance computation.

$d_{i,j} = d\,(F^i_Q, F^j_T)$

$$= \sqrt{\sum \left(hsv_Q^i - hsv_T^j\right)^2 + \sum \left(glcm_Q^i - glcm_T^j\right)^2} \qquad (3.8)$$

Sub-block R^i_Q, the i^{th} sub-block of I_Q, is compared with every sub-block R^j_Q (j= *1,2, ..., n*) of I_T. This results in '*n*' comparisons for a single sub-block in I_Q and *n* distance measures. The minimum distance between R^i_Q and the sub-blocks of I_T (D_i) is given by

$$D_i = \min(d_{i,1}, d_{i,2}, \ldots, d_{i,n}) \qquad (3.9)$$

Now, the minimum distance between the salient sub-blocks of the query and the target image, *D*, is computed as:

$$D = \sum_{i=1}^{m} D_i \qquad (3.10)$$

Thus, out of the $m \times n$ distances m lowest distances are added to get the distance *D*. This means that if image I_Q is compared with itself, D will be equal to zero. The minimum distance computation is summarized as follows:

Algorithm 3.2: Minimum distance computation

Input: Salient sub-block features $F_Q^{1,\ldots,m}$ and $F_T^{1,\ldots,n}$ of the query and the target images, I_Q and I_T

Output: D; minimum distance between salient sub-blocks of I_Q and I_T.

begin

D = 0;

for each sub-block in the query image I_Q, i=1 to *m*

for each sub-block in the target image I_T, j=1 to *n*

Compute distance $d_{i,j}$ using equation (3.8);

end

$D_i = \min(d_{i,1}, d_{i,2}, \ldots, d_{i,n})$;

$D = D + D_i$

end

end begin

3.3.1 Overall Similarity Computation

The overall similarity between the query image I_Q and the target image I_T, $D_{IQ,IT}$, is computed by considering the similarities of the local salient sub-blocks of all configurations, central block and the global features extracted from the images.

$$D_{IQ,IT} = D_{3x3} + D_{horizontal} + D_{vertical} + D_{central} + D_{global} \quad (3.11)$$

where, D_{3x3}, $D_{horizontal}$ and $D_{vertical}$ are the distance between the query and target image for the 3x3, horizontal and vertical grid configurations, computed using the minimum distance algorithm. $D_{central}$ is the feature similarity between the central blocks of I_Q and I_T and D_{global} is their global feature (represented using HSV colour histogram and EHD texture

descriptor) similarity. Both $D_{central}$ and D_{global} are also computed using Euclidean distance measure. $D_{IQ,IT}$ is used for filtering and ranking the resultant images.

3.4 Performance Evaluation

Experiments are carried out to analyse the following:

- Performance of the minimum distance algorithm for region based/ block based image retrieval.
- Performance of the salient sub-block method using minimum distance algorithm.

First, performance of the minimum distance algorithm is evaluated to check its effectiveness in region matching. Second, retrieval is performed using salient sub-block approach, employing minimum distance algorithm to determine region similarity.

Two publically available benchmark databases for CBIR, the Wang's image database and Coil-100 database, described in section 2.4 of Chapter 2, are used for carrying out the experiments.

Experimental Setup

The experiments are carried out in a personal computer with Intel i3 processor, 2GB RAM and 3.06 GHz clock. The program is coded in Matlab 2011b.

Evaluation Metrics

The performance of the algorithm is evaluated by computing the precision and recall measures. For each retrieval, a preselected number of images are retrieved which are illustrated and listed in the ascending order of the distance between the query and the target images. A retrieved image is considered correct if and only if it is in the same category as the query.

3.4.1 Evaluation of Minimum Distance Algorithm for Image Similarity Computation

The Wang's database is used for carrying out the experiments to analyse the performance of the minimum distance algorithm. The region matching algorithms, integrated region matching algorithm (IRM) and integrated image matching algorithm, described in section 2.2.3.2 of Chapter 2, are used for comparison purpose. The images in the database are represented with the HSV colour and GLCM texture features of the sub-blocks of the 3x3-grid configuration. All the 9 sub-blocks are taken into consideration to obtain the maximum retrieval time, which is also the worst-case scenario. Moreover, integrated image matching algorithm requires equal number of regions in the images to be compared. Since IRM similarity measure needed significance value for each sub-block for similarity computation, the white pixel density in each sub-block is considered for the same.

Table 3.1 shows the percentage average precision of the retrieved images for different categories when varying number of images are retrieved (k=20 and k=100) using the three methods.

Table 3.1 Average precision of the retrieved images using various algorithms for different number of retrieved images (k)

Image category	Average precision for varying values of k					
	k=20			k=100 (Recall)		
	Integrated image matching	Integrated region matching	Minimum distance method	Integrated image matching	Integrated region matching	Minimum distance method
Africa	62.07	69.90	68.99	41.89	**48.89**	43.08
Beaches	36.85	38.45	**41.20**	26.35	24.28	**30.17**
Buildings	38.15	48.15	**53.70**	23.99	30.81	**32.82**
Bus	72.55	80.30	**81.35**	49.81	57.21	**60.11**
Dinosaur	99.55	99.44	**99.70**	93.17	84.17	**95.61**
Elephant	51.80	61.10	57.70	31.60	34.17	33.45
Flowers	89.25	90.45	**92.35**	60.05	62.3	**64.20**
Horse	84.70	88.35	**91.60**	52.55	52.26	**60.55**
Mountains	30.70	37.27	35.30	23.66	24.83	**25.33**
Food	52.40	68.60	60.25	33.50	**42.87**	39.11
Average	**61.80**	**68.20**	**68.21**	**43.66**	**47.51**	**48.44**

*k=100 is recall, as each category in the dataset contains 100 images each.

Figure 3.5 and Figure 3.6 graphically depicts the performance of the three algorithms. It can be seen that the minimum distance algorithm performs best among the three algorithms. It has presented competent results with that of IRM and has an edge over IRM in some of the categories and outperformed integrated image matching algorithm in all the categories.

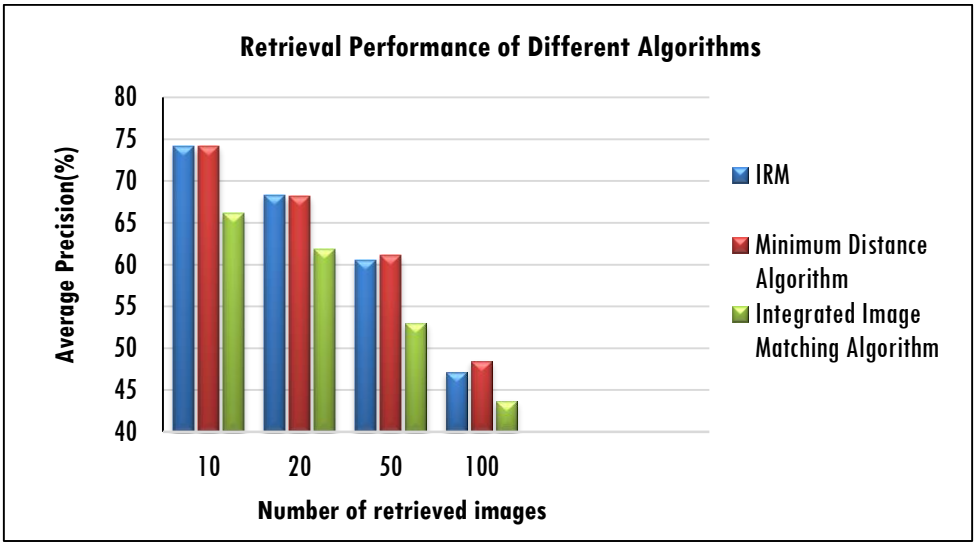


Figure 3.5 Average precision of the different region matching algorithms with varying number of retrieved images.

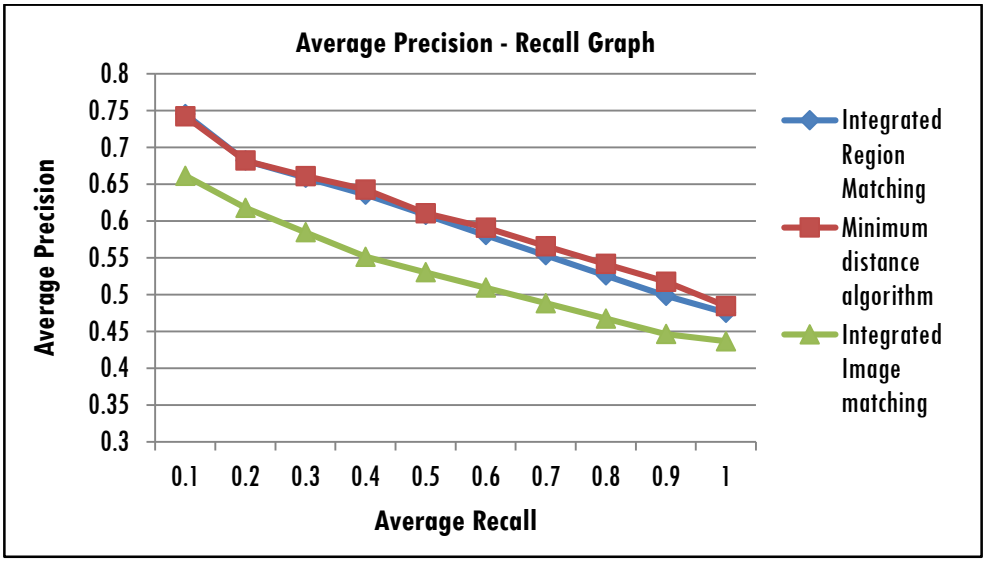


Figure 3.6 Average precision - recall graph showing the retrieval performance of different region matching algorithms.

Table 3.2 shows the average time taken by different algorithms for retrieval. Here also, the minimum distance algorithm outperforms the other two methods. It is faster than integrated image matching algorithm and takes only 1/3rd of the retrieval time of IRM algorithm.

Table 3.2 Average time (in seconds) to retrieve images using different algorithms

	Average time to retrieve images (in seconds)		
	Integrated image matching algorithm	**IRM algorithm**	**Minimum distance algorithm**
Time (s)	41.89	98.78	32.44

Figure 3.7 shows the top retrieved results of the three algorithms for a randomly selected sample query image (top left corner). The marked images are irrelevant images retrieved by the system for different algorithms. It is clear that the retrieval result of the minimum distance algorithm is better than the other two algorithms.

(a) Retrieval using Integrated Image Matching algorithm

(b) Retrieval using Integrated Region Matching algorithm

(c) Retrieval using Minimum distance algorithm

Figure 3.7 Top retrieved images using the three region matching algorithms. On top left corner is the query and the marked images are the irrelevant images.

3.4.2 Evaluation of Salient Sub-block Approach

The previous section exemplifies that the minimum distance algorithm can be used for region comparison in RBIR systems for faster retrieval. In this section, we analyse the performance of retrieval by employing the salient sub-blocks only, using minimum distance algorithm for region matching.

For performing retrieval in response to a query image, initially the salient sub-blocks of every image in the database including the query image are identified using the salient sub-block identification algorithm. The number of identified sub-blocks in each image depends on the salient regions in the image. It can be 5 to 9 in the case of 3x3 grid configuration, and 1to 3 in the case of horizontal and vertical configurations.

The experiments are carried out to evaluate the following:

- Retrieval performance of the system with salient sub-blocks
- Robustness of the proposed approach to different feature combinations
- Analysing the performance of the system in different datasets
- Performance comparison with other image retrieval systems

3.4.2.1 Retrieval Performance with Salient Sub-blocks

The initial experiments are carried out to investigate the performance of retrieval considering the salient sub-blocks of the 3x3 grid configuration only. The HSV colour and GLCM texture features extracted from these

sub-blocks are used to characterize the images. The similarity is computed by finding the distance between the sub-block features of the respective images using minimum distance algorithm (the minimum distance algorithm only is considered as it has presented better performance than the other two algorithms in terms of speed and efficiency). The retrieval results are depicted in Table 3.3.

Table 3.3 Average precision (in %) of the retrieved images for varying values of 'k' when salient sub-blocks of 3x3 configuration only is used for retrieval

Image Category	k=100	k=50	k=20	k=10
Africa	44.39	57.88	70.00	77.47
Beaches	25.76	30.44	37.30	43.50
Buildings	34.50	42.84	52.60	59.50
Bus	61.53	75.24	83.00	87.60
Dinosaur	84.17	96.60	98.85	99.50
Elephant	34.68	44.48	60.10	72.40
Flowers	59.58	76.44	89.30	94.40
Horse	59.78	78.30	91.70	95.20
Mountains	24.33	48.66	35.65	41.70
Food	42.80	53.64	65.10	70.40
Average	**47.15**	**60.45**	**68.36**	**74.17**

It can be seen that even without using all the sub-blocks of an image, comparable results can be obtained with salient sub-block approach, i.e., competent performance can be attained using a subset of the original sub-block set at a lower computational cost. Though the recall is slightly lower, the precision at lower values of *k* (number of retrieved images) is better than that of the results obtained by using all the sub-blocks. The average response time for retrieval is found to be 23.46s, 28% reduction from that of

using all the sub-blocks, confirming that the salient sub-block approach can speed up the retrieval considerably and hence can be effectively used for block based image retrieval.

On the basis of the above results, retrieval is again performed considering the salient sub-blocks of all the configurations namely, 3x3, horizontal, vertical, central block, and global features of the whole image. The global features considered here are the HSV colour histogram and EHD descriptor. The central block is also represented using the same. Table 3.4 depicts the results and Figure 3.8 shows the average precision recall graph. It can be seen that the retrieval performance can be substantially improved by extracting more information from the images. Figure 3.7 (a) and (b) shows the top retrieval results of two randomly picked query images using the proposed approach.

Table 3.4 Average precision (in %) of the retrieved images for varying values of 'k', when sub-blocks of various configurations are used for retrieval.

Image Category	k=100	k=50	k=20	k=10
Africa	46.18	59.94	71.57	76.87
Beaches	29.65	36.48	46.15	53.80
Buildings	31.83	42.64	56.20	63.10
Bus	66.49	80.30	87.40	91.70
Dinosaur	98.34	99.82	99.95	99.90
Elephant	33.57	44.00	58.45	72.50
Flowers	68.04	85.70	95.15	97.60
Horse	64.08	81.74	92.65	96.90
Mountains	25.18	50.36	35.65	40.60
Food	41.74	53.76	67.35	76.00
Average	**50.51**	**63.47**	**71.05**	**76.90**

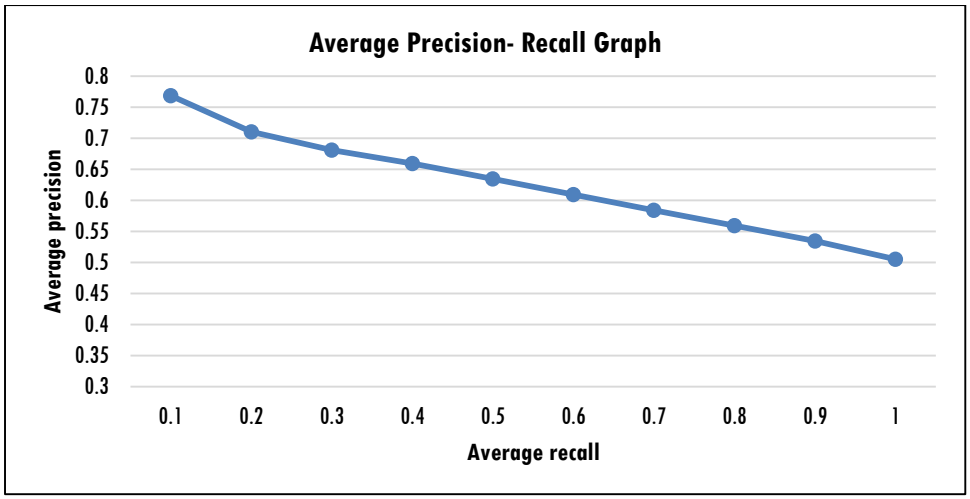
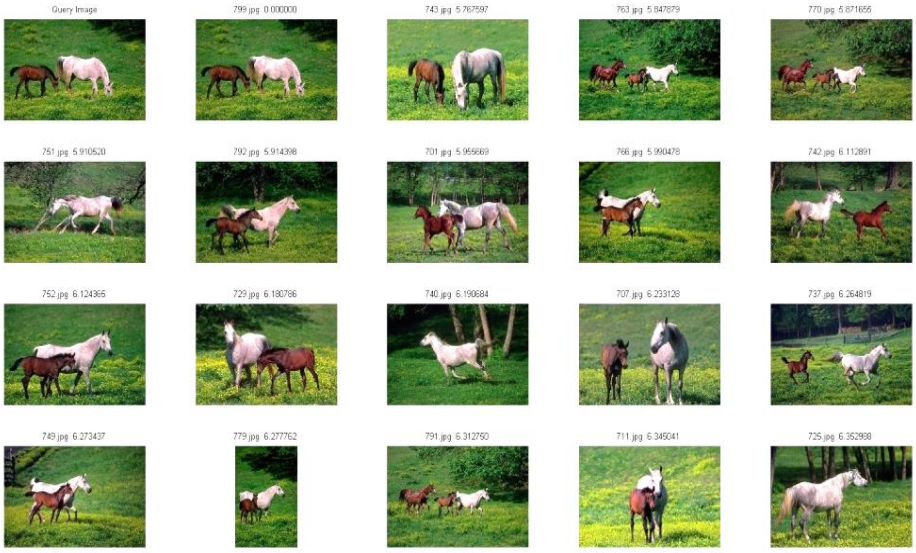


Figure 3.8 Average precision - recall graph for the retrieved images (Wang's dataset)

(a) Sample retrieved images using the proposed approach



(b) Sample retrieved images using the proposed approach

Figure 3.9 Top retrieved images in response to two sample queries from Wang's dataset using the proposed approach. The image on the top left corner depicts the query.

3.4.2.2 Retrieval Evaluation with CEDD Feature Descriptor

To analyse the robustness of the method to different image feature representations, the retrieval is performed by representing the images both locally and globally using CEDD descriptor. Here also, Euclidean distance is used for computing the feature similarity. Table 3.5 shows the resultant results. It is observed that retrieval using CEDD descriptor also provides similar results to that by using HSV, GLCM and EHD features indicating the robustness of the method. Comparing Tables 3.4 and 3.5, it can be seen

that the CEDD descriptor tends to provide better results for higher values of k and for lower values of k, the former feature combination performs well.

Table 3.5 Average precision (in %) of the retrieved results for different number of retrieved images (k=10, 20, 50 and 100)

Image Category	k=100	k=50	k=20	k=10
Africa	40.64	55.64	61.21	71.62
Beaches	34.16	41.82	50.30	55.00
Buildings	39.00	48.24	62.45	66.10
Bus	45.84	62.26	69.70	79.00
Dinosaur	93.91	99.02	99.80	99.80
Elephant	30.11	49.66	53.80	68.90
Flowers	72.70	86.72	93.40	96.50
Horse	72.33	85.38	94.55	94.30
Mountains	38.76	51.98	58.35	63.70
Food	42.07	56.88	62.60	72.50
Average	**50.95**	**63.76**	**70.61**	**76.74**

3.4.2.3 Image Retrieval in Coil1-100 Dataset

To study the performance of the salient sub-block approach in different datasets, the retrieval is carried out in the Coil 100 object dataset. All the images of the first 20 image categories are used as query for the experiment and the retrieval is performed on the entire dataset. The first 20 image categories consisting of 72 images in each, only are used because of the high retrieval time incurring due to the multiple sub-block comparison. The local and global features of the image are represented with the CEDD descriptor. The percentage average precision when varying number of images are retrieved (k),

is shown in Table 3.6. Figure 3.10 shows average precision recall graph of the retrieved results and Figure 3.11 and 3.12 depict the top 72 retrieved results in response to two random queries. It is seen that for query1 (Figure 3.11), all 72 relevant images in the dataset are retrieved and for query2 (Figure 3.12), there is only one irrelevant image in the retrieved result proving the effectiveness of the method.

Table 3.6 Average precision (in %) of the retrieved images for different number of retrieved images (k=10, 20, 50, 72) on Coil100 dataset

Image category	k=72 (Recall)*	k=50	k=20	k=10
1-20	76.12	83.04	91.97	96.74

* K=72 is recall as there are 72 images in each category.

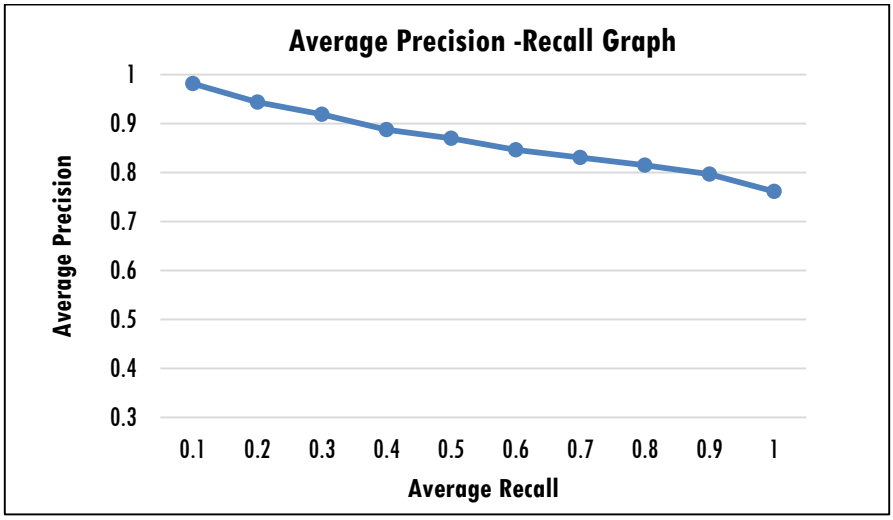


Figure 3.10 Average precision- recall graph of the retrieved images for Coil 100 dataset (first 20 categories)

Figure 3.11 Top retrieved images in response to the query from Coil100 dataset. The result shows that all the 72 relevant images in the dataset has been retrieved.

Figure 3.12 Top 72 retrieved images in response to the query from Coil100. The marked result shows the wrongly retrieved image.

3.4.3 Performance Comparison with Other Systems

The performance of the proposed salient sub-block approach is compared with other similar retrieval systems and the results are shown in Tables 3.7 and 3.8. Table 3.7 shows the percentage average precision of the top 20 retrieved images for various systems and Table 3.8 shows the percentage average recall. The Wang's database is chosen for comparison, as it is one of the most widely used benchmark dataset consisting of diverse image categories.

Table 3.7 Average precision (in %) of retrieved images using different methods (k=20)

Image Category	Average Precision of retrieved images when k=20					
	Method in (Jhanwar et al., 2004)	Method in (Hung & Dai, 2003)	Method in (Banerjee, Kundu & Maji, 2009)	CTDCIRS (Rao, Rao, & Govardhan, 2011)	Proposed method with CEDD descriptor	Proposed method with HSV colour and Texture (GLCM and EHD)
Africa	45.25	42.40	60.25	56.20	61.21	71.56
Beaches	39.75	44.55	55.23	53.60	50.30	46.15
Buildings	37.35	41.05	60.5	61.00	62.45	56.20
Bus	74.10	85.15	70.59	89.30	69.70	87.40
Dinosaur	91.45	58.65	95.00	98.40	99.80	99.95
Elephant	30.40	42.55	75.50	57.80	53.80	58.45
Flowers	85.15	89.75	80.50	89.90	93.40	95.15
Horse	56.80	58.90	90.00	78.00	94.55	92.65
Mountains	29.25	28.5	65.80	51.20	58.35	35.65
Food	36.95	42.65	55.80	69.40	62.60	67.35
Average	**52.64**	**53.24**	**70.91**	**70.48**	**70.61**	**71.05**

Table 3.8 Average recall (in %) of retrieved images using different methods (k=100)

Image Category	Average recall of retrieved images							
	Simplicity(Wang , Li & Wiederhold, 2001)	**Method in (Murala and Wu, 2014)**	**Lin, Chen & Chan, 2009**	**Hiremath & Pujari, 2008**	**Manipoonchelvi & Muneeswaran, 2014**	**Murala et al., 2013**	**Proposed method with HSV colour Texture (GLCM and EHD)**	**Proposed method with CEDD**
Africa	48	42.9	42.1	48	45.8	43.58	46.18	40.64
Beaches	32	32.6	32.1	34	39.5	35.77	29.65	34.16
Buildings	35	34.3	36.5	36	25.5	34.89	31.83	39.00
Bus	36	78.4	61.7	61	43.4	63.39	66.49	45.84
Dinosaur	95	96.2	94.1	95	99.4	92.78	98.34	93.91
Elephant	38	41.3	33.1	48	47.5	30.31	33.57	30.11
Flowers	42	66.1	75.0	61	63.0	64.59	68.04	72.70
Horse	72	42.0	47.6	74	39.6	66.55	64.08	72.33
Mountains	35	27.6	27.7	42	59.5	32.09	25.18	38.76
Food	38	36.9	49.0	50	41.8	45.12	41.74	42.07
Average	**47**	**49.9**	**49.89**	**49.9**	**50.5**	**50.91**	**50.51**	**50.95**

Table 3.7 and Table 3.8 show the average precision and recall of the retrieved images with respect to different categories for different methods. It is seen that for most of the categories the proposed method provides better or comparable results with that of the other methods. For a few categories like ‘Beaches’, ‘Mountains’ and ‘Elephant’ the performance of the proposed method is lower than that of some of the compared methods because of the similarity in the foreground and background of the images in these

categories. Figure 3.13 depicts the result of retrieval for a random query from 'Beaches' category, evidently showing the similarity in colour and texture content of the images, which are used for exemplifying the image regions. This implies that more localised colour-texture descriptors are to be used for better characterization of the image regions. However, it is evident from the Tables 3.7 and 3.8 that the aforementioned image categories are challenging for most of the retrieval approaches used for comparison.

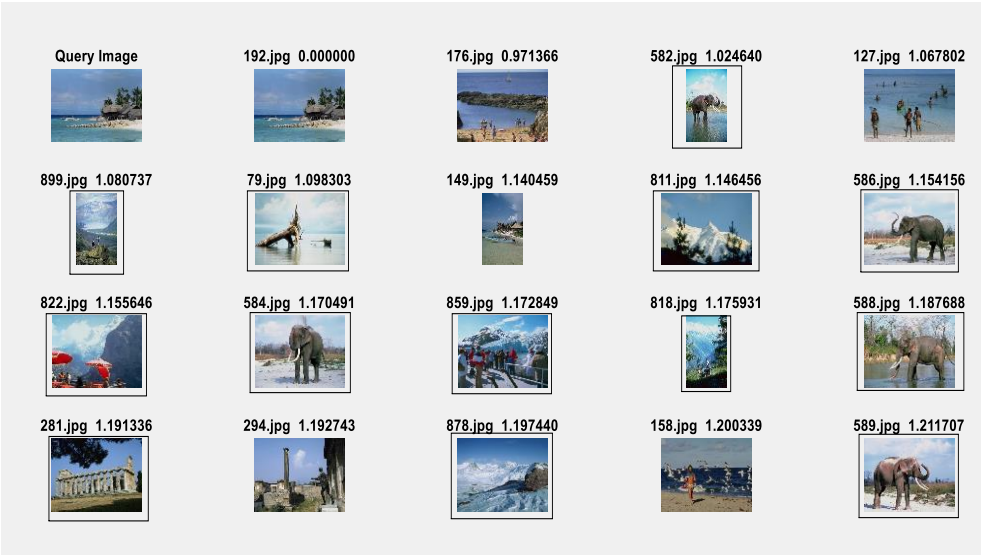


Figure 3.13 Top retrieved images in response to a sample query from 'Beaches' category of Wang's dataset. The image on the top left corner depicts the query. The marked images show the wrongly retrieved results which appear to be almost similar in colour and texture content.

For the categories 'Dinosaur' and 'Flowers' the average precision when k=20 is very high. This means that, for images with single object, the proposed algorithm works better than the compared algorithms. The retrieval performance in Coil 100 dataset, shown in Table 3.6 corroborates the same. Hence, though the method shows subservient performance for some of the image categories, its overall performance is competent with other similar retrieval approaches.

3.5 Summary

A new salient sub-block based approach and a minimum distance method for image matching are described in this chapter. Unlike other sub-block based methods, that involves all the sub-blocks of the query image to be compared with that of the candidate images, the proposed system requires only selected sub-blocks for similarity measurement, reducing the number of comparisons, computational cost and retrieval time without compromising the retrieval rate. It is observed that 28% reduction in the response time can be attained by using the salient sub-block approach than using all the sub-blocks for image matching. Evaluations using different feature sets, datasets and comparison with other retrieval systems, prove the effectiveness of the approach implying that the salient sub-blocks with minimum distance method can be effectively used for retrieval in small to medium datasets. However, it cannot be recommended for retrieval in large datasets because of the still high response time resulting from the region

matching process. Hence, the subsequent chapters explore global feature based approaches for improving retrieval efficacy.

Chapter

4 INTEGRATION OF MULTIPLE CUES IN BAG OF VISUAL WORDS FRAMEWORK

Contents

The bag-of-visual words has been applied to myriad of recognition problems in computer vision such as object recognition, scene classification and image retrieval due to its scalability and high precision. However, their performance is subservient in natural image datasets mainly due to the lack of consideration of image cues like colour, texture etc. which are not prime features while computing invariant descriptors, on which BoVW models are generally built on. Hence, this chapter describes a multi-cue fusion approach for BoVW framework, exploiting both early and late fusion methods, to improve the retrieval performance, mainly in natural image datasets. For this, a composite edge and colour descriptor is proposed to describe the local regions of the image along with the invariant feature descriptor SURF. Independent vocabularies are built based on these descriptors and images in the dataset are encoded to form two histograms using the respective vocabularies. The histograms are further fused to form a joint histogram to characterize the image. The retrieval is carried out by matching the histograms. Experimental results show that significant increment in the average precision can be attained by combining the proposed descriptor with invariant descriptors. Incorporation of spatial information further boosts the performance.

4.1 Introduction

The Bag of Visual Words framework has been highly exploited in recent years for scalable image retrieval and is considered to be one of the most successful approaches for scene and object recognition. Despite their success, the retrieval performance of BoVW based systems built with invariant descriptors such as SURF and SIFT is limited compared to region-based systems in natural image datasets. This is because most of the local invariant descriptors are computed based on the intensity information of the image and do not consider other important cues such as colour, texture etc., which are rich in natural images. Incorporation of multiple cues through feature fusion techniques such as early fusion and late fusion have been proposed to overcome this issue (Yuan et al., 2011; Yu et al., 2013; Velmurugan & Baboo, 2011; Fan et al., 2009; Vigo et al., 2010; Chu & Smeulders, 2010; Fu et al., 2012; Khan, Van De Weijer & Vanrell, 2012). A brief review of these methods are outlined in Chapter 2. In general, the early fusion approaches fuse multiple image cues at the feature level so that a joint feature vocabulary is formed, whereas in late fusion, vocabularies are created independently for different image cues. The former method is reported to be suitable for image categories with feature dependencies while the latter is preferred for categories with independent features. In this work, an attempt is made to integrate the colour and edge features of an image in the BoVW framework along with the invariant descriptors, exploiting both early and late fusion strategies, with the aim to boost the performance of retrieval.

4.2 Proposed Method

In the proposed method, early fusion approach is adopted to integrate the edge and colour information at patch level while late fusion approach is used to combine the resultant histogram with the histogram built over the keypoint descriptors. The following section describes a typical BoVW framework with multiple image cues. The terminologies and notations used in (Khan, Van De Weijer & Vanrell, 2012; Khan et al., 2013) are adopted here.

Let N be the number of images in the training set. Let a number of local features $f_{ij,}$ j = 1, 2, …, M^i, be detected for each image I^i, i = 1, 2, …, N, where, M^i is the total number of features in image *i*. These local features are represented in visual vocabularies, which describe different image cues such as color, texture, shape etc. Assume that visual vocabularies are available for different cues and be denoted by $W^k = \{w_1^k, \dots, w_{V^k}^k\}$, where w_n^k represents the n^{th} visual word from the visual vocabulary for cue k, and V^k is the total number of visual words in the vocabulary for k, i.e., n= 1, 2, …, V^k.

The quantization of the local features, $f_{ij,}$ is different for the early fusion and late fusion approaches. For example, if the image cues under consideration are colour (c), texture (t) and composite colour-texture (ct), i.e., k ∈{c, t, ct} then, for late fusion the visual words are represented as w_{ij}^c, w_{ij}^t and w_{ij}^{ct} for early fusion. Thus, $w_{ij}^k \in W^k$ is the j^{th} quantized feature of the i^{th} image for a visual cue k. For a standard single cue BoVW,

the images can be represented by the frequency distribution over these visual words as:

$$h\left(w_n^k|I^i\right) \alpha \sum_{j=1}^{M^i} \delta\left(w_{ij}^k, w_n^k\right), \quad (4.1)$$

$$\text{with } \delta(a,b) = \begin{cases} 0 \; for \; a \neq b \\ 1 \; for \; a = b \end{cases}$$

For early fusion, the histogram for the composite colour-texture feature can be represented as $h\left(w^{ct}\middle|I^i\right)$. For late fusion, the histogram are computed separately as $h\left(w^c\middle|I^i\right)$ and $h\left(w^t\middle|I^i\right)$ and are combined later.

As aforementioned, the proposed work adopts early fusion approach to integrate the image cues, colour and edge information, to form a joint feature vocabulary. A separate vocabulary is created using SURF features extracted from keypoints. These vocabularies are then used to build independent histograms for the respective features, which are further combined through late fusion to characterize the images.

Various steps involved in the characterization of images are summarized as follows:

1. Extraction of colour and edge information at patch level, their integration through early fusion, construction of a joint feature vocabulary and image histogram generation.
2. Extraction of invariant descriptors at patch level, vocabulary creation and generation of image histogram.
3. Image level integration of histograms.

4. Histogram matching and retrieval.

The feature extraction methods are described in the following sections.

4.2.1 Composite Edge and Colour Feature Extraction

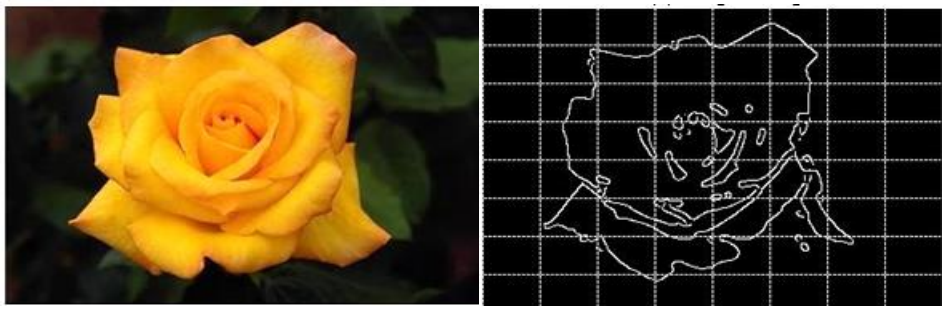
The edge and colour information are extracted by dividing the image into equally sized patches.

Edge Feature Extraction

For extracting the local edge distribution in an image, initially the edge map is detected using canny filter (Figure 4.1). The resultant edge map is then divided into *l* x *l* patches and for each patch a 1x5 edge orientation histogram is computed, in which the first four bins corresponds to the distribution of edges in the horizontal, vertical, 45^o diagonal, 135^o diagonal directions respectively and the fifth bin represents the frequency of non-directional edges. Figure 4.2 shows the filters used for identifying the edges. If $\{r_1, r_2, \ldots, r_n\}$ represents the *n* directed edges extracted from an image patch, and $E=\{e_1, e_2, ..., e_5\}$ represents the five different types of edges, then the edge distribution histogram is represented as:

$$h_{(e_i)} = \sum_{j=1}^{n} \begin{cases} 1 & if\ e_i = r_j,\ \ i=1,2,3,\ldots,5 \\ 0 & Otherwise. \end{cases} \qquad (4.2)$$

In other words, h_{e_i} represents the frequency of occurrence of directed edges in an image patch.



(a) (b)

Figure 4.1 (a) Original image. (b) The edge map of the image divided into equally sized patches

1	2	1
0	0	0
-1	-2	-1

-1	0	1
-2	0	2
-1	0	1

2	2	-1
2	-1	-1
-1	-1	-1

-1	2	2
-1	-1	2
-1	-1	-1

-1	0	1
0	0	0
1	0	-1

Figure 4.2 Filters for identifying the five types of edges

The normalized histogram is then obtained as:

$$h'_e = {}^{h_e}/_{n} \qquad (4.3)$$

where, n is the total number of edges in the image patch under consideration. Hence, for a single image patch, a 1 x 5 edge orientation histogram is computed.

Colour Feature Extraction

For extracting the colour features, the image is initially converted to CIE Lab colour space. Each of the three channels are then superimposed with the edge map so that only the edge regions in the channels are extracted, i.e., the pixels that are part of an edge only is considered. Next, for each image patch having edge, the first two colour moments, mean (μ_i) and standard deviation (σ_i), are computed for each channel, leading to a 1x6 dimensional colour feature vector. The colour moments are computed using the following formulae.

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij} \tag{4.4}$$

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N} (p_{ij} - \mu_i)^2} \tag{4.5}$$

where, N is the number of pixels contributing to edges in a patch under consideration and p_{ij} is the intensity of the j^{th} pixel in the i^{th} channel.

4.2.2 Joint Edge and Colour Feature Based BoVW

Once the colour and edge directivity features are computed, every patch in the image having edge information is represented with a joint descriptor computed by concatenating the features. Therefore, the joint edge-colour (ec) descriptor of an image patch is a 1x11 vector, in which the first five values describe the edge distribution and the next six values describe the colour information.

For constructing the vocabulary of visual words, a set of randomly selected images from the dataset are chosen to form the training set. For each image in the training set, patch level image features are extracted and a vocabulary is constructed by k-means clustering. The images in the database are then encoded using this vocabulary and characterized in the form of histograms, $h(w^{ec} \mid I^{i})$. Therefore, the histogram of the i^{th} image computed using the early fusion of edge and colour information, $h_{I^i}^{ec}$, can be represented as:

$$h_{I^i}^{ec} = h(w^{ec} \mid I^{i}), \tag{4.6}$$

where, w^{ec} is the visual words in the vocabulary of joint edge-colour descriptor.

4.2.3 SURF Based BoVW

BoVW models based on invariant descriptors have shown outstanding performance in object classification and recognition (Vedaldi et al., 2009; Arandjelovic, Zisserman, 2012; Vedaldi, & Zisserman, 2012). For this, the initial step involves identification of interesting local patches in an image either by dense sampling or by interest-point detectors. Since previous studies have shown that sampling on regular grids outperforms other methods (Chum, 2010), we have used the dense grid sampling method for local patch identification. These local patches are then represented with SURF descriptors, which are further clustered using k-means algorithm to form a vocabulary of visual words, w^{SURF}. This vocabulary is then used to

build histograms to represent the images, i.e., the histogram of the i^{th} image constructed with visual words of SURF features $h_{I^i}^{SURF}$ is:

$$h_{I^i}^{SURF} = h(w^{SURF} \mid I^i) \quad (4.7)$$

4.2.4 Joint Histogram

Once the histograms for an image are computed as aforementioned, they are concatenated through late fusion to form a cumulative / joint histogram, h_{I^i}, as shown in equation 4.8.

$$h_{I^i} = [h_{I^i}^{ec}, h_{I^i}^{SURF}] \quad (4.8)$$

The retrieval is performed by matching the image histograms.

4.3 Incorporation of Spatial Information

Spatial information is incorporated with the aim of improving the efficacy of retrieval. A variant of spatial pyramid matching method (Lazebnik, Schmid & Ponce, 2006) is adopted for this purpose. To reduce the computational overhead, images with 3 resolutions, 100%, 50% and 25%, only are taken into account. For each resolution, the image is divided into 2^k sub images, where, k=2. Descriptors are extracted from each of these sub-images and encoded with the codebook to create histograms. Hence, for a single image there will be histograms for each resolution and four additional histograms for each sub-image of the respective resolution.

4.4 Similarity Computation

All images in the database are represented in the form of histograms computed as explained in the earlier sections. The similarity between the query and the candidate images are then computed using Euclidean distance measure. If h_{I^q} and h_{I^i} represent the histograms of the query and the candidate image, then the distance, *D*, between them is calculated as:

$$D(I^q, I^i) = \sqrt{\sum_k (h_{I^q}(k) - h_{I^i}(k))^2} \quad (4.9)$$

where, k is the bins of the query and the target images.

Since edge-colour descriptor is not rotation invariant, more weightage is given to SURF descriptor while calculating similarity. Therefore, equation 4.9 is modified as:

$$D(I^q, I^i) = \alpha \sqrt{\sum_m (h_{I^q}^{ec}(m) - h_{I^i}^{ec}(m))^2} + \beta \sqrt{\sum_n (h_{I^q}^{SURF}(n) - h_{I^i}^{SURF}(n))^2} \quad (4.10)$$

Here, *m* and *n* are the bins of edge-colour feature based histogram and SURF feature based histogram respectively. α and β are the weights, with α+β =1 and α < β. The value of β is chosen to be greater than that of α, considering the robustness and invariance property of SURF descriptor. Additionally, SURF based histograms are generally built with higher number of visual words, as numerous descriptors are extracted from images, whereas edge-colour descriptors are extracted from limited number of patches of the training set images leading to smaller codebooks and hence less number of bins in the respective histogram.

Further, when spatial information is incorporated, the equation 4.10 is modified to accommodate weightage to the images with various resolutions and their respective sub-images. Hence, $D'(I^q, I^i)$, the distance between the query and the image in the dataset with spatial information is given as:

$$D'(I^q, I^i) = \sum_{k=0}^{2} \frac{1}{2^k} \left(D_k(I^q, I^i) + \frac{1}{2} \sum_{j=1}^{4} D_{kj}(I^q, I^i) \right) \quad (4.11)$$

where, k=0, 1, 2 represents the image with different resolutions, k=0 being the original image, with k=1 and k=2, the 50% and 25% resolutions of the original image.

4.5 Experimental Results and Discussions

To analyse the performance of the integrated feature approach, experiments are carried out with individual histograms and joint histograms. The codebooks for SURF based histogram and joint edge-colour feature based histogram are constructed by extracting respective features from a set of randomly selected training images in the datasets. The Wang's dataset of 1000 images, Corel5K dataset consisting of 5000 images and Coil 100 object dataset of 7200 images are used for the experiments.

4.5.1 Performance Evaluation in Various Datasets

Dataset 1 (Wang's dataset): Detailed performance evaluation is carried out using the Wang's dataset because of its compactness. To study the impact of the size of the codebooks for individual features, experiments

are conducted by varying the codebook sizes for different features. For constructing the codebook, 10% of the images in the dataset are randomly selected and both SURF and edge-colour features are extracted. For extracting the edge and colour information, different patch sizes are considered (l=8, 16, 32 and 64) and better performance was found to be for l=32. Hence, the patch size *'l'* is taken as 32 for all the experiments. The SURF and edge-colour features are then separately clustered employing k-means clustering algorithm to form the bag of visual words or codebooks.

Table 4.1 Average precision of the retrieval results for varying codebook sizes constructed using SURF descriptors

Category	100 words	200 words	300 words	500 words
Africa	54.14	55.10	53.79	55.30
Beaches	37.65	35.35	35.65	36.50
Buildings	36.15	36.40	36.80	36.85
Bus	82.50	84.40	82.80	85.10
Dinosaur	98.45	98.45	98.15	98.45
Elephant	54.55	57.15	58.70	58.20
Flowers	81.20	84.10	84.00	82.95
Horse	81.35	83.50	86.00	85.05
Mountain	32.50	35.40	35.75	34.45
Food	39.50	41.85	43.05	43.35
Average	**59.79**	**61.17**	**61.47**	**61.62**

Table 4.1 shows the percentage average precision of the retrieval results, when the top 20 images are retrieved with varying number of words in the codebook, considering the SURF features only. Codebooks of size 100, 200, 300 and 500 words are considered and images are represented with

histograms built with them. Table 4.2 shows the results of retrieval when the combined edge-colour descriptor is used as feature descriptor.

Table 4.2 Average precision of the retrieval results for varying codebook sizes, constructed using edge-colour descriptor

Category	200 words	100 words	50 words
Africa	58.79	56.97	57.12
Beaches	31.65	35.15	39.00
Buildings	43.20	43.10	40.00
Bus	74.90	77.45	76.85
Dinosaur	90.45	90.95	90.80
Elephant	33.60	35.15	37.55
Flowers	88.70	87.50	88.15
Horse	70.20	72.00	70.35
Mountain	40.50	40.05	45.30
Food	51.80	55.70	55.55
Average	**58.37**	**59.40**	**60.06**

From Tables 4.1 and 4.2, it can be seen that the histograms built with codebook size of 500 words for SURF features and 50 words for the joint edge-colour features have better retrieval performance than the other codebooks. However, in the case of SURF descriptor, histogram with 200 words also provides comparable performance with that of 500 words. Hence, for the joint histogram, codebook size of 200 for SURF and 50 for edge-colour descriptors are selected. Table 4.3 shows the percentage average precision of the retrieval results when varying number of images (k) are retrieved using the joint histograms. Here, α is taken as 0.3 and β is taken as 0.7 for similarity computation. Further improvement in precision is

achieved when spatial information is also incorporated as shown in Table 4.4.

Table 4.3 Average precision of the retrieval results when images are represented with the joint histograms

Category	**k=100**	**k=50**	**k=20**	**k=10**
Africa	41.96	54.77	66.26	71.52
Beaches	30.18	37.10	46.45	53.60
Buildings	23.81	30.54	41.95	51.30
Bus	69.06	86.24	92.80	94.20
Dinosaur	92.16	99.48	99.95	100.00
Elephant	29.10	38.44	56.40	70.10
Flowers	76.25	89.78	94.85	96.70
Horse	54.44	74.26	87.65	95.30
Mountain	28.11	34.34	43.70	51.20
Food	39.54	50.86	62.00	68.60
Average	**48.46**	**59.58**	**69.20**	**75.25**

Table 4.4 Average precision of the results, when images are represented with the combined histograms and spatial information

Category	**k=100**	**k=50**	**k=20**	**k=10**
Africa	43.25	56.59	67.53	74.34
Beaches	28.95	36.24	44.85	52.40
Buildings	23.06	31.16	43.15	51.70
Bus	69.28	85.58	92.55	94.80
Dinosaur	96.21	99.94	100.00	100.00
Elephant	33.69	45.68	64.35	78.70
Flowers	78.70	92.84	96.90	97.70
Horse	54.63	75.18	90.20	96.80
Mountain	28.10	35.74	46.45	56.60
Food	39.31	48.84	59.25	66.70
Average	**49.52**	**60.78**	**70.52**	**76.97**

As seen from Table 4.1 to Table 4.4, 8.1% increment in the average precision can be achieved by the proposed method by combining the histograms of the descriptors. Also, an additional 1% increment in retrieval precision is observed when spatial information is incorporated. It should be noted that the proposed feature combination and fusion method boosts the performance of BoVW framework, comparable to that of the RBIR systems.

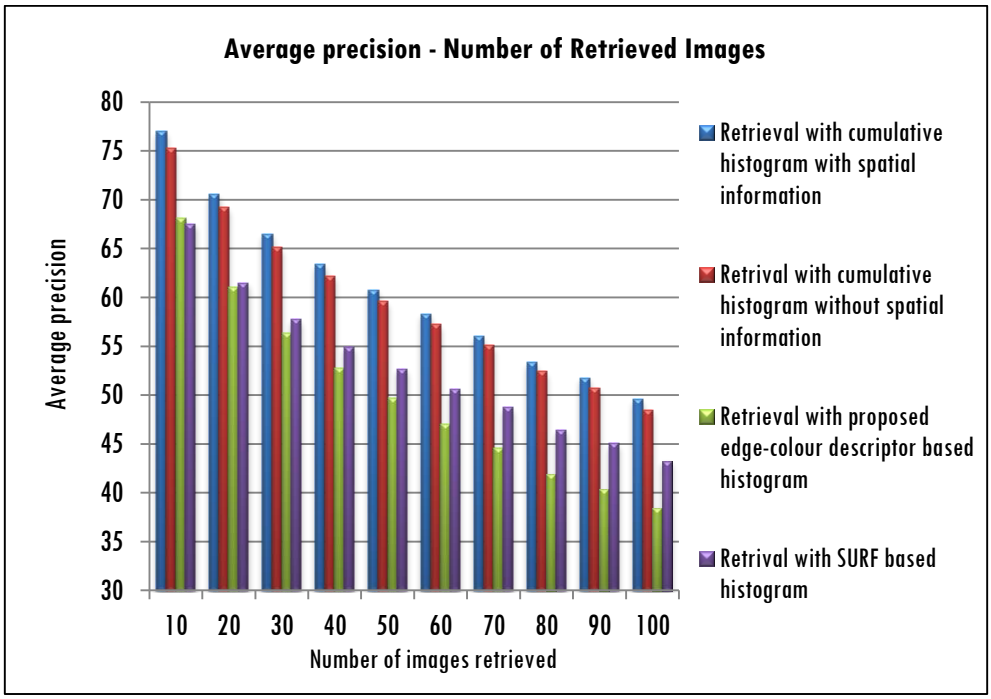


Figure 4.3 Average precision of the retrieved results (using Wang's dataset) with varying number of retrieved images for various descriptors and with spatial information

Figure 4.3 depicts the percentage average precision with varying number of retrieved images for different descriptor combinations and spatial

information. It can be seen that the proposed edge-colour descriptor has high precision competent with that of SURF descriptor with limited number of words (50 words to 200 words of SURF) when small number of images are retrieved. This is a desirable property as large datasets are more concerned about precision rather than recall. Figure 4.4 shows the precision recall graph for different combination of descriptors and with spatial information.

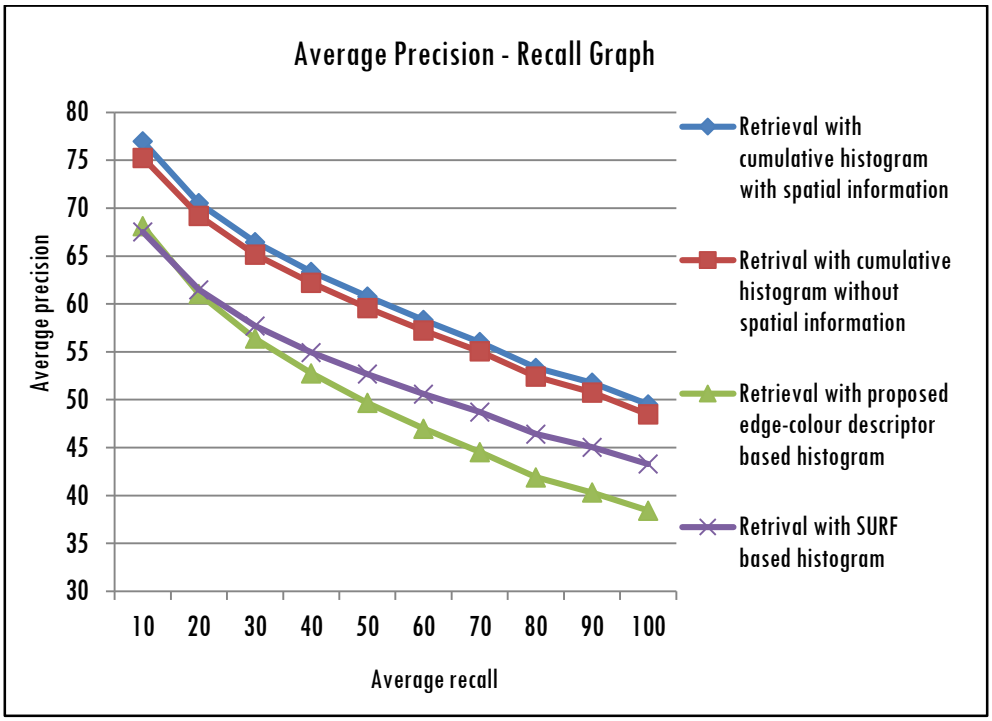
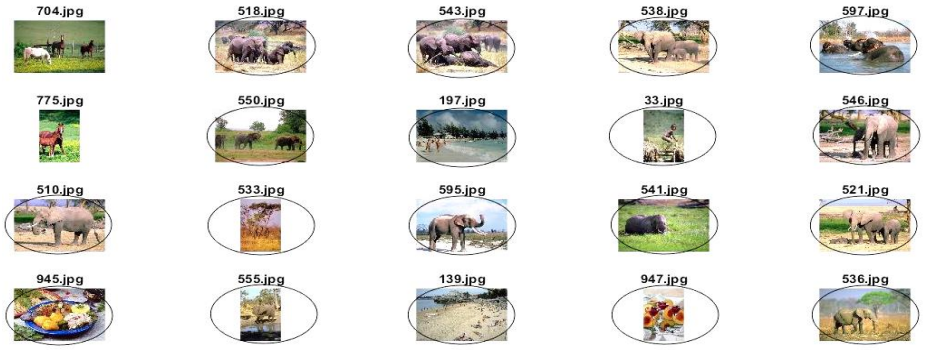
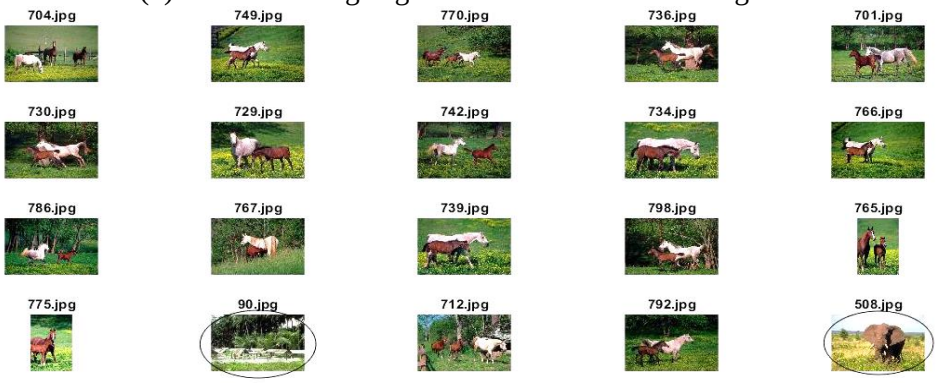


Figure 4.4 Average precision – recall graph for the various combination of descriptors and with spatial information.



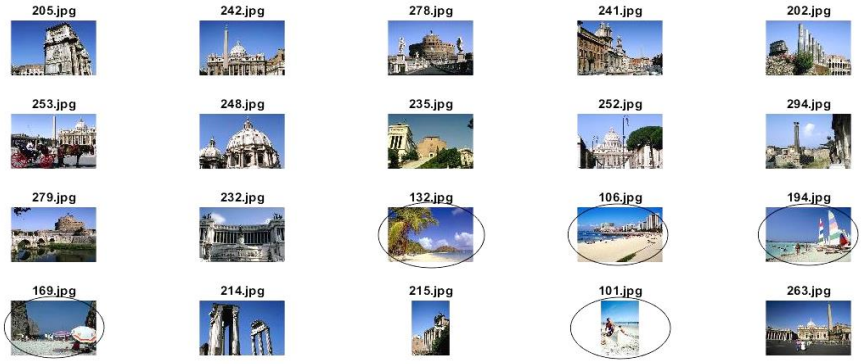
(a) Retrieval using SURF feature based histogram

(b) Retrieval using edge-colour feature based histogram

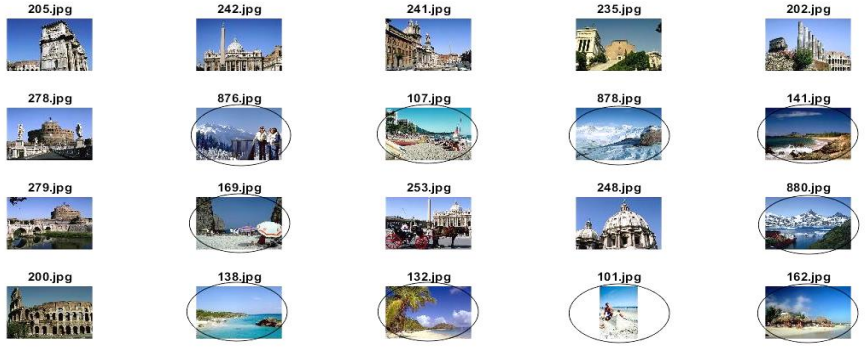


(c) Retrieval using cumulative feature (edge-colour and SURF) based histogram

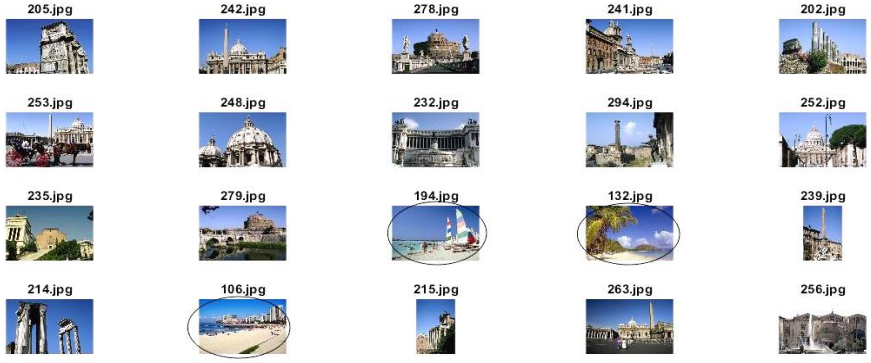
Figure 4.5 Retrieval results using individual and combined features. On top left corner is the query and marked images denote the false positives.



(a) Retrieval using SURF feature based histogram



(b) Retrieval using edge-colour feature based histogram



(c) Retrieval using cumulative features based histogram

Figure 4.6 Retrieval results using individual and combined features. On top left corner is the query and the marked images denote false positives.

Figure 4.5 and Figure 4.6 show the results of retrieval when two sample images are presented as query. It can be seen that the fusion of the features considerably improves the performance of retrieval.

Table 4.5 Performance comparison with other retrieval systems using Wang's dataset (percentage average precision when top 20 images are retrieved)

Category	SIFT based BoVW (Jurie & Triggs, 2005)	SIFTbased BoVW with SPM.(Yang et al., 2009)	Patch based SIFT-LBP (Yu et al., 2013)	Image based SIFT-LBP(Yu et al., 2013)	Proposed method without spatial information	Proposed method with spatial information
Africa	55	61	54	57	66.26	67.53
Beaches	47	49	39	58	46.45	44.85
Buildings	44	46	45	43	41.95	43.15
Bus	93	93	80	93	92.80	92.55
Dinosaur	98	99	93	98	99.95	100.00
Elephant	52	58	30	58	56.40	64.35
Flowers	77	83	79	83	94.85	96.90
Horse	65	65	54	68	87.65	90.20
Mountain	34	36	35	46	43.70	46.45
Food	52	51	52	53	62.00	59.25
Average	**61.7**	**64.1**	**56.1**	**65.7**	**69.20**	**70.52**

The performance of the proposed feature fusion method is compared with some of the recent systems employing different features, combination of features, and various methods reported to improve retrieval. In (Jurie & Triggs, 2005) and (Yang et al., 2009), SIFT feature based histograms are

used to represent the images. (Yang et al., 2009) employs linear SPM kernel based on SIFT sparse codes, an extension of the SPM method, implemented by generalizing vector quantization to sparse coding followed by multi-scale spatial max pooling. Two feature integration schemes are described in (Yu et al., 2013) using SIFT and LBP features, at the patch level and image level. The results of various methods (percentage average precision), when the top 20 images are retrieved, are presented in Table 4.5. It is observed that the proposed method outperforms the other systems under consideration even without spatial information.

The retrieval performance of the system in terms of recall is also compared with other methods employing feature fusion approaches and the results are shown in Table 4.6. In (Walia & Pal, 2014), a modified Color Difference Histogram (CDH) and Angular Radial Transform (ART) features are exploited to capture color, texture and shape information of an image. Retrieval is performed using the individual descriptors and the results are fused by means of three techniques namely min-max normalization, z-score normalization and Borda Count (a post-classification fusion technique). (Murala, Maheshwari & Balasubramanian, 2012) has combined LBP features with the edge information. In (Murala & Wu, 2014), a robust LBP feature, RLBP, is proposed, which combines robust sign LBP operator (RS_LBP), a generalized LBP operator, and robust magnitude LBP operator (RM_LBP). It is observed that the proposed method has an edge over various other methods under consideration.

Table 4.6 Performance comparison with other retrieval systems using Wang's dataset (percentage recall)

Category	Min-max fusion (Walia & Pal, 2014)	Z-score normalization (Walia & Pal, 2014)	Borda count fusion (Walia & Pal, 2014)	Method in (Murala, Maheshwari & Balasubramanian, 2012)	Method in (Murala & Wu, 2014)	Proposed method with spatial information
Africa	10.20	9.80	10.00	39.7	43.20	43.25
Beaches	18.00	18.00	17.60	37.3	35.10	28.95
Buildings	11.60	11.80	11.20	34.9	34.70	23.06
Bus	15.60	16.00	14.60	74.1	81.60	69.28
Dinosaur	20.00	20.00	20.00	88	88.40	96.21
Elephant	16.80	17.00	16.80	29.0	39.90	33.69
Flowers	20.00	19.80	19.80	70.8	65.20	78.70
Horse	20.00	20.00	20.00	41.7	40.80	54.63
Mountain	16.80	16.40	15.40	29.0	28.00	28.10
Food	7.60	7.20	7.40	47.0	38.50	39.31
Average	**15.66**	**15.60**	**15.28**	**49.16**	**49.50**	**49.52**

Dataset 2 (Corel 5K dataset): For the evaluation in the Corel 5K dataset also, the feature extraction and histogram creation are done similar to that of Wang's dataset, i.e., fusion of SURF based histograms of 200 bins and edge-colour based histogram of 50 bins are used. Leave-one-out method to query all 5,000 images, i.e., querying every image with the remaining 4,999 images in the database images is used. The performance is evaluated by calculating the percentage retrieval precision of the top *k* retrieved images averaged over the 5,000 queries. The results are shown in Table 4.7. Here also, better performance can be attained through combining the features.

Table 4.7 Average precision of the Corel 5K dataset for different features

Features used	Average precision for different number of top retrieved images (k)				
	k=1	**k=2**	**k=3**	**k=5**	**k=10**
SURF based histogram built with 200 words	57.94	43.17	35.57	27.65	22.02
Joint edge- colour based histogram built with 50 words	61.32	46.65	38.96	30.71	24.57
Cumulative histogram with 250 bins	64.27	50.75	43.14	34.68	28.38

Table 4.8 provides a comparative study of the proposed method with various features and fusion methodologies used in (Zhang et al., 2012a and Zhang et al., 2015b) on the Corel-5K dataset. In their work, both local and holistic features such as VOC, GIST and HSV colour features are considered for initial retrieval and the obtained results are further combined together through graph fusion to improve the precision. It can be seen that the proposed approach outperforms various feature combinations and boosting methods under consideration even with compact size and without spatial information.

Table 4.8 The top-1 average precision (in %) of the Corel-5K dataset for different methods

Method	***VOC***	***GIST***	***VOC graph***	***GIST graph***	***SVM fusion***	***Graph PageRank***	***Graph density***	***Three features*** *(VOC, GIST, HSV)*	**Joint edge-colour based histogram built with 50 words**	**Cumulative histogram with 250 bins**
Average precision	46.66	46.16	51.50	50.72	51.34	51.76	54.62	62	61.32	64.27

Dataset 3 (Coil 100 dataset): Coil 100 dataset is used for analyzing the performance of the proposed method for object retrieval. For building

the codebook, features are extracted from 10% of the randomly selected images. Similar to the other two datasets, here also the images are represented with 250 bins histograms; of which 200 bins are of SURF features and 50 bins are of edge-colour features. All the 7200 images in the dataset are issued as query and the average precision of the results, when varying number of images (k) are retrieved, are shown in Table 4.9. It is observed that the proposed feature combination and integration method can provide better retrieval results for object datasets also.

Table 4.9 Average precision of the retrieved images for different number of retrieved images (k=10, 20, 50, 72) on Coil100 dataset for different feature combinations

Features used	k=72 (Recall)	k=50	k=20	k=10
SURF feature based histogram built with 200 words	50.98	51.74	62.82	86.84
Joint edge-colour histogram built with 50 words	57.15	57.90	69.93	89.19
Cumulative histogram with 250 bins	63.74	64.46	76.44	94.22

4.5.2 Response Time for Retrieval in Various Datasets

Table 4.10 shows the average time required by the proposed system when retrieval is carried out in datasets of varying sizes- Wang's 1K dataset, Corel 5K dataset and Coil 100 dataset. The coding is done in Matlab 2014b and is run on a personal computer having Intel core i3 processor, 2GB RAM and 3.06GHz clock. It is seen that the response time of retrieval is lower compared to RBIR systems. Analysis of the computational time required for retrieval in various datasets revealed that, on an average, 80% of the time is

spent on calculating the similarity between images using Euclidean distance. Hence, the response time can be reduced by employing other distance measures such as histogram intersection distance, city-block distance etc.; but with slight reduction in the retrieval precision.

Table 4.10 Average response time for retrieval in datasets of different sizes

Features used	Average time required to retrieve images (s)		
	Wang's DB (1K)	Corel DB (5K)	Coil DB (7.2K)
SURF feature (200 bins histogram)	0.17	0.91	1.68
Edge-Color feature (50 bins histogram)	0.15	0.88	1.61
Cumulative histogram with 250 bins	0.315	1.57	2.67

4.6 Summary

In this chapter, a feature integration scheme, employing both early and late fusion strategies, is proposed to improve the retrieval performance of BoVW based image retrieval systems. A combined edge-colour descriptor extracted from image patches is introduced, which when fused with SURF descriptors through late fusion is observed to be providing an increment of 8.1%, 6.9% and 11% in average precisions of the top retrieved results of Wang's dataset, Corel 5K dataset and COIL100 dataset respectively, indicating the effectiveness of the method. Additionally, as the BoVW framework characterize the images with global features computed over local features, the response time for retrieval is also low compared to RBIR systems.

Chapter

5 FEATURE REPLACEMENT BASED MULTIPLE QUERY CBIR

Contents

Multiple queries are often used in CBIR systems to reduce the semantic gap problem by gathering additional information about the users' requirement and thereby to improve the retrieval efficiency. One major challenge here is the determination of feature similarity between the query set and the candidate images. This chapter introduces a feature replacement algorithm for this purpose, which computes the relevance of the candidate images to the query set from the cumulative displacements of the centroid caused by the replacement of elements of query set with the candidate images. Experimental results show that using the proposed algorithm, the average precision of the top retrieved results can be increased by 10% by having an additional image with the query, and can continue to provide improved precision with every additional image added to the query image set. Also, the system is found to exhibit superior performance compared to other multi-query systems employing query averaging, feature re-weighting and supervised learning methods for boosting retrieval.

5.1 Introduction

Multiple queries are often employed in image retrieval systems with the objective of enhancing the performance of retrieval, as they provide a more expressive formulation of the user's need than the information gathered from a single query. In a typical multiple query retrieval system, the features extracted from the queries are utilised, either for reforming the query to a better representation or to learn a model using supervised learning algorithms. An overview of various such methods are outlined in section 2.6 of Chapter 2. Despite the methodologies used to reform the query and various learnt models, ultimately the relevant images are retrieved by computing the feature similarities between the images. Hence, in this chapter a feature replacement algorithm is described that utilizes the features extracted from the images in the query set for similarity computation without any query refinement or model learning. The similarity is computed by considering the displacements of the centroid of the query set caused by the replacement of elements in it with an element from the target image set. This algorithm can also be effectively used with CBIR systems employing relevance feedback, as significant improvement in performance can be achieved with minimal user feedback.

5.2 Image Representation

Since the focus of the work is the computation of similarity between the query-set and the candidate images in the dataset, the images are represented with global features rather than local features. Here, various

features comprising of colour and texture are extracted from the images in the same way as described in section 3.2.2 of Chapter 3. The similarity is computed by matching the features. The following features are used for the holistic image representation.

- CEDD descriptor.
- HSV colour and EHD texture.
- HSV colour and GLCM texture.

The images are represented with any one of the above features for similarity computation and further retrieval.

5.3 Feature Replacement Algorithm

The proposed feature replacement algorithm is used for computing the similarity between the query image set and the target images in the database in a multi-query environment. The algorithm is based on the principle that if an element in set X is to be replaced with an element in set Y, it will cause minimum information change if the replaced element has high similarity with the element being replaced. Here, the user provides N_q positive images as the query image set. Positive images only are considered, as the aim of the system is to get improved performance with minimum number of query images and minimum effort from the user. Feature vectors are extracted from each element in the query image set and every image in the database. The centroid, C_q, of the query image set is then obtained by computing the mean of their feature vectors. Now, an element

x_i of the query set is replaced with an element y_j from the candidate image data set and the new centroid C_i' is computed (Figure 5.1). The displacement of C_i' from C_q is computed and this process is repeated by replacing other elements in the query set with the same element y_j from the candidate image database. The cumulative sum of these resultant displacements is computed, and is considered as the similarity of the candidate image with the query image set. The distance between the new centroids and C_q can be computed using any distance measure. Here, we have used Euclidean distance for this purpose.

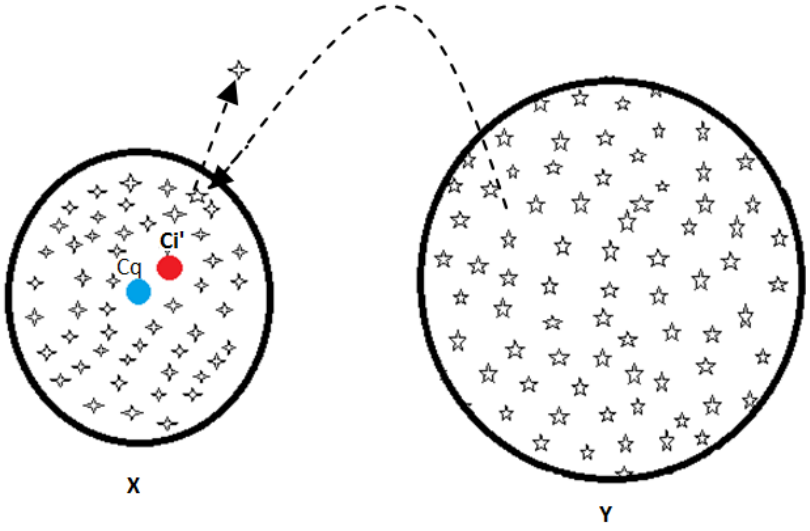


Figure 5.1 When an element in set X (query image set) is replaced with an element in set Y (candidate image set), the centroid shifts from Cq to Ci'. The algorithm computes the cumulative shifts caused by the replacement of every element in *X* with the same element from *Y*.

The algorithm is summarized as follows:

Let $X = \{x_1, x_2, \ldots, x_{N_q}\}$ be the set of feature vectors of the query images provided by the user, Nq being the number of images in the query set. The sum of feature vectors, Q_{sum} and centroid, C_q of the query images are computed using the following formulae:

$$Q_{sum} = \sum_{i=1}^{N_q} x_i \tag{5.1}$$

$$C_q = \frac{1}{N} Q_{sum} \tag{5.2}$$

where, x_i is the feature vector of the i^{th} image in the query image set. Let $Y = \{y_1, y_2, \ldots, y_M\}$ be the set of feature vectors of the candidate images in the database, M being their total number. For $y_i \in Y$, replace the feature vector of $x_i \in X$ one by one with y_j and calculate the new centroids C_1,' C_2,' …, C_{Nq}' for every replacement and compute the displacement D_i, the distance between C_q and C_i'.

$$C_i^{'} = \frac{1}{N_q}\left(Q_{sum} - x_i + y_j\right) \tag{5.3}$$

$$D_i = \sqrt{\left(Cq - C_i^{'}\right)^2} \tag{5.4}$$

The similarity of the j^{th} candidate image with the query image set, S_j is now computed as:

$$S_j = \sum_{i=1}^{N_q} D_i \quad (5.5)$$

The candidate images are now ranked according to these similarity measures. The algorithm is described as follows:

Algorithm 5.1. Feature replacement algorithm

Input:

1. The N_q query images with their feature vectors; X= {x_1, x_2, ..., x_{N_q}}.
2. M candidate images in the database with their feature vectors; Y={y_1, y_2, ..., y_M}.

Output:

Similarity measures, $S_{1...M}$; the cumulative distances between C_q and the N_q number of C_i's.

Begin

1. Find the centroid, C_q, of the query image set using equation 5.2.
2. **for** *j=1* to *M*
3. D=0;
4. **for** *i=1* to N_q
5. Replace $x_i \in X$ with $y_j \in Y$ so that X'= {y_j, x_2, …, x_{N_q}}.
6. $C_i' = \frac{1}{N_q}(Q_{sum} - x_i + y_j)$
7. $D = D + \sqrt{(C_q - C_i')^2}$
8. **end for**
9. $S_j = D$
10. **end for**

End

5.4 Experimental Results

To analyse the performance of the algorithm, the following criteria are evaluated:

- The performance of the feature replacement algorithm with varying number of images in the query set.
- The robustness of the algorithm to different feature representations.
- The response time of the algorithm to datasets of different sizes.

The experiments are conducted on the Wang's image database and Coil-100 database. The program is coded in Matlab 2014b and is run on a personal computer with Intel i3 processor having 2GB RAM and 3.06 GHz clock. The precision and recall measures are used as evaluation metrics.

5.4.1 Response of The Algorithm to Number of Images in The Query Set.

To analyse the sensitivity of the algorithm to the number of images in the query image set, Nq is given different values (Nq=1, 2, 3, 5, 10, 12, 15) and retrieval efficiency is computed for every image in the database. The CEDD feature descriptor is used for image representation and the sample images are randomly picked from each category according to the number of images (Nq) required for the query image set. The performance of the feature replacement algorithm on two different datasets are given below.

Dataset1 (Wang's dataset): To analyse the performance of the method, retrieval is performed on the entire dataset by varying the number of images in the query set. The query set is formed in such a way that every image in the dataset becomes a member in it at least once, i.e., each image in the dataset is issued as a query at least once by choosing the respective image and by randomly picking other images according to the required Nq. Table 5.1 and Table 5.2 show the percentage average precision of the retrieval results when '*k*' number of images are retrieved, with varying number of images in the query set (Nq). Table 5.1 depicts the results of retrieval when there is only one image in the query image set (N_q =1) and Table 5.2 shows the retrieval performance when N_q assumes different values. It is seen that by adding one more image to the query set, the percentage average precision is improved by 8.6% on an average, and by 10% on the top retrieved results (i.e., for lower values of '*k*'). Also, the system acquires better precision with each additional image added to the query image set. For lower values of Nq, there is significant improvement in the average precision compared to larger numbers of Nq, which is a desirable property, as improved precision can be achieved with lesser number of sample query images. Figure 5.2 graphically depicts the same. Figure 5.3. shows the average precision-recall graph for the retrieved images with varying number of images in the query image set, Nq. From the graphs, it can be confirmed that the feature replacement algorithm can be effectively used for boosting the efficiency of CBIR systems.

Table 5.1 Average precision of the results for different number of retrieved images (k), with single image in the query set, Nq=1.

Category	Number of retrieved images (k)									
	100	**90**	**80**	**70**	**60**	**50**	**40**	**30**	**20**	**10**
Africa	42.07	43.98	46.02	48.12	50.62	52.79	55.23	58.62	62.32	69.29
Beaches	38.51	39.77	40.95	42.47	44.80	46.54	48.18	50.83	54.45	59.80
Buildings	35.59	36.89	38.49	40.39	42.73	45.08	48.80	52.47	57.95	65.50
Bus	47.90	50.22	52.94	55.94	59.45	62.96	66.83	70.70	74.55	81.00
Dinosaur	87.74	90.67	92.86	95.04	96.27	97.54	98.75	99.43	99.85	100.0
Elephant	26.52	27.57	28.64	30.10	31.85	34.22	36.90	41.03	46.85	58.50
Flowers	48.92	51.74	54.88	59.03	62.68	66.74	70.83	76.17	82.20	89.90
Horse	74.01	77.24	80.06	82.61	85.23	87.36	88.95	90.27	91.30	94.40
Mountain	38.65	39.58	40.51	41.60	43.20	77.30	47.13	49.40	52.70	60.30
Food	41.63	42.79	44.00	45.77	47.53	49.72	51.98	55.30	59.20	66.30
Average	**48.1**	**50.0**	**51.9**	**54.1**	**56.4**	**58.7**	**61.3**	**64.4**	**68.1**	**74.4**

Table 5.2 Average precision of retrieved images for different values of k when Nq=1, 2, 3,5,10, 12 and 15

No. of images in the query image set (Nq)	Average precision for different values of k									
	100	**90**	**80**	**70**	**60**	**50**	**40**	**30**	**20**	**10**
1	48.15	50.04	51.93	54.10	56.43	58.76	61.35	64.42	68.13	74.49
2	55.21	57.54	59.28	62.42	65.07	67.76	70.56	73.74	77.86	84.51
3	58.42	60.98	63.44	65.97	68.69	71.46	74.17	75.88	81.88	88.17
5	61.85	64.54	67.13	69.62	72.41	75.23	78.07	81.29	85.37	90.98
10	64.74	67.44	69.92	72.44	75.16	77.82	80.52	83.52	87.47	91.55
12	65.43	68.20	70.80	73.35	76.21	78.91	81.48	84.11	87.73	91.63
15	65.54	68.37	70.90	73.49	76.26	79.03	81.52	84.32	87.80	91.82

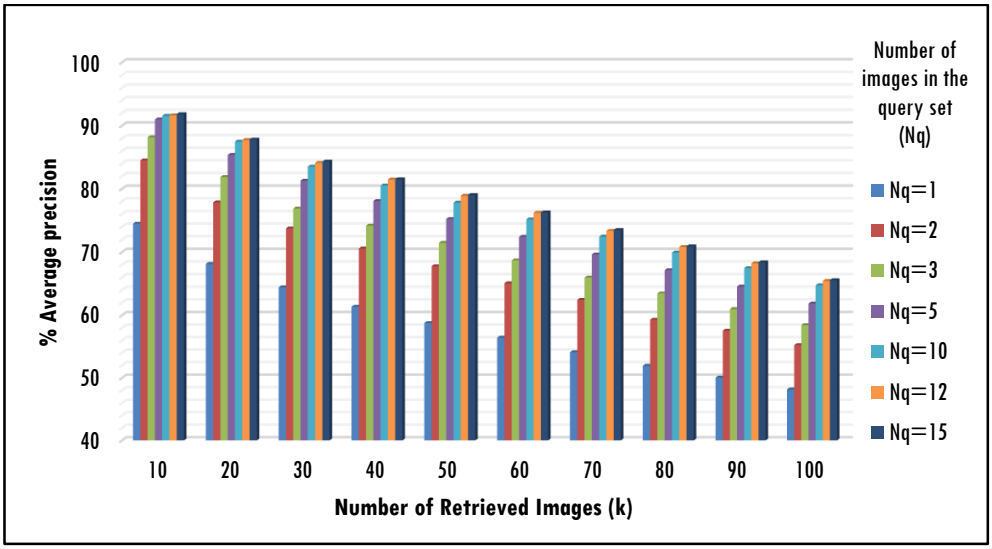


Figure 5.2 Average precision of the retrieved images for different values of k with varying number of images (Nq) in the query set.

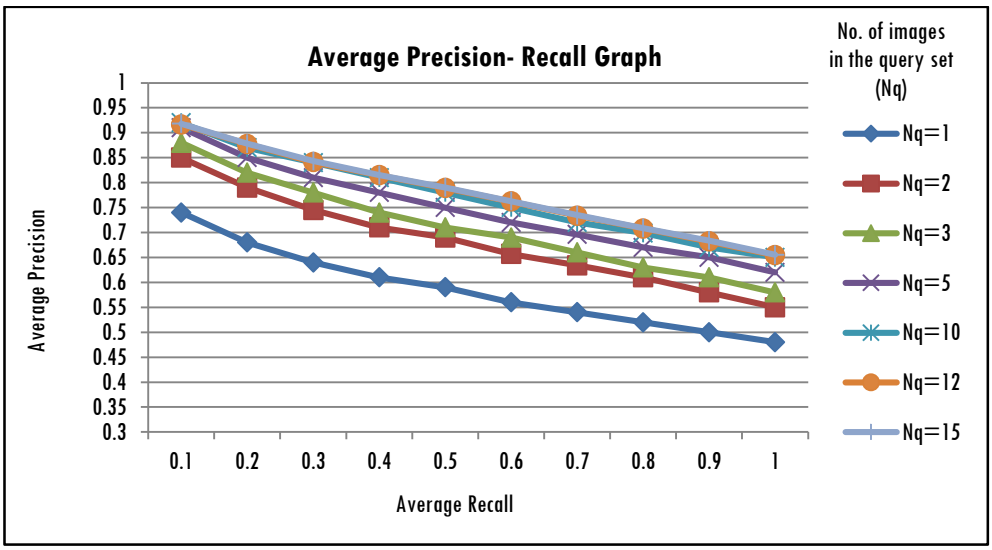


Figure 5.3 Average precision of the retrieved images for different values of k with varying number of images, Nq in the query set.

Dataset 2 (Coil 100 DB): For the Coil 100 DB, the images are represented with CEDD descriptor and the retrieval is performed for all the 7200 images in the database, considering Nq=1 and 2. Table 5.3 shows the results of retrieval. The average precision recall graph for the same is depicted in Figure 5.4. As with dataset 1, here also, the retrieval performance is enhanced by applying feature replacement algorithm.

Table 5.3 Average Precision of the retrieved images for different number of retrieved images (k=10, 20, 50, 72) on Coil100 dataset for different feature combinations

Number of images in the query set (Nq)	k=72 (Recall)	k=50	k=20	k=10
1	70.86	79.33	91.78	96.29
2	76.99	85.37	94.78	98.15

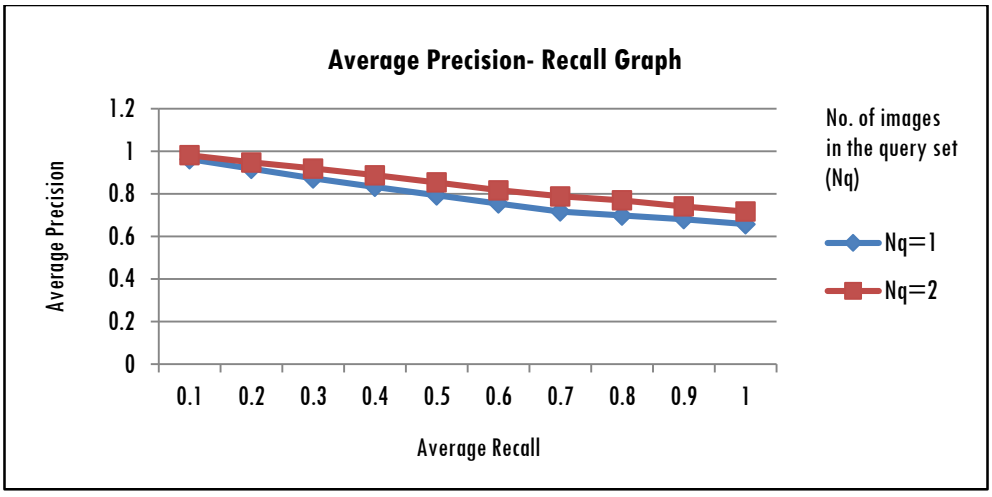


Figure 5.4 Average precision recall graph for the retrieved images (dataset2) with different values of Nq.

5.4.2 Response of the Algorithm to Different Feature Combinations

To study the response of the algorithm to different feature representations, the experiments are conducted on the Wang's database by representing images with different features. The Wang's database is chosen, as it is compact and consists of diverse categories of images. The following combinations of features are considered:

- HSV colour feature with EHD
- HSV colour with GLCM

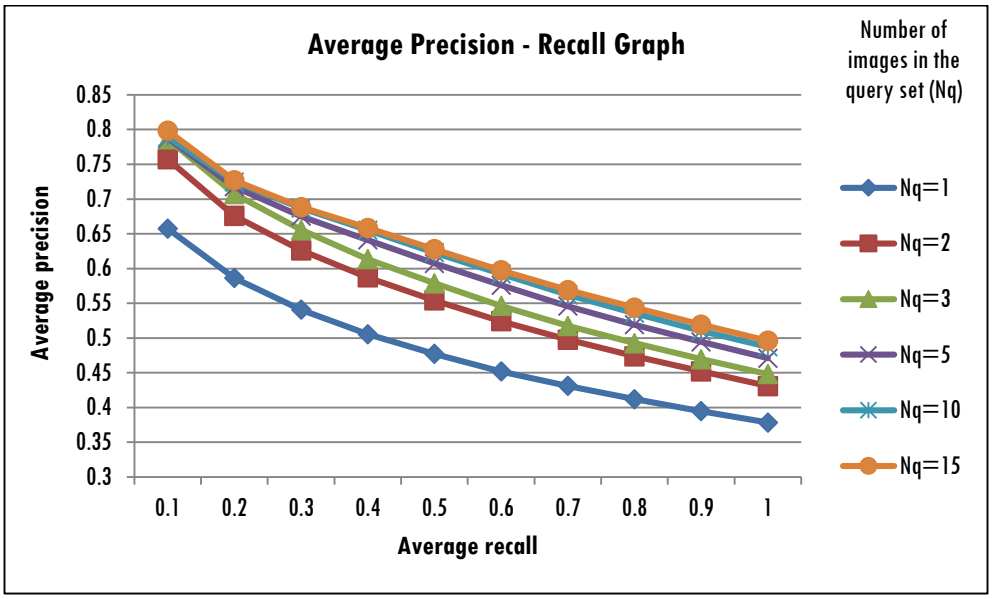


Figure 5.5 Average precision recall graph for the retrieved images (dataset1) with different values of Nq with HSV colour and EHD texture features.

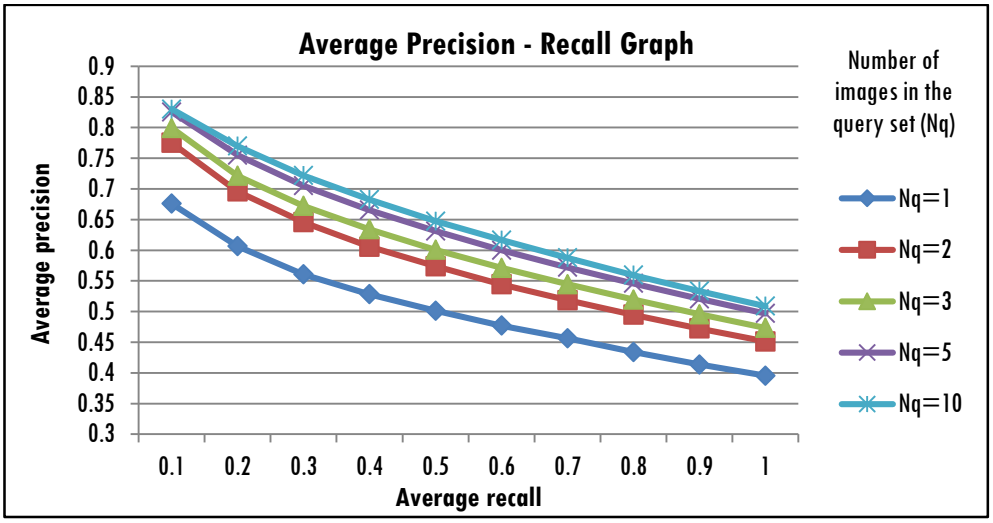


Figure 5.6 Average precision-recall graph for the retrieved images (dataset1) with different values of Nq with HSV colour and GLCM texture features.

Figures 5.5 and 5.6 show the average precision recall graphs of the retrieved images for the above two feature representations. It can be seen that for different feature representations also the performance of retrieval can be boosted by applying the proposed feature replacement algorithm.

5.4.3 Response Time of the Algorithm in Databases of Different Sizes

To analyse the average time required by the algorithm to retrieve images, the retrieval is performed on two different databases containing different number of images, i.e., 1K (Wang's DB) and 7.2K (Coil DB). The CEDD feature is used to represent the images in both the cases. The average time required by the algorithm with varying number of images in the query set is shown in Table 5.4. It is seen that the response time increases with the

number of images in the query set. Hence, the algorithm is suitable for retrieval in moderate image data sets. It should be noted that, though the response time increases with the size of data set, it is significantly low compared to the region-based retrieval systems.

Table 5.4 Average response time of the algorithm with different number of images in the query set.

No. of images in the query image set (Nq)	Average time required to retrieve images (s)	
	Wang's DB (1K)	Coil 100 DB (7.2K)
1	0.34	1.45
2	0.6	3.46
3	0.79	4.82
5	1.18	7.63
10	2.14	14.6
15	2.99	21.6

5.4.4 Performance Comparison with Other Systems

The performance of feature replacement algorithm is compared with other recent systems employing multiple images (Table 5.5). These images are either provided by the user initially or obtained through relevance feedback in response to the initial retrieved results. Retrieval is performed in the Wang's dataset and the percentage average precision of the top 20 retrieved images are used as evaluation metric for comparison. In (Liu and Peng, 2014), quantum particle swarm optimization relevance feedback algorithm based on weight adjustment is used for improving the retrieval performance and the results obtained after sixth iteration is shown in the Table. In (Banerjee & Kundu, 2009), a feature-reweighting scheme is

employed, obtaining relevant and irrelevant images through relevance feedback from the user on the top 20 results of the initial retrieval. (Irtaza et al., 2013) employs support vector machines in a grid computing environment for learning a classification model and for further retrieval. Here, for initial training, 30% of images of the concerned category are taken as positive samples and 30% of the images from the remaining categories are used as negative samples. In the joint query averaging method (Arandjelovic & Zisserman, 2012; Chatfield et. al, 2015), the feature vectors of the images in the query set are averaged to form the representative query and retrieval is performed by computing the similarity of the images in the database with this refined query vector.

Table 5.5 Performance comparison with systems employing multiple images

Category	Average precision of the retrieved images (k=20)				
	Feature adaptation method (Liu & Peng, 2014)	**Feature reweighting (Banerjee et al., 2009)**	**SVM in grid computing environment (Irtaza et al., 2013)**	**Joint Query average (Arandjelovic & Zisserman, 2012; Chatfield et al., 2015)**	**Proposed method when Nq=2**
Africa	63.8	61.00	69	72.53	78.13
Beaches	58.4	56.23	60	66.50	68.20
Buildings	50.2	63.67	56	68.05	68.90
Bus	84.4	72.77	89	85.80	87.85
Dinosaur	98	95.00	96	100.00	100.0
Elephant	48.4	77.00	64	49.70	53.55
Flowers	40.4	83.00	97	88.45	90.85
Horse	92.1	95.00	66	98.25	97.80
Mountain	42	68.00	57	62.55	63.55
Food	70.4	57.00	86	66.85	69.85
Average	**64.81**	**72.86**	**74**	**75.8**	**77.86**

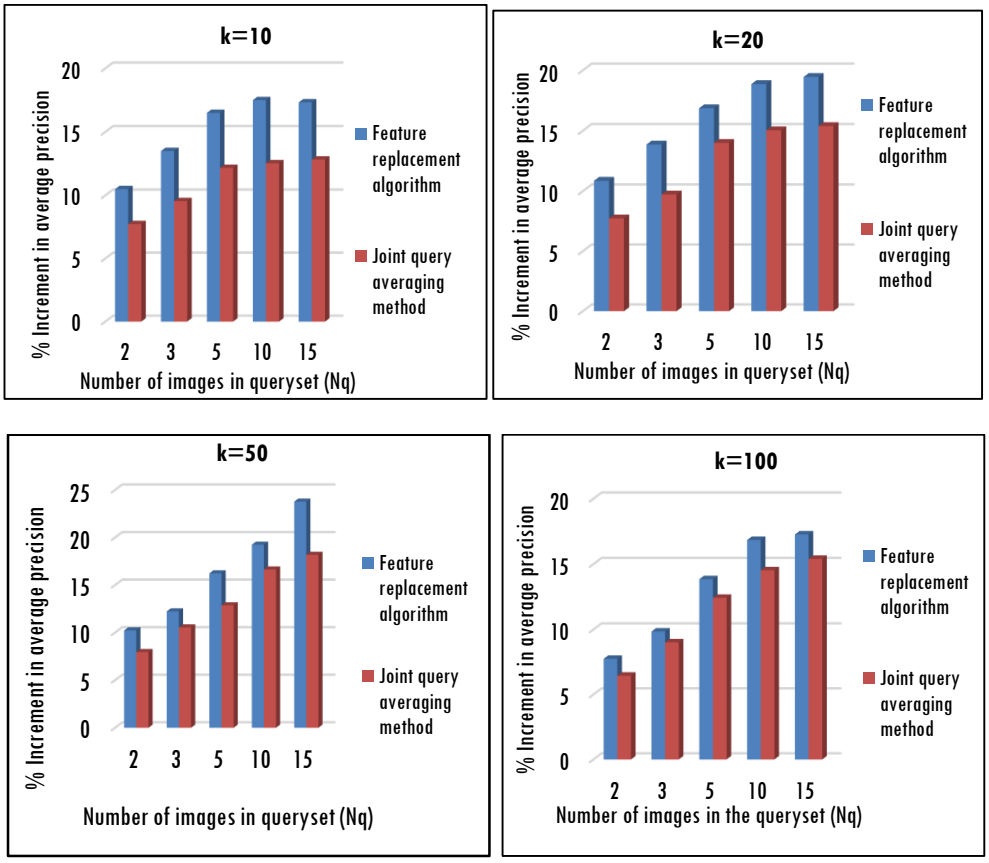
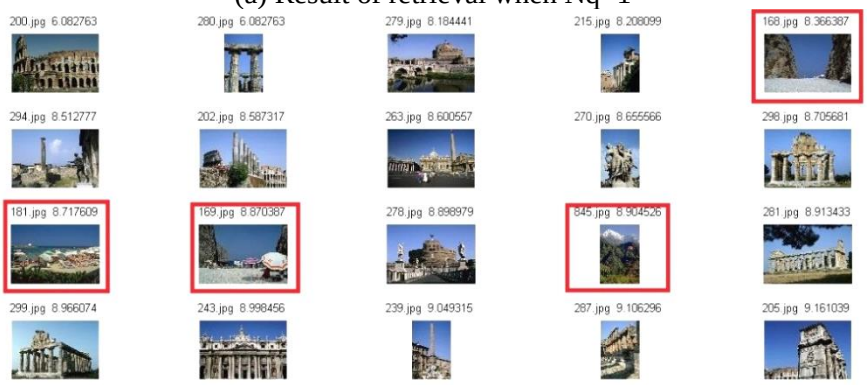


Figure 5.7 Comparison of joint query averaging method and the proposed method at different precision values with varying number of images in the query set.

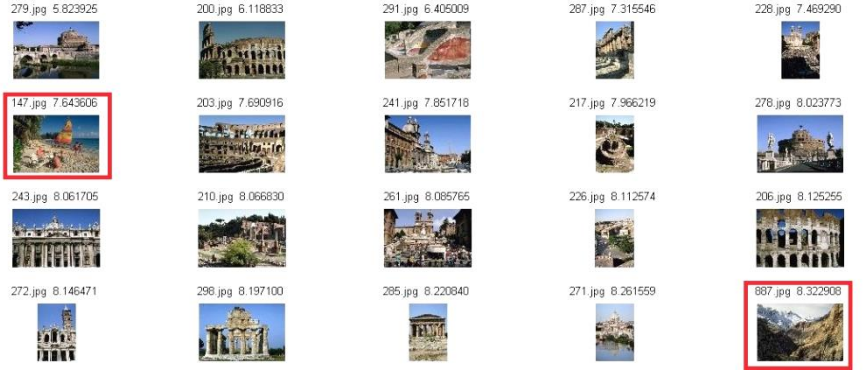
It is seen from Table 5.5 that the proposed method employing feature replacement algorithm out-performs all other considered methods. It should be noted that for both joint query averaging method and the proposed method, there are only two positive images in the query set, showing the efficacy of the algorithm in achieving better precision with minimum number of positive samples.

As joint query averaging method has shown better performance compared to other methods, detailed study is carried out to compare the performance of the proposed feature replacement algorithm with the same by varying the number of images in the query set. The images are characterized with CEDD features for this purpose. Figure 5.7 shows the percentage increment in average precision, attained by adding additional number of images to the query-set using the proposed method and joint query averaging method, when varying number of images are retrieved (k). It can be seen that the proposed algorithm out-performs joint query averaging method at every stage of retrieval. For example, when top 10 images are retrieved with two images in the query-set (Nq=2), the proposed method shows an increment of 10.5% while that of joint query average method is 7.73%. For both the cases, the increment in average precision is computed with respect to the result of retrieval when single image is issued as query, i.e., when Nq=1. However, as the number of images in the query set increases, the algorithm tends to behave more like joint-query averaging method. This is because, the feature replacement algorithm works by summing up the displacement of centroids from the initial centroid of the query-set and shows more discriminative property when there are less number of images in the query set. As the number of images in the query-set increases, the resultant sum of displacements tend to become the initial query-set centroid. However this is advantageous in multi-query systems, because, in real life the user usually have limited number of relevant images for the initial query-set.

(a) Result of retrieval when Nq=1



(b) Result of retrieval when Nq=2



(c) Result of retrieval when Nq=3

Figure 5.8 Sample output of the proposed system using (a) Single query, Nq=1 (b) Retrieval result when Nq=2 (c) Retrieval result when Nq=3.

Figure 5.8 depicts the results of retrieval for a sample query when varying number of images are there in the query set. It can be seen that the performance of retrieval improves with every additional image in the query set.

5.5 Summary

In this chapter, a feature replacement algorithm is proposed for computing the similarity between the query image set and the candidate images in the dataset in a multiple query environment. Experimental results show that by using a few positive query images, the retrieval precision of a CBIR system can be improved significantly. Compared to the other multi-query systems employing query-averaging methods, feature weighting methods and other supervised learning methods, the proposed system is found to have better performance with limited number of query images. It requires only positive images while most of the supervised learning approaches require substantial number of both positive and negative samples for learning the model for satisfactory performance. The algorithm is found to be effective for retrieval in smaller datasets as the computational time increases with the size of the dataset and the number of images in the query set. However, compared to region-based retrieval systems, the proposed method is found to be significantly faster. The algorithm can be

applied on larger datasets by choosing limited number of images in the query set to improve retrieval time. Since the method is proved to attain better performance with limited number of positive images relevant to the query, it can be effectively used in systems employing relevance feedback, as minimum effort is needed from the user.

Chapter

6 CONCLUSION AND FUTURE DIRECTIONS

Contents

This chapter sums up the results, highlights the achievements and points out some of the limitations of the research work carried out. A few suggestions for future work are also outlined.

6.1 Thesis Summary and Conclusions

Content Based Image Retrieval has drawn significant attention of the researchers in recent years due to the exponential growth of digital imagery generated, accumulated and stored in numerous fields with the advancements in multimedia technologies, widespread use of Internet and declining cost of storage devices. A variety of techniques has been developed in the past for CBIR, exploiting various imagery features ranging from global features to region and local features. In this research work, various unsupervised CBIR schemes are explored, especially region based and local feature based schemes, and some methods are proposed to

improve the retrieval efficiency of generic CBIR systems mainly focusing on the scalability, retrieval precision and response time.

Chapter 2 has presented an overview of the CBIR architecture and a review of various CBIR approaches. While exploring the region based retrieval schemes, the major challenges noted were the identification of significant regions in images and the computational load incurred during the similarity matching between images. Hence, to address these issues, a salient sub-block based approach, utilizing both local and global features of the image along with a minimum distance algorithm to match the image regions, is proposed in Chapter 3. Unlike other block based retrieval systems, which requires all the sub-blocks of an image for similarity computation and retrieval, the proposed approach involved only selected sub-blocks for this purpose, leading to less number of comparisons during region matching process and reducing the computational time by 28% in addition to maintaining competent retrieval efficiency. The minimum distance algorithm, though simple, plays a major role in reducing the retrieval time. Experimental results corroborates that the proposed method yields competent results with other systems. It is also proven that the minimum distance algorithm can be successfully employed for region comparison in RBIR systems without compromising the retrieval precision while considerably reducing the retrieval time.

Despite the advantages of the above mentioned salient sub-block scheme, like the other RBIR systems, it might not be suitable for retrieval in

large data sets, which is often the case in real-life, because of the still high response time arising from the multiple sub-block comparisons. Hence, focusing on holistic image representation methods, the bag of visual words framework was explored, which characterize the image as a histogram built on local features.

The BoVW based approaches are known to have high scalability and are widely used for object recognition, classification and retrieval purposes. However, their performance is found to be incompetent with RBIR schemes in natural image datasets mainly due to the presence of rich colour and texture information, which are not prime factors considered in BoVW based image characterization. This is because, bag of visual words are mostly built on key-point based invariant descriptors like SIFT or SURF, which are generally computed based on the intensity information of the image. Hence, various methods to incorporate multiple features in BoVW framework, such as late and early fusion approaches are explored and a combined approach, exploiting both early and late fusion, which integrates colour and edge distribution in the image with invariant descriptor, is introduced and described in Chapter 4. The new descriptor combined colour and edge distribution in the image through early fusion, and the histogram built based on this descriptor is joined with invariant descriptor (SURF) based histogram through late fusion. The resultant fusion histogram is used to represent the image. Experiments carried out in Wang's, Corel5K and COIL 100 datasets showed 8.1%, 6.9% and 11% increments respectively in the average precisions of the top retrieved results using the combined

histograms to that of SURF based histograms. Also, it is observed that the proposed method outperforms many of the recent feature fusion methods that integrate SIFT with LBP, Color Difference Histogram (CDH) with Angular Radial Transform, LBP with edge information etc. However, the datasets considered here for carrying out the experiments were static and had well-defined categories with considerable number of images in each category. Because of this, codebooks with limited size only were needed to represent the images for obtaining competent retrieval results with recent similar systems. Nevertheless, in real-life scenario, the retrieval is usually performed in databases with assorted images. In such cases, larger codebooks are needed for effective representation and hence retrieval of images.

With the intention of further improving the retrieval efficacy, a feature replacement algorithm has been presented for similarity computation in a multi-query environment. Multiple queries are often used in CBIR systems with the idea of gathering additional information about the user's requirement. In a generic multi-query system, the features gathered from the query image set are used for learning a discriminative classification model (if both positive and negative images are included in the query set) or methods such as query averaging, query point movement, feature reweighting etc. are employed to rank the images in the dataset for retrieving relevant images. Despite the methodology used, ultimately the retrieval is performed based on the similarity of the candidate images with the query. Hence, the proposed feature replacement algorithm focused only on computing the similarity

between the query set and the candidate images. The query set included only positive images, reducing the burden of the user in providing negative images also. The algorithm exploited the fact that if an image in the query set is replaced with a target image from the dataset, it will cause minimum displacement of the centroid of the query set, if it is a candidate for retrieval. The feature replacement algorithm hence computed the cumulative displacement of the centroid caused by replacing the elements of the query set with the target image. Experimental results showed 10% increment in the average precision simply by having two images in the query set. This is a desirable property in general CBIR as well as in systems employing relevance feedback, as better results can be obtained with limited input from the user.

6.2 Limitations

There were some constraints while developing and evaluating the proposed approaches. Some of them are listed below:

- Numerous features are available in literature for characterizing the images for CBIR and related applications. We have not analysed all those features while selecting features for this research work. We have employed some of the well explored and commonly used features pertaining to colour, texture and interest point based invariant descriptors such as SURF to represent the images. Hence, there may be other feature combinations, which can provide more effective and faster retrieval performance.

- For evaluating different methods proposed in this work, publically available datasets that are commonly used for CBIR evaluation are used. All these datasets consisted of well-defined categories with sizable number of images in each category, which might not be the real life scenario.
- The response time recorded for various experiments here are hardware dependant. Hence, for some of the proposed methodologies, the time taken for retrieval could not be compared with that of other systems, as code for the works were not freely and publically available and the reported results were obtained using different hardware architecture and experimental settings.

6.3 Future Work

The present work was an attempt to develop some methodologies to improve the retrieval efficacy of general CBIR systems mainly focusing on the scalability, response time and performance aspects. This section discusses some potentially promising directions for future work, which we believe, will render further improvements.

- In various approaches considered in this thesis, we have used statistical features for exemplifying the images. Many works employing transformation domain features based on ridgelets, curvelets, wavelets, ripplets etc. have exhibited good performance, which can be tried using the proposed methods to attain improved results.

- In the integrated feature BoVW framework, we have incorporated spatial information only to improve the performance of retrieval. Techniques to improve the quality of feature bags such as TF-IDF, Singular Value Decomposition etc. are not considered, employing which may lead to the formation of more meaningful feature bags leading to improved retrieval.

  It is observed that the BoVW are quite sensitive to the visual vocabulary and the dataset based on which it is built. Finding ways to construct vocabularies independent of datasets will be very useful for applications employing local features.

  Visual vocabularies are usually constructed with static datasets, wherein real-life, the repositories are dynamic and undergo frequent updations. Finding methods to construct dynamic-adaptable vocabularies will be beneficial for real-world applications.

- Distance metrics play an important role in image matching and retrieval. Images represented using same features and matched using different distance metrics generally yield varying results and response time. Some systems try to overcome this problem by employing majority voting and fusion schemes, which consider the outputs of more than one methods in formulating the results. However, the computational cost and complexity involved are much higher. Finding ways to tackle this problem will be a great benefit.

- Region-based approaches are used in many imaging applications. Despite their capability in retrieving relevant images effectively, due to the high response time, the usage is confined to small datasets. Hence, large scale retrieval applications prefer global image feature representations to region based approaches. Devising techniques to combine local and region based approaches to represent images in the form of a holistic compact descriptors will be advantageous.
- The computational time for feature extraction and distance computation can be reduced by exploiting parallel computation capabilities of GPUs, clusters and clouds.
- For real word applications, meta data can be incorporated as an additional information to enhance retrieval.

BIBLIOGRAPHY

[1] (Abdel-Hakim & Farag, 2006) Abdel-Hakim, A. E., & Farag, A. A. (2006). CSIFT: A SIFT descriptor with colour invariant characteristics. In Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (Vol. 2, pp. 1978-1983). IEEE.

[2] (Altmann & Reitbock, 1984) Altmann, J., & Reitbock, H. J. (1984). A fast correlation method for scale-and translation-invariant pattern recognition. IEEE transactions on pattern analysis and machine intelligence, (1), 46-57.

[3] (Alzu'bi, Amira & Ramzan, 2015) Alzu'bi, A., Amira, A., & Ramzan, N. (2015). Semantic content-based image retrieval: A comprehensive study. Journal of Visual Communication and Image Representation, 32, 20-54.

[4] (Arandjelović & Zisserman, 2012a) Arandjelović, R., & Zisserman, A. (2012, June). Three things everyone should know to improve object retrieval. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012 (pp. 2911-2918). IEEE.

[5] (Arandjelović & Zisserman, 2012b) Arandjelovic, R., & Zisserman, A. (2012, September). Multiple queries for large scale specific object retrieval. In Proceedings of British Machine Vision Conference (BMVC) (pp. 1-11).

[6] (Aulia,, 2005) Aulia, E. (2005). Hierarchical Indexing for Region based image retrieval (Doctoral dissertation, Faculty of the Louisiana State University and Agricultural and Mechanical College In partial fulfilment of the Master of Science in Industrial Engineering in The Department of Industrial and Manufacturing Systems Engineering by Eka Aulia BS, Louisiana State University).

[7] (Bach et al., 1996) Bach, J. R., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B., Humphrey, R., & Shu, C. F. (1996, March). Virage image search engine: an open framework for image management. In Electronic Imaging: Science & Technology (pp. 76-87). International Society for Optics and Photonics.

[8] (Banerjee, Kundu & Maji, 2009) Banerjee, M., Kundu, M. K., & Maji, P. (2009). Content-based image retrieval using visually significant point features. J. Fuzzy Sets and Systems, 160(23), 3323-3341.

[9] (Bay et al., 2008) Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). J. Computer Vision and Image Understanding, 110(3), 346-359.

[10] (Bhagavathy & Chhabra, 2007) Bhagavathy, S., &Chhabra, K. (2007). A wavelet-based image retrieval system. Journal of University of California, Santa Barbara, ECE A, 278.

[11] (Bosch, Zisserman & Muñoz, 2008) Bosch, A., Zisserman, A., & Muñoz, X. (2008). Scene classification using a hybrid generative/discriminative approach. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(4), 712-727.

[12] (Calonder et al., 2010) Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010, September). Brief: Binary robust independent elementary features. In Proceedings of European Conference on Computer Vision (pp. 778-792). Springer Berlin Heidelberg.

[13] (Carson et al., 1999) Carson, C., Thomas, M., Belongie, S., Hellerstein, J. M., & Malik, J. (1999, June). Blobworld: A system for region-based image indexing and retrieval. In Proceedings of International Conference on Advances in Visual Information Systems (pp. 509-517). Springer Berlin Heidelberg.

[14] (Chatfield et al., 2015a) Chatfield, K., Arandjelović, R., Parkhi, O., & Zisserman, A. (2015). On-the-fly learning for visual search of large-scale image and video datasets. International Journal of Multimedia Information Retrieval, 4(2), 75-93.

[15] (Chatfield et al., 2011b) Chatfield, K., Lempitsky, V. S., Vedaldi, A., & Zisserman, A. (2011, September). The devil is in the details: an evaluation of recent feature encoding methods. In Proceedings of British machine Vision Conference (BMVC) (Vol. 2, No. 4, p. 8)

[16] (Chatzichristofis & Boutalis, 2008a) Chatzichristofis, S. A., & Boutalis, Y. S. (2008, May). CEDD: colour and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In Proceeding of International Conference on Computer Vision Systems (pp. 312-322). Springer Berlin Heidelberg.

[17] (Chatzichristofis & Boutalis, 2008b) Chatzichristofis, S. A., & Boutalis, Y. S. (2008, May). FCTH: Fuzzy colour and texture histogram-a low-level feature for accurate image retrieval. In Proceedings of 2008, Ninth International Workshop on Image Analysis for Multimedia Interactive Services (pp. 191-196). IEEE.

[18] (Chen, Hu & Shen, 2009) Chen, X., Hu, X., & Shen, X. (2009, April). Spatial weighting for bag-of-visual-words and its application in content-based image retrieval. In Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 867-874). Springer Berlin Heidelberg.

[19] (Chen, Li & Wang, 2006) Chen, Y., Li, J., & Wang, J. Z. (2006). Machine Learning and Statistical Modeling Approaches to Image Retrieval (Vol. 14). Springer Science & Business Media.

[20] (Chen et al., 2012) Chen, Y., Li, X., Dick, A., & van den Hengel, A. (2012). Boosting object retrieval with group queries. IEEE Signal Processing Letters, 19(11), 765-768.

[21] (Chen, Wang & Krovetz, 2005) Chen, Y., Wang, J. Z., & Krovetz, R. (2005). CLUE: cluster-based retrieval of images by unsupervised learning. IEEE transactions on Image Processing, 14(8), 1187-1201.

[22] (Cheng, Kuo & Chen, 2006) Cheng, S. C., Kuo, C. T., & Chen, H. J. (2006, October). Invarint Image Retrieval using Block-Based Visual Pattern Matching. In Proceedings of 2006 International Conference on Image Processing (pp. 1461-1464). IEEE.

[23] (Chu & Smeulders, 2010) Chu, D. M., & Smeulders, A. W. (2010, September). Color invariant surf in discriminative object tracking. In Proceedings of European Conference on Computer Vision (pp. 62-75). Springer Berlin Heidelberg.

[24] (Chum, 2010) Chum, O. (2010). Large-scale discovery of spatially related images. IEEE transactions on pattern analysis and machine intelligence, 32(2), 371-377.

[25] (Chum et al., 2007) Chum, O., Philbin, J., Sivic, J., Isard, M., & Zisserman, A. (2007, October). Total recall: Automatic query expansion with a generative feature model for object retrieval. In proceedings of 2007 IEEE 11th International Conference on Computer Vision (pp. 1-8). IEEE.

[26] (Dalal & Triggs, 2005) Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE.

[27] (Datta et al., 2008) Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys (CSUR), 40(2), 5.

[28] (Deselaers, Keysers & Ney, 2008) Deselaers, T., Keysers, D., & Ney, H. (2008). Features for image retrieval: an experimental comparison. J. Information Retrieval, 11(2), 77-107.

[29] (Do & Vetterli, 2003) Do, M. N., & Vetterli, M. (2003). The finite ridgelet transform for image representation. IEEE Transactions on Image Processing, 12(1), 16-28.

[30] (Everingham et al., 2015) Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. International Journal of Computer Vision, 111(1), 98-136.

[31] (Fakheri et al., 2013) Fakheri, M., Sedghi, T., Shayesteh, M. G., & Amirani, M. C. (2013). Framework for image retrieval using machine learning and statistical similarity matching techniques. J. IET Image Processing, 7(1), 1-11.

[32] (Fan et al., 2009) Fan, P., Men, A., Chen, M., & Yang, B. (2009, November). Color-SURF: A surf descriptor with local kernel colour histograms. In Proceedings of 2009 IEEE International Conference on Network Infrastructure and Digital Content (pp. 726-730). IEEE.

[33] (Fang et al., 2012) Fang, M. Y., Kuan, Y. H., Kuo, C. M., & Hsieh, C. H. (2012). Effective image retrieval techniques based on novel salient region segmentation and relevance feedback. International Journal of Multimedia Tools and Applications, 57(3), 501-525.

[34] (Farooq, 2016) Farooq, J. (2016) Object Detection and Identification using SURF and BoW Model. In Proceedings of 2016 International Conference on Computing, Electronic and Electrical Engineering (pp. 318-323)

[35] (Fei-Fei & Perona, 2005) Fei-Fei, L., & Perona, P. (2005, June). A bayesian hierarchical model for learning natural scene categories. In Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (Vol. 2, pp. 524-531). IEEE.

[36] (Fernando & Tuytelaars, 2013) Fernando, B., & Tuytelaars, T. (2013). Mining multiple queries for image retrieval: On-the-fly learning of an object-specific mid-level representation. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2544-2551).

[37] (Flickner, 1995) Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B. & Steele, D. (1995). Query by image and video content: The QBIC system. J. Computer, 28(9), 23-32.

[38] (Fu et al., 2012) Fu, J., Jing, X., Sun, S., Lu, Y., & Wang, Y. (2012, May). C-surf: Colored speeded up robust features. In International Conference on Trustworthy Computing and Services (pp. 203-210). Springer Berlin Heidelberg.

[39] (Gadkari, 2004) Gadkari, D. (2004). Image quality analysis using GLCM.

[40] (Gehler & Nowozin, 2009) Gehler, P., &Nowozin, S. (2009, September). On feature combination for multiclass object classification. In Proceedings of 2009 IEEE 12th International Conference on Computer Vision (pp. 221-228). IEEE.

[41] (Giouvanakis & Kotropoulos, 2014) Giouvanakis, E., & Kotropoulos, C. (2014, August). Saliency map driven image retrieval combining the bag-of-words model and PLSA. In Proceedings of 2014 19th International Conference on Digital Signal Processing (pp. 280-285). IEEE.

[42] (Gordo et al., 2016) Gordo, A., Almazan, J., Revaud, J., & Larlus, D. (2016). Deep Image Retrieval: Learning global representations for image search. arXiv preprint arXiv:1604.01325.

[43] (Grigorova et al., 2007) Grigorova, A., De Natale, F. G., Dagli, C., & Huang, T. S. (2007). Content-based image retrieval by feature adaptation and relevance feedback. IEEE Transactions on Multimedia, 9(6), 1183-1192.

[44] (Hafner et al., 1995) Hafner, J., Sawhney, H. S., Equitz, W., Flickner, M., & Niblack, W. (1995). Efficient colour histogram indexing for quadratic form distance functions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(7), 729-736.

[45] (Haralick & Shanmugam, 1973) Haralick, R. M., & Shanmugam, K. (1973). Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, (6), 610-621.

[46] (Hervé & Boujemaa, 2007) Hervé, N., & Boujemaa, N. (2007, July). Image annotation: which approach for realistic databases? In Proceedings of the sixth ACM international conference on Image and video retrieval (pp. 170-177). ACM.

[47] (Hiremath & Pujari, 2007) Hiremath, P. S., & Pujari, J. (2007, December). Content based image retrieval using colour, texture and shape features. In Proceedings of the International Conference on Advanced Computing and Communications, 2007. ADCOM 2007. (pp. 780-784). IEEE.

[48] (Hiremath & Pujari, 2008) Hiremath, P. S., & Pujari, J. (2008). Content-based image retrieval using colour boosted salient points and shape features of an image. International Journal of Image Processing, 2(1), 10-17.

[49] (Hsiao et al., 2007) Hsiao, J. H., Chen, C. S., Chien, L. F., & Chen, M. S. (2007). A new approach to image copy detection based on extended feature sets. IEEE Transactions on Image Processing, 16(8), 2069-2079.

[50] (Hsiao et al., 2010) Hsiao, M. J., Huang, Y. P., Tsai, T., & Chiang, T. W. (2010). An efficient and flexible matching strategy for content-based image retrieval. Life Science Journal, 7(1), 99-106.

[51] (Hu, 1962) Hu, M. K. (1962). Visual pattern recognition by moment invariants. IRE Transactions on Information Theory, 8(2), 179-187.

[52] (Huang, Gao, & Chan, 2010) Huang, W., Gao, Y., & Chan, K. L. (2010). A review of region-based image retrieval. Journal of Signal Processing Systems, 59(2), 143-161.

[53] (Huang & Dai, 2003) Huang, P. W., & Dai, S. K. (2003). Image retrieval by texture similarity. J. Pattern Recognition, 36(3), 665-679.

[54] (Huang et al., 1997) Huang, J., Kumar, S. R., Mitra, M., Zhu, W. J., & Zabih, R. (1997, June). Image indexing using colorcorrelograms. In Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (pp. 762-768). IEEE.

[55] (Ion, Stanescu & Burdescu, 2007) Ion, A. L., Stanescu, L., & Burdescu, D. (2007, September). Semantic based image retrieval using relevance feedback. In Proceedings of EUROCON, 2007. International Conference on Computer as a Tool, (pp. 303-310). IEEE.

[56] (Irtaza et al., 2013) Irtaza, A., Jaffar, M. A., & Mahmood, M. T. (2013, August). Semantic image retrieval in a grid computing environment using support vector machines. The Computer Journal, Vol.57, issue 2, pp. 205-216.

[57] (Jalilvand, Boroujeni & Charkari, 2011) Jalilvand, A., Boroujeni, H. S., & Charkari, N. M. (2011, July). CH-SIFT: A local kernel colour histogram SIFT based descriptor. In Proceedings of 2011 International Conference on Multimedia Technology (ICMT), (pp. 6269-6272). IEEE.

[58] (Jégou, Douze & Schmid, 2010) Jégou, H., Douze, M., & Schmid, C. (2010). Improving bag-of-features for large-scale image search. International Journal of Computer Vision, 87(3), 316-336.

[59] (Jhanwar et al., 2004) Jhanwar, N., Chaudhuri, S., Seetharaman, G., & Zavidovique, B. (2004). Content based image retrieval using motif cooccurrence matrix. J. Image and Vision Computing, 22(14), 1211-1220.

[60] (Jurie & Triggs, 2005) Jurie, F., &Triggs, B. (2005, October). Creating efficient codebooks for visual recognition. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 (Vol. 1, pp. 604-610). IEEE.

[61] (Kanimozhi & Latha, 2015) Kanimozhi, T., & Latha, K. (2015). An integrated approach to region based image retrieval using firefly algorithm and support vector machine. J. Neurocomputing, 151, 1099-1111.

[62] (Kaplan, Murenzi, & Namuduri, 1997) Kaplan, L. M., Murenzi, R., & Namuduri, K. R. (1997, December). Fast texture database retrieval using extended fractal features. In Proceedings of Photonics West'98 Electronic Imaging (pp. 162-173). International Society for Optics and Photonics.

[63] (Ke & Sukthankar, 2004) Ke, Y., & Sukthankar, R. (2004, June). 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. (Vol. 2, pp. II-506). IEEE.

[64] (Khan et al., 2013) Khan, F. S., Anwer, R. M., Van De Weijer, J., Bagdanov, A. D., Lopez, A. M., & Felsberg, M. (2013). Coloring action recognition in still images. International journal of Computer Vision, 105(3), 205-221.

[65] (Khan, Van De Weijer & Vanrell, 2009) Khan, F. S., Van De Weijer, J., & Vanrell, M. (2009, September). Top-down colour attention for object recognition. In Proceedings of 2009 12th IEEE International Conference on Computer Vision (pp. 979-986). IEEE.

[66] (Khan, Van De Weijer & Vanrell, 2012) Khan, F. S., Van de Weijer, J., & Vanrell, M. (2012). Modulating shape features by colour attention for object recognition. International Journal of Computer Vision, 98(1), 49-64.

[67] (Khotanzad & Hong, 1990) Khotanzad, A., & Hong, Y. H. (1990). Invariant image recognition by Zernike moments. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(5), 489-497.

[68] (Kimura et al., 2011) Kimura, P. A., Cavalcanti, J. M., Saraiva, P. C., Torres, R. D. S., & Gonçalves, M. A. (2011). Evaluating retrieval effectiveness of descriptors for searching in large image databases. Journal of Information and Data Management, 2(3), 305.

[69] (Krizhevsky, Sutskever & Hinton, 2012) Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Proceedings of Advances in Neural Information Processing Systems (pp. 1097-1105).

[70] (Lakshmi, Nema & Rakshit, 2015) Lakshmi, A., Nema, M., & Rakshit, S. (2015). Long term Relevance Feedback: A Probabilistic axis re-weighting update scheme. IEEE Signal Processing Letters, 22(7), 852-856.

[71] (Law, Thome & Cord, 2014) Law, M. T., Thome, N., & Cord, M. (2014). Bag-of-words image representation: Key ideas and further insight. In Fusion in Computer Vision (pp. 29-52). Springer International Publishing.

[72] (Lazebnik, Schmid & Ponce, 2006) Lazebnik, S., Schmid, C., & Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (Vol. 2, pp. 2169-2178). IEEE.

[73] (Leutenegger, Chli & Siegwart, 2011) Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011, November). BRISK: Binary robust invariant scalable keypoints. In Proceedings of the 2011

International conference on computer vision (pp. 2548-2555). IEEE.

[74] (Li & Allinson, 2013) Li, J., &Allinson, N. M. (2013). Relevance feedback in content-based image retrieval: a survey. In Handbook on Neural Information Processing (pp. 433-469). Springer Berlin Heidelberg.

[75] (Li, Wang & Wiederhold, 2000) Li, J., Wang, J. Z., & Wiederhold, G. (2000, October). IRM: integrated region matching for image retrieval. In Proceedings of the eighth ACM International Conference on Multimedia (pp. 147-156). ACM.

[76] (Lin, Chen & Chan, 2009) Lin, C. H., Chen, R. T., & Chan, Y. K. (2009). A smart content-based image retrieval system based on colour and texture feature. J. Image and Vision Computing, 27(6), 658-665.

[77] (Lin et al., 2015) Lin, K., Yang, H. F., Hsiao, J. H., & Chen, C. S. (2015). Deep learning of binary hash codes for fast image retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 27-35).

[78] (Liu & Peng, 2014) Liu S, & Peng J. (2014) A Novel Image Retrieval Algorithm Based on Adaptive Weight Adjustment and Relevance Feedback. Journal of Computers, 9(11):2720-2726

[79] (Liu & Picard, 1996) Liu, F., & Picard, R. W. (1996). Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(7), 722-733.

[80] (Liu, 2013) Liu, J. (2013). Image retrieval based on bag-of-words model. arXiv preprint arXiv:1304.5168.

[81] (Liu et al., 2009) Liu, Y., Chen, X., Zhang, C., & Sprague, A. (2009). Semantic clustering for region-based image retrieval. Journal of Visual Communication and Image Representation, 20(2), 157-166.

[82] (Liu, Zhang & Lu, 2008) Liu, Y., Zhang, D., & Lu, G. (2008). Region-based image retrieval with high-level semantics using decision tree learning. J. Pattern Recognition, 41(8), 2554-2570.

[83] (Liu et al., 2005) Liu, Y., Zhang, D., Lu, G., & Ma, W. Y. (2005, January). Region-based image retrieval with high-level semantic colour names. In Proceedings of the 11th International Multimedia Modelling Conference (pp. 180-187). IEEE.

[84] (Liu et al., 2007) Liu, Y., Zhang, D., Lu, G., & Ma, W. Y. (2007). A survey of content-based image retrieval with high-level semantics. J. Pattern recognition, 40(1), 262-282.

[85] (Long, Zhang & Feng, 2003) Long, F., Zhang, H., & Feng, D. D. (2003). Fundamentals of content-based image retrieval. In

Multimedia Information Retrieval and Management (pp. 1-26). Springer Berlin Heidelberg.

[86] (Lowe, 1999) Lowe, D. G. (1999). Object recognition from local scale-invariant features. In Proceedings of the 1999 Seventh IEEE International Conference on Computer vision (Vol. 2, pp. 1150-1157). IEEE.

[87] (Lowe, 2004) Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.

[88] (Lu, Li & Burkhardt, 2006) Lu, Z. M., Li, S. Z., & Burkhardt, H. (2006). A content-based image retrieval scheme in JPEG compressed domain. International Journal of Innovative Computing, Information and Control, 2(4), 831-839.

[89] (Manipoonchelvi & Muneeswaran, 2014) Manipoonchelvi, P., & Muneeswaran, K. (2014). Multi region based image retrieval system. J. Sadhana, 39(2), 333-344.

[90] (Manipoonchelvi & Muneeswaran, 2015) Manipoonchelvi, P., & Muneeswaran, K. (2015). Significant region-based image retrieval. International Journal of Signal, Image and Video Processing, 9(8), 1795-1804

[91] (Manjunath & Ma, 1996) Manjunath, B. S., & Ma, W. Y. (1996). Texture features for browsing and retrieval of image data. IEEE

Transactions on Pattern Analysis and Machine Intelligence, 18(8), 837-842.

[92] (Manjunath et al., 2001) Manjunath, B. S., Ohm, J. R., Vasudevan, V. V., & Yamada, A. (2001). Color and texture descriptors. IEEE Transactions on Circuits and Systems for Video Technology, 11(6), 703-715.

[93] (Manjunath, Salembier & Sikora, 2002) Manjunath, B. S., Salembier, P., & Sikora, T. (2002). Introduction to MPEG-7: multimedia content description interface (Vol. 1). John Wiley & Sons.

[94] (Matas et al., 2004) Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. J. Image and Vision Computing, 22(10), 761-767.

[95] (Mehtre, Kankanhalli & Lee, 1997) Mehtre, B. M., Kankanhalli, M. S., & Lee, W. F. (1997). Shape measures for content based image retrieval: a comparison. J. Information Processing & Management, 33(3), 319-337.

[96] (Mukundan & Ramakrishnan, 1998) Mukundan, R., & Ramakrishnan, K. R. (1998). Moment functions in image analysis: theory and applications (Vol. 100). Singapore: World Scientific.

[97] (Mukundan, Ong & Lee, 2001) Mukundan, R., Ong, S. H., & Lee, P. A. (2001). Image analysis by Tchebichef moments. IEEE Transactions on Image Processing, 10(9), 1357-1364.

[98] (Murala & Wu, 2014) Murala, S., & Wu, Q. J. (2014). Expert content-based image retrieval system using robust local patterns. Journal of Visual Communication and Image Representation, 25(6), 1324-1334.

[99] (Murala, Maheshwari & Balasubramanian, 2012) Murala, S., Maheshwari, R. P., & Balasubramanian, R. (2012). Directional local extrema patterns: a new descriptor for content based image retrieval. International Journal of Multimedia Information Retrieval, 1(3), 191-203.

[100] (Murala et al., 2013) Murala, S., Wu, Q. J., Balasubramanian, R., & Maheshwari, R. P. (2013, March). Joint histogram between colour and local extrema patterns for object tracking. In Proceedings of IS&T/SPIE Electronic Imaging (pp. 86630T-86630T). International Society for Optics and Photonics.

[101] (Ngo, Pong & Chin, 2001) Ngo, C. W., Pong, T. C., & Chin, R. T. (2001). Exploiting image-indexing techniques in DCT domain. International Journal of Pattern Recognition, 34(9), 1841-1851.

[102] (Niblack et al., 1993) Niblack, C. W., Barber, R., Equitz, W., Flickner, M. D., Glasman, E. H., Petkovic, D. &Taubin, G.

(1993, April). QBIC project: querying images by content, using colour, texture, and shape. In IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology (pp. 173-187). International Society for Optics and Photonics.

[103] (Park, Park & Won, 2000) Park, S. J., Park, D. K., & Won, C. S. (2000). Core experiments on MPEG-7 edge histogram descriptor. Technical Report, ISO/IEC JCT1/SC29/WG11-MPEG2000/M5984(2000)

[104] (Pass & Zabih, 1996) Pass, G., & Zabih, R. (1996, December). Histogram refinement for content-based image retrieval. In Proceedings 3rd IEEE Workshop on Applications of Computer Vision, 1996. WACV'96. (pp. 96-102). IEEE.

[105] (Patil & Kokare, 2011) Patil, P. B., & Kokare, M. B. (2011). Relevance Feedback in Content Based Image Retrieval: A Review. Journal of Applied Computer Science & Mathematics, (10).

[106] (Penatti, Valle & Torres, 2012) Penatti, O. A., Valle, E., & Torres, R. D. S. (2012). Comparative study of global colour and texture descriptors for web image retrieval. Journal of Visual Communication and Image Representation, 23(2), 359-380.

[107] (Pentland, Picard & Sclaroff, 1996) Pentland, A., Picard, R. W., & Sclaroff, S. (1996). Photobook: tools for content based image retrieval. International Journal of Computer Vision, 18(3), 233-254.

[108] (Perd'och, Chum & Matas, 2009) Perd'och M, Chum O & Matas J (2009) Efficient representation of local geometry for large scale object retrieval. Proceedings of CVPR. 9-16.

[109] (Philbin et al., 2007) Philbin, J., Chum, O., Isard, M., Sivic, J., & Zisserman, A. (2007, June). Object retrieval with large vocabularies and fast spatial matching. In 2007 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-8). IEEE.

[110] (Rao, Rao & Govardhan, 2011) Rao, M. B., Rao, B. P., & Govardhan, A. (2011). CTDCIRS: Content-based image retrieval system based on dominant colour and texture features. International Journal of Computer Applications, 18(6), 40-46.

[111] (Roth & Winter, 2008) Roth, P. M., & Winter, M. (2008). Survey of appearance-based methods for object recognition. Inst. for Computer Graphics and Vision, Graz University of Technology, Austria, Technical Report ICGTR0108 (ICG-TR-01/08).

[112] (Rublee et al., 2011) Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011, November). ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision (pp. 2564-2571). IEEE.

[113] (Rubner, Tomasi & Guibas, 2000) Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision, 40(2), 99-121.

[114] (Sajjanhar & Lu, 1997) Sajjanhar, A., & Lu, G. (1997). A grid-based shape indexing and retrieval method. Australian Computer Journal, 29(4), 131-140.

[115] (Sajjanhar, Lu & Wright, 1997) Sajjanhar, A., Lu, G., & Wright, J. (1997, April) b. An experimental study of moment invariants and Fourier descriptors for shape based image retrieval. In Proceedings of the second Australia document computing symposium, Melbourne, Australia (pp. 46-54).

[116] (Shahabi & Safar, 2007) Shahabi, C., & Safar, M. (2007). An experimental study of alternative shape-based image retrieval techniques. International Journal of Multimedia Tools and Applications, 32(1), 29-48.

[117] (Shi & Malik, 2000) Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 888-905.

[118] (Sikora, 2001) Sikora, T. (2001). The MPEG-7 visual standard for content description-an overview. IEEE Transactions on Circuits and Systems for Video Technology, 11(6), 696-702.

[119] (Singh, 2015) Singh, A. V. (2015). Content-Based Image Retrieval using Deep Learning.

[120] (Simonyan & Zisserman, 2014) Simonyan, K., &Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In Advances in Neural Information Processing Systems (pp. 568-576).

[121] (Sivic & Zisserman, 2003) Sivic, J., &Zisserman, A. (2003, October). Video Google: A text retrieval approach to object matching in videos. In Computer Vision, 2003. Proceedings of the Ninth IEEE International Conference on (pp. 1470-1477). IEEE.

[122] (Sivic & Zisserman, 2009) Sivic, J., &Zisserman, A. (2009). Efficient visual search of videos cast as text retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(4), 591-606.

[123] (Smeulders et al., 2000) Smeulders, A. W., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12), 1349-1380.

[124] (Smith, 1997) Smith, J. R. (1997). Integrated spatial and feature image systems: retrieval, compression and analysis (Doctoral dissertation, PhD thesis, Columbia University).

[125] (Stricker & Orengo, 1995) Stricker, M. A., &Orengo, M. (1995, March). Similarity of colour images. In IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology (pp. 381-392). International Society for Optics and Photonics

[126] (Su et al., 2011) Su, J. H., Huang, W. J., Philip, S. Y., & Tseng, V. S. (2011). Efficient relevance feedback for content-based image retrieval by mining user navigation patterns. IEEE Transactions on Knowledge and Data Engineering, 23(3), 360-372.

[127] (Su, Chen & Lien, 2010) Su, W. T., Chen, J. C., & Lien, J. J. J. (2010). Region-based image retrieval system with heuristic pre-clustering relevance feedback. J. Expert systems with Applications, 37(7), 4984-4998.

[128] (Subrahmanyam et al., 2013) Subrahmanyam, M., Wu, Q. J., Maheshwari, R. P., & Balasubramanian, R. (2013). Modified colour motif co-occurrence matrix for image indexing and retrieval. J. Computers & Electrical Engineering, 39(3), 762-774.

[129] (Suematsu et al., 2002) Suematsu, N., Ishida, Y., Hayashi, A., &Kanbara, T. (2002, May). Region-based image retrieval using wavelet transform. In Proceedings of the 15th international conf. on vision interface (pp. 9-16).

[130] (Swain & Ballard, 1991) Swain, M. J., & Ballard, D. H. (1991). Color indexing. International Journal of Computer Vision, 7(1), 11-32.

[131] (Tahaghoghi, Thom & Williams, 2001) Tahaghoghi, S. M., Thom, J. A., & Williams, H. E. (2001, January). Are two pictures better than one? In Proceedings of the 12th Australasian database conference (pp. 138-144). IEEE Computer Society.

[132] (Takala, Ahonen & Pietikäinen, 2005) Takala, V., Ahonen, T., & Pietikäinen, M. (2005, June). Block-based methods for image retrieval using local binary patterns. In Proceedings of the Scandinavian Conference on Image Analysis (pp. 882-891). Springer Berlin Heidelberg.

[133] (Tao & Grosky, 1998) Tao, Y., & Grosky, W. I. (1998, December). Delaunay triangulation for image object indexing: A novel method for shape representation. In Electronic Imaging'99 (pp. 631-642). International Society for Optics and Photonics.

[134] (Tombari, Franchi & Di Stefano, 2013) Tombari, F., Franchi, A., & Di Stefano, L. (2013). Bold features to detect texture-less objects. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1265-1272).

[135] (Torralba, Fergus & Weiss, 2008) Torralba, A., Fergus, R., & Weiss, Y. (2008, June). Small codes and large image databases for recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008. (pp. 1-8). IEEE.

[136] (Tsai, 2012) Tsai, C. F. (2012). Bag-of-words representation in image annotation: A review. ISRN Artificial Intelligence, 2012.

[137] (Vaca-Castano & Shah, 2015) Vaca-Castano, G., & Shah, M. (2015, October). Semantic Image Search from Multiple Query Images. In Proceedings of the 23rd ACM international conference on Multimedia (pp. 887-890). ACM.

[138] (Vadivel, Sural & Majumdar, 2007) Vadivel, A., Sural, S., &Majumdar, A. K. (2007). An integrated colour and intensity co-occurrence matrix. Pattern Recognition Letters, 28(8), 974-983.

[139] (Valle & Cord, 2009) Valle, E., & Cord, M. (2009, October). Advanced techniques in CBIR: local descriptors, visual dictionaries and bags of features. In Computer Graphics and Image Processing (SIBGRAPI TUTORIALS), 2009 Tutorials of the XXII Brazilian Symposium on (pp. 72-78). IEEE.

[140] (Van De Weijer & Schmid, 2006) Van De Weijer, J., & Schmid, C. (2006, May). Coloring local feature extraction. In European Conference on Computer Vision (pp. 334-348). Springer Berlin Heidelberg.

[141] (Van De Weijer et al., 2009) Van De Weijer, J., Schmid, C., Verbeek, J., & Larlus, D. (2009). Learning colour names or real-world applications. IEEE Transactions on Image Processing, 18(7), 1512-1523.

[142] (van den Broek et al., 2008) van den Broek, E. L., Kok, T., Schouten, T. E., &Vuurpijl, L. G. (2008, February). Human-centered content-based image retrieval. In Electronic Imaging 2008 (pp. 68061L-68061L). International Society for Optics and Photonics

[143] (Vedaldi & Zisserman, 2012) Vedaldi, A., & Zisserman, A. (2012). Efficient additive kernels via explicit feature maps. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(3), 480-492.

[144] (Vedaldi et al., 2009) Vedaldi, A., Gulshan, V., Varma, M., & Zisserman, A. (2009, September). Multiple kernels for object detection. In Proceedings of 2009 IEEE 12th international Conference on Computer Vision (pp. 606-613). IEEE.

[145] (Velmurugan & Baboo, 2011) Velmurugan, K., & Baboo, L. D. S. S. (2011). Content-based image retrieval using SURF and colour moments. Global Journal of Computer Science and Technology, 11(10).

[146] (Verma, Raman & Murala, 2015) Verma, M., Raman, B., & Murala, S. (2015). Local extrema co-occurrence pattern for colour and texture image retrieval. J. Neurocomputing, 165, 255-269.

[147] (Vigo et al., 2010) Vigo, D. A. R., Khan, F. S., Van De Weijer, J., & Gevers, T. (2010, August). The impact of colour on bag-of-words based object recognition. In Proceedings of the 20th International Conference on Pattern Recognition (ICPR, 2010), (pp. 1549-1553). IEEE.

[148] (Walia & Pal, 2014) Walia, E., & Pal, A. (2014). Fusion framework for effective colour image retrieval. Journal of Visual Communication and Image Representation, 25(6), 1335-1348.

[149] (Wan et al., 2014) Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., Zhang, Y., & Li, J. (2014, November). Deep learning for content-based image retrieval: A comprehensive study. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 157-166). ACM.

[150] (Wang, Li & Wiederhold, 2001) Wang, J. Z., Li, J., & Wiederhold, G. (2001). SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(9), 947-963.

[151] (Wang et al., 2011) Wang, J., Li, Y., Zhang, Y., Wang, C., Xie, H., Chen, G., &Gao, X. (2011). Bag-of-features based medical image retrieval via multiple assignment and visual words weighting. IEEE Transactions on Medical Imaging, 30(11), 1996-2011.

[152] (Wang, Song & Elyan, 2012) Wang, L., Song, D., &Elyan, E. (2012, October). Improving bag-of-visual-words model with spatial-temporal correlation for video retrieval. In Proceedings of the 21st ACM international conference on Information and knowledge management (pp. 1303-1312). ACM.

[153] (Wang, Yu & Yang, 2011) Wang, X. Y., Yu, Y. J., & Yang, H. Y. (2011). An effective image retrieval scheme using colour, texture and shape features. Computer Standards & Interfaces, 33(1), 59-68.

[154] (Weber & Mlivoncic, 2003) Weber, R., & Mlivoncic, M. (2003, November). Efficient region-based image retrieval. In Proceedings of the twelfth international conference on Information and knowledge management (pp. 69-76). ACM.

[155] (Wei, X., Phung, S. L., & Bouzerdoum, 2016) Wei, X., Phung, S. L., & Bouzerdoum, A. (2016). Visual descriptors for scene categorization: experimental evaluation. Artificial Intelligence Review, 45(3), 333-368.

[156] (Wengert, Douze & Jégou, 2011) Wengert, C., Douze, M., & Jégou, H. (2011, November). Bag-of-colors for improved image search. In Proceedings of the 19th ACM international conference on Multimedia (pp. 1437-1440). ACM.

[157] (Winn, Criminisi & Minka, 2005) Winn, J., Criminisi, A., & Minka, T. (2005, October). Object categorization by learned universal visual dictionary. In proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 (Vol. 2, pp. 1800-1807). IEEE.

[158] (Won, 2004) Won, C. S. (2004, November). Feature extraction and evaluation using edge histogram descriptor in MPEG-7. In Proceedings of the Pacific-Rim Conference on Multimedia (pp. 583-590). Springer Berlin Heidelberg.

[159] (Won, Park & Park, 2002) Won, C. S., Park, D. K., & Park, S. J. (2002). Efficient use of MPEG-7 edge histogram descriptor. ETRI Journal, 24(1), 23-30.

[160] (Wong & Pun, 2008) Wong, C. F., & Pun, C. M. (2008, May). Content-based image retrieval based on rectangular segmentation. In M. Demiralp, W. B. Mikhael, A. A. Caballero, N. Abatzoglou, M. N. Tabrizi, R. Leandre, ... & R. S. Chroas (Eds.), WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering (No. 7). World Scientific and Engineering Academy and Society.

[161] (Xiao et al., 2010) Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010, June). Sun database: Large-scale scene recognition from abbey to zoo. In Proceedings of the 2010 IEEE

conference on Computer vision and pattern recognition (CVPR), (pp. 3485-3492). IEEE.

[162] (Xu et al., 2010) Xu, S., Fang, T., Li, D., & Wang, S. (2010). Object classification of aerial images with bag-of-visual words. IEEE Geoscience and Remote Sensing Letters, 7(2), 366-370.

[163] (Yang et al., 2007) Yang, J., Jiang, Y. G., Hauptmann, A. G., & Ngo, C. W. (2007, September). Evaluating bag-of-visual-words representations in scene classification. In Proceedings of the International Workshop on Multimedia Information Retrieval (pp. 197-206). ACM.

[164] (Yang et al., 2009) Yang, J., Yu, K., Gong, Y., & Huang, T. (2009, June). Linear spatial pyramid matching using sparse coding for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009. (pp. 1794-1801). IEEE.

[165] (Yang et al., 2008) Yang, N. C., Chang, W. H., Kuo, C. M., & Li, T. H. (2008). A fast MPEG-7 dominant colour extraction with new similarity measure for image retrieval. Journal of Visual Communication and Image Representation, 19(2), 92-105.

[166] (Yang & Newsam, 2010) Yang, Y., & Newsam, S. (2010, November). Bag-of-visual-words and spatial extensions for

land-use classification. In Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems (pp. 270-279). ACM.

[167] (Yang & Newsam, 2011) Yang, Y., & Newsam, S. (2011, November). Spatial pyramid co-occurrence for image classification. In Proceedings of the 2011 International Conference on Computer Vision (pp. 1465-1472). IEEE.

[168] (Yu et al., 2013) Yu, J., Qin, Z., Wan, T., & Zhang, X. (2013). Feature integration analysis of bag-of-features model for image retrieval. J. Neurocomputing, 120, 355-364.

[169] (Yuan et al., 2011) Yuan, X., Yu, J., Qin, Z., & Wan, T. (2011, September). A SIFT-LBP image retrieval model based on bag of features. In Proceedings of the IEEE International Conference on Image Processing.

[170] (Zand et al., 2015) Zand, M., Doraisamy, S., Halin, A. A., & Mustaffa, M. R. (2015). Texture classification and discrimination for region-based image retrieval. Journal of Visual Communication and Image Representation, 26, 305-316.

[171] (Zhang et al., 2012) 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.

[172] (Zhang & Mayo, 2010) Zhang, E., & Mayo, M. (2010, November). Improving bag-of-words model with spatial information. Proceedings of the 25th International Conference of Image and Vision Computing New Zealand (IVCNZ), (pp. 1-8). IEEE.

[173] (Zhang et al., 2010) Zhang, S., Huang, J., Huang, Y., Yu, Y., Li, H., & Metaxas, D. N. (2010, June). Automatic image annotation using group sparsity. In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (pp. 3312-3319). IEEE.

[174] (Zhang et al., 2012a) Zhang, S., Yang, M., Cour, T., Yu, K., & Metaxas, D. N. (2012). Query specific fusion for image retrieval. In Proceedings of Computer Vision–ECCV 2012 (pp. 660-673). Springer Berlin Heidelberg.

[175] (Zhang et al., 2012b) Zhang, S., Yang, M., Cour, T., Yu, K., & Metaxas, D. N. (2015). Query specific rank fusion for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(4), 803-815.

[176] (Zhao et al., 2007) Zhao, G., Chen, L., Song, J., & Chen, G. (2007, September). Large head movement tracking using sift-based registration. In Proceedings of the 15th ACM international conference on Multimedia (pp. 807-810). ACM.

[177] (Zhou et al., 2014) Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning deep features for scene recognition using places database. In Advances in neural information processing systems (pp. 487-495).

[178] https://en.wikipedia.org/wiki/Lab_color_space

[179] http://www.vlfeat.org/

[180] http://wang.ist.psu.edu/docs/related/

LIST OF PUBLICATIONS

Papers in International Journals

- E R Vimina, K Poulose Jacob, "An Evaluation of Image Matching Algorithms for Region Based Image Retrieval", International Journal of Advancements in Computing Technology, Volume 6, Number 6, pp. 75-85, November 2014.
- E R Vimina, K Poulose Jacob, "Content Based Image Retrieval Using Low Level Features of Automatically Extracted Regions of Interest", Journal of Image and Graphics, Volume 1, No.1, pp. 7-11, March 2013.
- E R Vimina, K Poulose Jacob, "A Sub-block Based Image Retrieval Using Modified Integrated Region Matching", International Journal of Computer Science Issues", Volume 10, issue 1, No. 2, pp. 686-692, 2013.
- E R Vimina, K Poulose Jacob, "CBIR Using Local and Global properties of Image Sub-blocks", International Journal of Advanced Science and Technology, Vol. 48, pp. 11-21, 2012.
- E R Vimina, K Poulose Jacob, "Enhancing Retrieval Efficiency with Image Replacement Based Relevance Feedback" (Communicated to International Journal)

- E R Vimina, K Poulose Jacob, "A Feature Replacement Based Multi-query System for Content Based Image Retrieval (Communicated to International Journal).
- E R Vimina, K Poulose Jacob, "A Multi-Cue Fusion Method Using BoVW Framework for Enhancing Image Retrieval" (Communicated to International Journal).

Book Chapters

- E R Vimina, K Poulose Jacob, "Image Retrieval Using Low Level Features of Object Regions with Application to Partially Occluded images", 17th Ibero-american Congress on pattern Recognition-CIARP 2012 (Argentina), Lecture Notes in Computer Science, Springer Berlin Heidelberg, Vol. 7441, pp. 422-429, 2012.
- E R Vimina, K Poulose Jacob, "Image Retrieval using Local Colour and Texture Features", Proceedings of ICMET 2011 (London, UK), Advances in Intelligent and Soft Computing, Springer Berlin Heidelberg, Vol. 125, pp. 767-772, 2011.

Papers in International Conferences

- E R Vimina, K Poulose Jacob, "Integrating Multiple Image Cues for Enhanced Image Retrieval in Bag of visual Words Framework", IEEE TENCON 2016 (Accepted).
- E R Vimina, K Poulose Jacob, Navya Nandakumar, "Boosting Retrieval Efficiency with Image Replacement Based Relevance

Feedback", In Proceedings of the 4th International Conference on Advances in Computing, Communications and Informatics - ICACCI-2015 (Kochi, India), IEEE Computer Society Press, pp. 2250-2255, August 2015.

- E R Vimina, K Ramakrishnan, N Nandakumar, K Poulose Jacob, "An Efficient Multi Query System for Content Based Image Retrieval Using Query Replacement", In Proceedings of the 16th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing-SNPD 2015 (Takamatsu, Japan), IEEE Computer Society Press, pp. 1-5, 2015.
- E R Vimina, K Poulose Jacob, "Image Retrieval using Local and Global Properties of Image Regions With Relevance Feedback", In Proceedings of the International Conference on Advances in Computing, Communications and Informatics-ICACCI-2012 (Chennai, India), ACM, pp. 683-689, August 2012.
- "Image retrieval using colour and texture features of Regions Of Interest", International Conference on Information Retrieval & Knowledge Management-CAMP 2012 (Kuala Lumpur, Malaysia), IEEE Computer Society Press, pp. 240-243, March 2012.