



Benha University  
Faculty of Engineering (Shoubra)  
Electrical Engineering Department

# Quantitative Image Search based on Feature Integration

A Thesis Submitted to the Faculty of Engineering "Shoubra"

Benha University

In Partial Fulfillment of the Requirements

For the M.Sc. Degree in Computer System Engineering

(Computer Systems Engineering Branch)

(Electrical Engineering Department)

*By*

***Mohamed Ahmed Ebrahim Helala***

B.Sc. in Computer Systems Engineering

**Supervised By**

**Assoc. Prof. Dr. Hala Helmy Zayed**

Faculty of Computers and Informatics  
Benha University

**Dr. Mazen Mohamed Selim**

Faculty of Engineering "Shoubra"  
Benha University

**Cairo - Egypt**

**2010**

# Acknowledgments

Praise to Allah for helping us to accomplish this work, may Allah accept our good deeds.

I wish to express my highest appreciation and gratitude to my advisors Prof. Dr. Hala Zayed and Dr. Mazen selim for their encouragement, collaboration and suggestions.

I would like to thank the current and former colleagues of computer engineering Dept., shoubra faculty of engineering, benha university for their hospitality, friendship and cooperation.

I would like to thank my father and mother for their great support and my fiance for helping me organizing the references.

An extended appreciation to Prof. Dr. Ashraf Abo Shosha for the helpful suggestions.

# Contents

<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>x</b>
<b>List of Abbreviations</b>	<b>xi</b>
<b>Abstract</b>	<b>xiv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Research Motivation . . . . .	1
1.2 Problem Statement . . . . .	3
1.3 Main Contributions . . . . .	7
1.4 Thesis Outline . . . . .	7
<b>2 Background and Related Work</b>	<b>9</b>
2.1 Text Based Image Retrieval . . . . .	9
2.1.1 Uncontrolled Vocabulary . . . . .	9
2.1.2 Controlled Vocabulary . . . . .	9
2.1.3 Assistance by Content-Based Data . . . . .	10
2.2 Content Based Image Retrieval . . . . .	10
2.3 Architecture of CBIR System . . . . .	12
2.4 Visual Content Description Techniques . . . . .	13
2.4.1 Color Feature Representation . . . . .	14
2.4.1.1 Color Space . . . . .	14

2.4.1.2	Color Histograms . . . . .	16
2.4.1.3	Color Moments . . . . .	21
2.4.1.4	Color Coherence Vector . . . . .	21
2.4.1.5	Color Correlograms . . . . .	22
2.4.1.6	Color Co-occurrence Matrix . . . . .	22
2.4.1.7	MPEG-7 Color Descriptors . . . . .	22
2.4.2	Texture Feature Representation . . . . .	24
2.4.2.1	Statistical Based Methods . . . . .	25
2.4.2.2	Model Based Methods . . . . .	25
2.4.2.3	Spectral Based Methods . . . . .	26
2.4.2.4	MPEG-7 Edge Histogram . . . . .	26
2.4.3	Shape Feature Representation . . . . .	27
2.4.3.1	Boundary-based Methods . . . . .	28
2.4.3.2	Region-based Methods . . . . .	29
2.5	Salient Feature Representation . . . . .	31
2.5.1	Wavelet Based . . . . .	33
2.5.2	Corner Based . . . . .	34
2.5.3	Scale Space Based . . . . .	35
2.5.4	Color Saliency Map . . . . .	38
2.5.5	Composite Feature Representation . . . . .	38
2.6	Similarity Measurements Techniques . . . . .	40
2.6.1	Histogram Comparison Methods (Metrics) . . . . .	40
2.6.1.1	Bhattacharyya distance . . . . .	40
2.6.1.2	Divergence factor . . . . .	40
2.6.1.3	Euclidean distance . . . . .	40
2.6.1.4	Matusita distance . . . . .	41
2.6.1.5	The histogram intersection . . . . .	41
2.6.2	Object Matching . . . . .	41
2.6.3	Region based Matching . . . . .	41
2.6.3.1	Salient Feature Matching . . . . .	42

2.6.4	Learning-Based Matching . . . . .	43
2.6.5	Similarity Measures Fusion . . . . .	43
2.7	Examples of some CBIR Systems . . . . .	43
2.7.1	QBIC (Query by Image Content) . . . . .	43
2.7.2	SIMBA (Search Images by Appearance) . . . . .	44
2.7.3	SIMPLICITY . . . . .	44
2.7.4	SMURF (Similarity-based Multimedia Retrieval Framework)	45
2.7.5	MIRROR (MPEG-7 Image Retrieval Refinement based On Relevance feedback) . . . . .	45
2.7.6	Img(Rummager) An Interactive Content Based Image Retrieval System . . . . .	45
<b>3</b>	<b>Saliency Image Retrieval Approach</b>	<b>47</b>
3.1	Introduction . . . . .	47
3.2	Salient Feature Extraction . . . . .	49
3.3	Local Description . . . . .	52
3.4	Region Spatial Relationships . . . . .	58
3.5	Similarity Measurement . . . . .	60
3.6	Experimental Work . . . . .	62
3.6.1	Standard Datasets . . . . .	62
3.6.1.1	Wang 1000 . . . . .	62
3.6.1.2	UCID . . . . .	63
3.6.2	Performance Evaluation . . . . .	63
3.6.3	Results . . . . .	65
<b>4</b>	<b>Principal Regions Detection for Efficient Retrieval</b>	<b>75</b>
4.1	Introduction . . . . .	75
4.2	Morphological Operations . . . . .	76
4.2.1	Dilation . . . . .	78
4.2.2	Erosion . . . . .	78
4.2.3	Morphological Opening . . . . .	79

4.2.4	Morphological Closing . . . . .	80
4.2.5	Morphological Opening followed by a Closing . . . . .	80
4.3	HSV Color Quantization . . . . .	81
4.4	Over-Segmentation Handling . . . . .	82
4.4.1	Region Labeling . . . . .	83
4.4.2	Region Sorting . . . . .	83
4.4.3	Region Combination . . . . .	83
4.5	Description of Segmented Regions . . . . .	84
4.6	Spatial Graph Generation . . . . .	85
4.7	Experimental Work . . . . .	86
<b>5</b>	<b>Conclusion and FutureWork</b>	<b>93</b>
<b>Publications</b>		<b>105</b>

# List of Figures

1.1	Google image search results for the keyword "basketball" . . . . .	4
1.2	CBIR query formulation methods . . . . .	5
2.1	general architecture of content-based image retrieval system. . . . .	12
2.2	Color description techniques . . . . .	14
2.3	Relating HSV to RGB color spaces . . . . .	15
2.4	Quantization of HSV color space components . . . . .	17
2.5	Generating the cummulative histogram. . . . .	18
2.6	Generated fuzzy regions for the L*,a* and b* color components. . .	19
2.7	Membership functions of the output of the fuzzy system. . . . .	19
2.8	(a) Query image 1 and (b) its fuzzy linked histogram. . . . .	21
2.9	Texture description techniques. . . . .	24
2.10	Constructing 80 bins edge histogram. . . . .	27
2.11	Shape description techniques . . . . .	27
2.12	Obtaining The same frequency spectrum of a pattern invariant to rotation at a specific scale. . . . .	30
2.13	Watershed segmentation controlled by parameters r and h. . . . .	31
2.14	Salient description techniques. . . . .	31
2.15	An image and its diagonal (D). horizontal (H) and vertical (V) wavelet coefficients. . . . .	33
2.16	Detection of salient points using the wavelet transform. . . . .	34
2.17	Rejecting detected points having low contrast (sensitive to noise) or have a small ratio of principal curvature. . . . .	36
2.18	Keypoint description using gradient magnitudes and orientations. .	37

3.1	A flowchart showing the proposed approach. . . . .	48
3.2	Calculating the saliency interest regions at different scales. . . . .	49
3.3	The need of weighting the entropy by a measure of self-dissimilarity between successive scales. . . . .	51
3.4	Gaussin Images are subtracted to produce DOG images. . . . .	51
3.5	Stability of extracted salient regions. . . . .	52
3.6	Combination of fuzzy systems to generate a single color and texture histogram. . . . .	53
3.7	Edges extraction on hue (H) channel. . . . .	54
3.8	Membership functions used to generate the 10-bin fuzzy linking histogram. . . . .	55
3.9	Membership functions of channels S (Saturation) and V (Value) used to generate the 24-bin fuzzy linking histogram. . . . .	56
3.10	Extraction of three texture features that correspond to energies in high frequency bands of wavelet transforms. . . . .	57
3.11	Membership functions used to generate 8-bin texture histogram. . .	57
3.12	A Voronoi diagram of 7 points in a plane. . . . .	59
3.13	Resulting Voronoi diagram for two images. . . . .	59
3.14	Precision and and recall comparisions of salient feature descriptors performed on Wang database. . . . .	66
3.15	Precision and and recall comparisions of FCTH, HsvHistogram and entropy scale invariant salient regions descriptors performed on Wang database. . . . .	67
3.16	Comparison of the mean precision and recall of the top N results for the proposed approach toward other approaches evaluated over Wang database. . . . .	68
3.17	Comparison of the mean precision and recall of the top N results for the proposed approach toward other approaches evaluated over UCID database. . . . .	69

3.18	Drop of precision of proposed retrieval approach due to using the actual NN graph matching technique evaluated over Wang database.	70
3.19	The mean precision for first 5 categories of Wang database. . . . .	71
3.20	The mean precision for last 5 categories of Wang database. . . . .	72
3.21	The mean recall for first 5 categories of Wang database. . . . .	73
3.22	The mean recall for last 5 categories of Wang database. . . . .	74
4.1	A flowchart showing the proposed approach. . . . .	77
4.2	The effect of dilate morphological operation using a $9 \times 9$ structuring element of value 1. . . . .	78
4.3	The effect of erode morphological operation using a $9 \times 9$ structuring element of value 1. . . . .	79
4.4	The effect of morphological opening operation using a $9 \times 9$ structuring element of value 1. . . . .	79
4.5	The effect of morphological closing operation using a $9 \times 9$ structuring element of value 1. . . . .	80
4.6	The effect of morphological opening followed by Closing using a $9 \times 9$ structuring element of value 1. . . . .	81
4.7	Applying HSV color Quantization after morphological operations. .	82
4.8	The over-segmentation problem. . . . .	82
4.9	Sorting image segmented regions based on size. . . . .	83
4.10	Controlling the number of segmented regions by combining small regions to large regions. . . . .	84
4.11	Spatial graph generation, regions and their associated descriptors are nodes in the graph. . . . .	85
4.12	Top N precision and recall comparisons evaluated over Wang database.	87
4.13	Top N precision and recall comparisons evaluated over UCID database.	88
4.14	The mean precision for first 5 categories of Wang database. . . . .	89
4.15	The mean precision for last 5 categories of Wang database. . . . .	90
4.16	The mean recall for first 5 categories of Wang database. . . . .	91

4.17 The mean recall for last 5 categories of Wang database. . . . . 92

# List of Tables

3.1 Wang database categories. . . . .	63
---------------------------------------	----

# List of Abbreviations

1D	One dimensional
2D	Two dimensional
3D	Three dimensional
ANMRR	Average normalized modified retrieval rank
AV	Audio-Visual
BDIP	Block difference of inverse probabilities
BVLC	Block variation of local correlation coefficients
CBIR	Content-Based image retrieval
CBVIR	content-based visual information retrieval
CCM	Color co-occurrence matrix
CCV	Color coherence vector
CHKM	Color histogram for K-mean
CLD	Color layout descriptor
CORR	Color autocorrellogram
CSS	Curvature scale space
DBPSP	Difference between pixels of scan pattern
DCD	Dominant color descriptor
DCT	Discrete cosine transform
DDL	Description definition language
DOG	Difference-of-Gaussian
DSs	Description schemes
EGFD	Enhanced generic fourier descriptor
EHD	Edge histogram descriptor
EM-DD	Expectation maximization-diverse density

FCTH	Fuzzy color and texture histogram
FDs	Fourier descriptors
GCM	Grayscale co-occurrence matrix
GD	Grid descriptors
GFD	Generic fourier descriptor
GGD	Generalized gaussian density
GLCM	Gray level co-occurrence matrix
GMD	Geometric moments descriptors
GoF/GoP	Group of frames/group of pictures color
HSV	Hue saturation value
ICA	Independent component analysis
ICICM	Integrated color and intensity co-occurrence Matrix
IPM	Intersection points map
IRM	Integrated region matching
KL	Kullback-Leibler
MAA	Major axis algorithm
MF	Membership functions
MPEG7	Moving picture experts group 7
MRF	Markov random field
nCCHs	Normalized color co-occurrence histograms
NN	Nearest neighbor
OPAC	Online public access catalogue
PCA	Principal component analysis
QBE	Query by example image
QBIC	Query by image content
RCWFs	Rotated complex wavelet filters
RGB	Red green blue
SCD	Scalable color descriptor
SIFT	Scale invariant feature transform
SOM	Self-Organizing map

SWT	Stationary wavelet transform
UCID	Uncompressed color image Database
WBIIS	Wavelet-Based image indexing and searching
WT	Watershed transform
ZMD	Zernike moment descriptors

# Abstract

Image search engines attempt to give access to the wide range of images available on the Internet. Many engines have been developed based on text and context strategies or on associated catalogue entries. There have been a number of attempts to build image content-based image retrieval engines to enhance image search results. Content-Based Image Retrieval (CBIR) relies on the characteristics of the image itself, for example its shapes, colors and textures to provide more accurate ranking of the results . The current approaches of CBIR differ in terms of which image features are extracted. This research is intended to enhance the precision and accuracy of CBIR systems.

This framework addresses efficient retrieval of images in large image collections. An evaluation of current CBIR techniques is provided. This evaluation concluded that image features need to be integrated to provide more accurate description of image content and better image retrieval accuracy. In this context, this thesis presents two image retrieval approaches that provide more accurate retrieval accuracy than the previous approaches.

The first proposed approach integrates techniques of salient, color and texture features. This approach extracts interest salient regions that work as local descriptors. A greedy graph matching algorithm with a proposed modified scoring function is applied to determine the final image rank. The proposed approach is appropriate for accurately retrieving images even in distortion cases such as geometric deformations and noise.

The second proposed approach discusses a new effective generalized multi-resolution region-based segmentation scheme that provides better segmentation and treats problems existed in previous region-based segmentation algorithms such as over segmentation. The new scheme is extended to provide a region based image

retrieval approach. This approach segments an image to the most general principal regions that work as local descriptors. A spatial graph is constructed from the principal regions and a greedy graph matching algorithm is applied to determine the final image rank.

The approaches are tested on standard image databases. Also a case study is developed where our approach is tested on images retrieved from Google keyword based image search engine. The results show that a combination of our approaches as local image descriptors with another global descriptor provides more accurate results than previous approaches.

The proposed retrieval approaches can be used in personal and web image collections. It also can be used as image descriptors for the purpose of classification and categorization. In addition, they can be extended to provide automatic annotations of images and story illustration.

# Chapter 1

## Introduction

This chapter presents the motivation concepts for this thesis; moreover, it introduces the concept of content based image retrieval. Finally the chapter presents the main contributions of the research work and the thesis outline.

### 1.1 Research Motivation

Recent years have seen a massive growth of the storage of digital images. These large image databases will be useless unless they are stored and accessed using efficient techniques of searching, indexing and retrieval. Generally there are three ways for searching multimedia data:

1. Free Browsing: users continue browsing a multimedia collection until finding the desired image.
2. Text Based Retrieval: users use text retrieval techniques for searching a multimedia repository. These retrieval techniques rely on the keywords around the multimedia data and the filenames. Also some retrieval techniques rely on manually annotating images with keywords describing their content.
3. Content Based Retrieval: Users search a multimedia repository using information about the content of the query multimedia object. The search process uses signatures from the actual content of the query object to search for and retrieve candidate multimedia objects.

The first two searching techniques have several drawbacks. The Free Browsing techniques is time consuming and inappropriate for large scale multimedia databases. The text based retrieval techniques rely on the keywords around the images and the filenames, which produces a lot of randomness in the search results. Also some text retrieval techniques rely on manually annotating images with keywords describing their content. This has the drawback of producing inaccurate descriptions due to differences in human perception of images being annotated. Other drawbacks rise from differences in human culture and poor description of image content.

Text based image search engines are typically web-based services that collect and index images from other sites on the internet. Image searching is offered by general search engines, like Google or Yahoo, and by specialized image search engines - services devoted to the searching of images or multimedia. In addition, there are meta- search engines, which pass on search requests to more than one search engine and then bring back the results.

Sometimes people use 'Image Search Engine' to refer to collection-based search engines, services that index a single or small number of image collections, Large digital libraries, commercial stock photo collections, or community-based collections like Flickr that typically offer their own search engine-like facilities.

In order to overcome the limitations of text based image retrieval techniques, researchers started studying methods that efficiently and accurately index huge images repository. A new research area had been constructed by researchers named Content Based Image Retrieval (CBIR). CBIR techniques tries to index visual characteristics of an image; such as its salient, color, shape and textures to search for a specific image in a huge amount of images. Salient features describe the most complex and interest points or regions in an image [1]. Color features describe the color spatial structure of an image. Texture features describe the spatial variation

---

in pixel intensities of an image; it can be described by the coarseness, contrast, directionality, line-likeness and roughness of object surfaces [2]. The shape features describe boundaries or regions of different objects in an image.

The image search domains were classified as narrow domains and broad domains [3]. This classification holds until now and remains important for the sake of system design. In narrow domain search engines, images have less variation in terms of geometric transformations, occlusion between objects and changes in illuminations; so we have well defined visual features, for example face databases. In broad domain search engines, images have high and unpredictable variations of geometric, illumination and object occlusion characteristics. This research work address the broad domain search engines.

This thesis aims to address the following issues:

1. Improve image search results retrieved from keyword based queries by integrating image features. for achieving our goal, we address the implementation of a CBIR system that retrieves images from a professional keyword based image search engine as a first filtering stag, a second stage will arrange images upon their content to a given query image.
2. Improve the precision and recall of the best ranked image list retrieved for a query. To achieve this goal, the framework addresses the feature extraction and matching techniques.

## 1.2 Problem Statement

Keyword based image search is the most intuitive image query method given by professional keyword based search engines like Google and Yahoo. For extracting keywords, they use image surrounding text. However, images frequently lack descriptive text which degrades the accuracy of text-based searching.

A keyword based image search engine uses images descriptive text to provide an image dataset that are relevant to the vocabulary of the user. But one can view this dataset as a random image dataset which needs to be reranked based on a user supplied image containing the required objects that meet the user perception. For example the following dataset (figure 1.1) is the retrieved results from Google image search engine for the keyword "basketball".

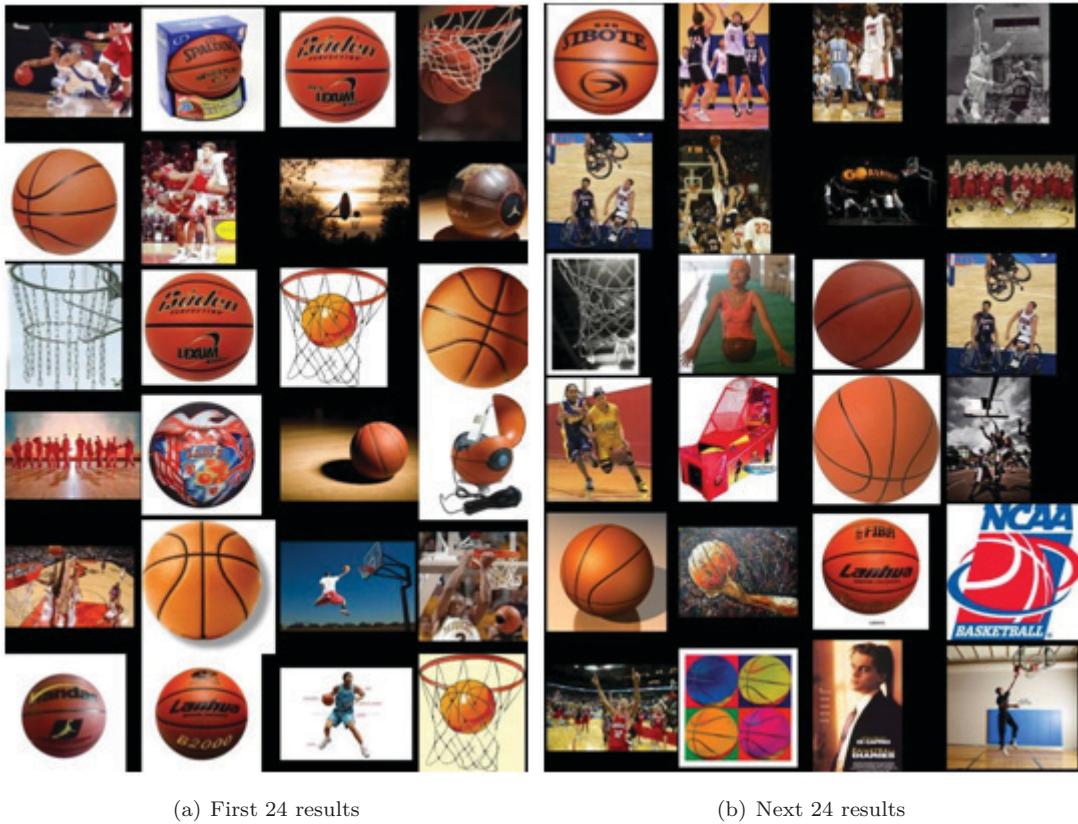


Figure 1.1: Google image search results for the keyword "basketball"

Suppose the user intention of this query is an image containing only a basketball. A user views this query retrieved results as a random retrieved set. This is because the search engine retrieves the images based on descriptive text that may be inaccurate and doesn't care about the content of these images. These random results can be used as a first web filtering stage. A second stage will be a content based image retrieval system that reranks the random dataset based on a user supplied sample image.

Several systems were designed that use text based queries as initial filtering stage, for example WebSeek [4] and ImageRover [5]. WebSeek uses both text based and color based queries through a catalogue of images and videos collected from the web. ImageRover system combines textual and visual statistics in a single index for content-based search of a WWW image database.

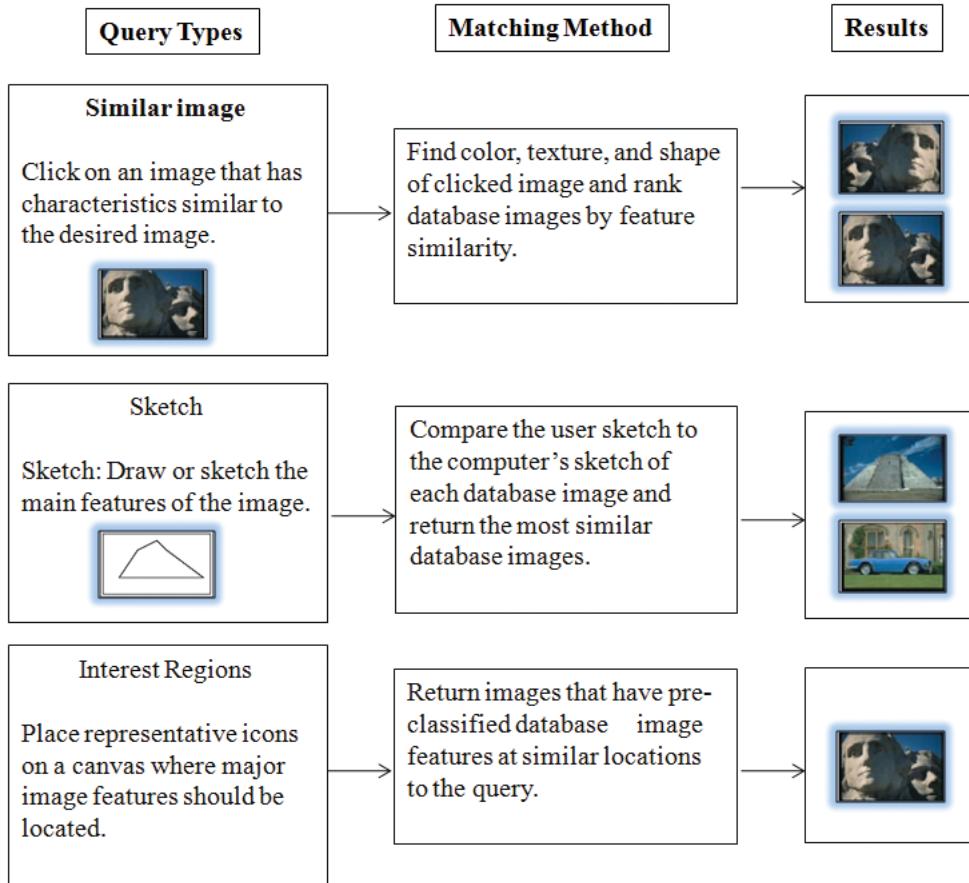


Figure 1.2: CBIR query formulation methods

In content based image retrieval, users can formulate queries using several techniques. Figure 1.2 summarizes these techniques. The most widely used technique for query formulation is the query by example image (QBE). In this technique, a user supplies a query image to search for candidate images in a huge image repository. This repository can be either an online image files collected and indexed by a web crawler, or an offline huge image database. The CBIR system index repository images using their content to speed up the retrieval process. A vector of extracted

features is used as the image content for indexing. The system extracts the same features vector from the supplied query image. A matching process matches the features of the query image to the indexed features of the database images based on the distance between feature vectors. The matched results retrieved are displayed to the user ordered by their matching rank. The existence of geometric deformations and changes in illuminations affect the quality of the features and hence the accuracy of retrieval results of the CBIR system. The need of features that are invariant to deformations and noise is critical for having better retrieval results.

The integration of several image features is critical in searching huge image repositories. We cannot rely on one image feature and ignore others. Many systems have been developed that extracts and combine several features; for example color, shape and texture features are used in QBIC system [6] for handling the retrieval problem, VIRAGE system [7] supports arbitrary combinations of these features with weights attached to each, MIRROR [8] image retrieval system Investigates MPEG7 visual descriptors [9]. A new Fuzzy Color and Texture Histogram (FCTH) was proposed by [10]. This approach combines in a single histogram both color and texture features. The authers of this approach compared it to other MPEG7 feature descriptors where the same dataset of MIRROR system was used for comparisons. Their results showed better accuracy.

The main limitation of feature integration in most existing CBIR systems is the heavy involvement of the user, who not only must select the features to be used for each individual query, but also must specify their relative weights. An interactive CBIR system needs to simplify this problem. The extraction and integration of image features should be automated to simplify the matching process for users.

The type and quality of features being integrated is critical for CBIR systems. Image features chosen for integration must be cumulative toward achieving better matching quality that meets user perception. The need to improve the precision

---

of retrieved images needs to be addressed for constructing better CBIR systems.

### 1.3 Main Contributions

1. This research implements a java based CBIR system that uses image results retrieved from Google keyword based image search engine as a web filtering stage. A second stage uses a query image to rerank the results.
2. This thesis proposes a new approach that automatically integrates techniques of salient, color and texture features. The technique is appropriate for accurately retrieving images even in distortion cases such as geometric deformations and noise. The technique is fully implemented and tested in the implemented CBIR system. The MPEG7 standard image retrieval techniques are being used for comparisons. Preparatory image databases collected by the research community for testing and benchmarking purposes are being used to show the effectiveness of our approach toward others. Experiments show that this technique provides efficient precision when dealing with large image datasets.
3. The Principal Regions Image Retrieval (PRIR) approach is proposed as a new generalized region based image retrieval technique. This proposed technique provides an efficient multi-resolution region based segmentation that extracts the general principal regions of an image, moreover it treats problems existed in previous segmentation algorithms such as over segmentation. Experiments show the effectiveness of this approach toward other approaches.
4. Several color and texture retrieval techniques are implemented and integrated for testing and comparison purposes.

### 1.4 Thesis Outline

The remainder of the thesis is organized as follows:

**Chapter 2:** presents a survey on the related work and notes about the proposed framework.

**Chapter 3:** explains the proposed saliency image retrieval approach that form a local descriptor for salient extracted regions. It also describes a modification of the greedy graph matching algorithm for similarity measurement.

**Chapter 4:** discusses the PRIR approach as a new generalized region based image retrieval technique. Moreover it describes a new automatic region based segmentation technique that extracts the general principal regions of an image and its usage in constructing an efficient image retrieval approach.

**Chapter 5:** presents the conclusions of the work and provides some important notes for the continuity of the research and future directions in this subject.

# **Chapter 2**

## **Background and Related Work**

This chapter reviews related work on image retrieval. Firstly the chapter explores the text based image retrieval methods and their related limitations. Secondly the basic concepts and the general architecture of the CBIR system are dicussed. the chapter continue with a survey of CBIR techniques. Finnaly an overview of the recent commercial and academic CBIR systems is provided.

### **2.1 Text Based Image Retrieval**

Text based image retrieval can be classified into three categories, uncontrolled vocabulary, controlled vocabulary and assistance methods.

#### **2.1.1 Uncontrolled Vocabulary**

This method describes an image using descriptive words such as caption, file name, or some background information [11]. The user can formulate queries using homogeneous descriptions such as keywords. This method can provide inaccurate results because of poor description of image content. The term, "uncontrolled vocabulary", is used to distinguish this method from other systematic methods.

#### **2.1.2 Controlled Vocabulary**

This method tries to control the number of image descriptive words by developing indexing systems [11]. "A controlled vocabulary can be built during the indexing

process as a custom made list of terms based upon prose descriptions. The terms on such a list may in turn be used to build new descriptions” (Van den Berg, 1995 [12])

### 2.1.3 Assistance by Content-Based Data

The University of California, Berkeley developed Chabot system (1995) [13]. The chabot system combined text based descriptions with content-based analysis in retrieving images from a collection of photographs. The textual information included the shooting date, the picture location, and the perspective of the photo. The content based feature was the color histogram. For matching, more than 50% of the pixels in an image must have the requested color.

## 2.2 Content Based Image Retrieval

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR). It can be defined as ”the application of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large databases” [14]. The term ”Content-based” means that the search will address the actual contents of the image. The image content can be its shape, color, texture or any other signatures that can be derived.

Smeulders et al. (2000) [15] classified the domains for image search into broad domains and narrow domains. Broad domains address searching image repositories having high variability and unpredictability in their visual characteristics. In broad domains variations in illumination, occlusion and clutter of scene objects and viewpoint changes have large impact on the content representation and search techniques. Examples of broad domains are large photo stocks and random image collections. Narrow domains on the other hand address searching image repositories having limited variability and better-defined visual characteristics such as face

databases. The Narrow domain collections are usually recorded under constant illumination using a fixed viewpoint and no occlusion.

Content based image retrieval (CBIR) is easier to be formulated for narrow domains because of the limited variability in visual characteristics. For broad domains, the need for techniques that deals with the geometric and photometric variations is critical for searching huge image repositories.

A gap [15] exists between objects in the real world and the description of a scene recorded for these objects. This gap was called the sensory gap. The sensory gap can provide uncertainty in object scene description; this can be made by variations in the scene recording conditions, such as geometric deformations, changes in illumination and existence of occlusion and clutter. The sensory gap can cause different 2D image shots of 3D objects to have the same representation due to missing information about the recording conditions.

Several levels of features can be used to represent the content of an image. Three levels of features need to be addressed [16], the low level features, the semantic features and the subjective features. The low-level features rely on the image perceptual properties such as color, texture and shape to provide a representation of image content. The semantic features bring semantic meanings into the search. They depend on the low level features, extract and represent objects and define their roles in scenes. The abstract and subjective features describe the impressions, emotions and the meaning of perceptual properties, for example we may want to retrieve pictures of a particular birthday celebration or a happy beautiful woman.

Another gap [16] exists between the feature representations of objects in images and the actual intent of a user. This is called the semantic gap. For example a user may have a specific meaning about an image. The CBIR system needs to

locate images having the same meaning depending on the feature representation of images. Current research intend to integrate high level semantics to provide better retrieval results.

The goal of content based image retrieval (CBIR) is to minimize the semantic and sensory gaps. This can be achieved by using invariant and high quality visual features of images and relevant domain knowledge to support various search categories.

## 2.3 Architecture of CBIR System

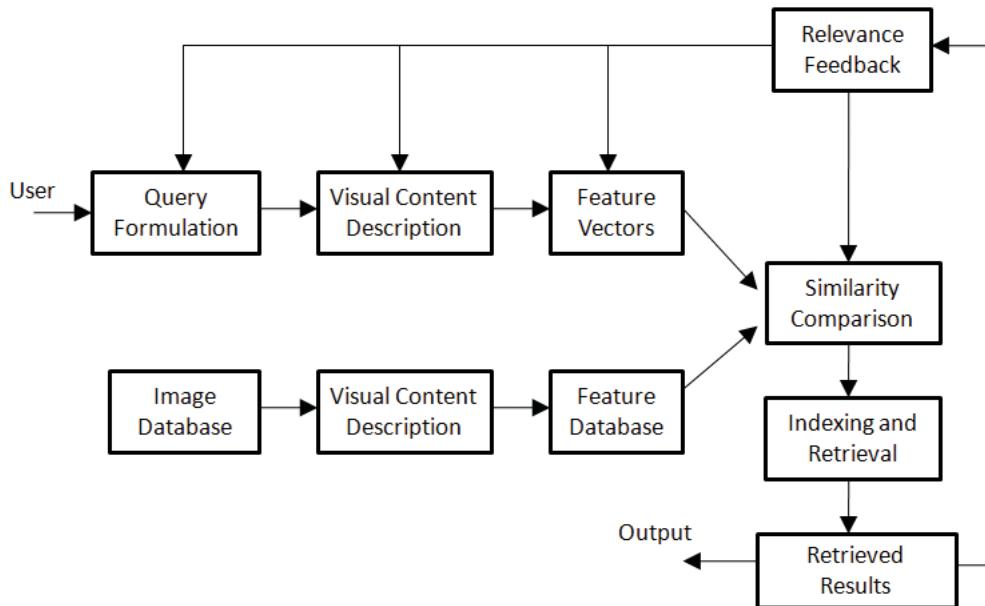


Figure 2.1: general architecture of content-based image retrieval system.

In a typical content-based image retrieval system [15] (Figure 2.1), the visual features of an image repository are represented by multi-dimensional feature vectors. These feature vectors form a feature database. To retrieve images, a user provides an example query image to the retrieval system. The system then extracts the visual features of the query image and forms a query feature vector. A similarity measure is used to compare the query feature vector and the feature vectors stored in the database. To speed up the retrieval process, an indexing scheme is

used to index feature vectors in the database. The query relevant images are fast retrieved with the aid of the indexing scheme. Recent retrieval systems include a relevance feedback stage. In this Stage, a user provides feedback by marking wrong results. This modifies the retrieval process in order to retrieve more accurate results next time.

There are several criteria that differentiate CBIR systems. These criteria are:

- The kinds of features used.
- The similarity/matching techniques used to retrieve results.
- The used indexing data structures.
- The way retrieved results presented to the user.
- The relevance feedback process used.
- Type of query formulation.

A presentation of the current research performed to enhance these criteria is provided along the next sections.

## 2.4 Visual Content Description Techniques

To better describe image content, a visual content descriptor should be invariant to variations introduced by the imaging process. However, a tradeoff exists between the invariance and the discriminative power of visual features. A descriptor that has a very wide class of invariance may lose the ability to discriminate between essential differences [2].

Visual description techniques use color, texture, shape and salient features to provide low level representation of an image. The low level representation forms a base for choosing indexing and matching techniques and selection of appropriate similarity measures.

### 2.4.1 Color Feature Representation

Color analysis is a fundamental component for representing image content. Color features are one of the most important and extensively used low-level features in visual content description and retrieval. Its popularity comes from the robustness to noise, resolution, orientation and resizing. Due to their little semantic meaning and its compact representation, color features tend to be more domain independent compared to other features. Figure 2.2 shows current techniques used to describe color features. To construct an efficient color descriptor, an appropriate color space must be selected.

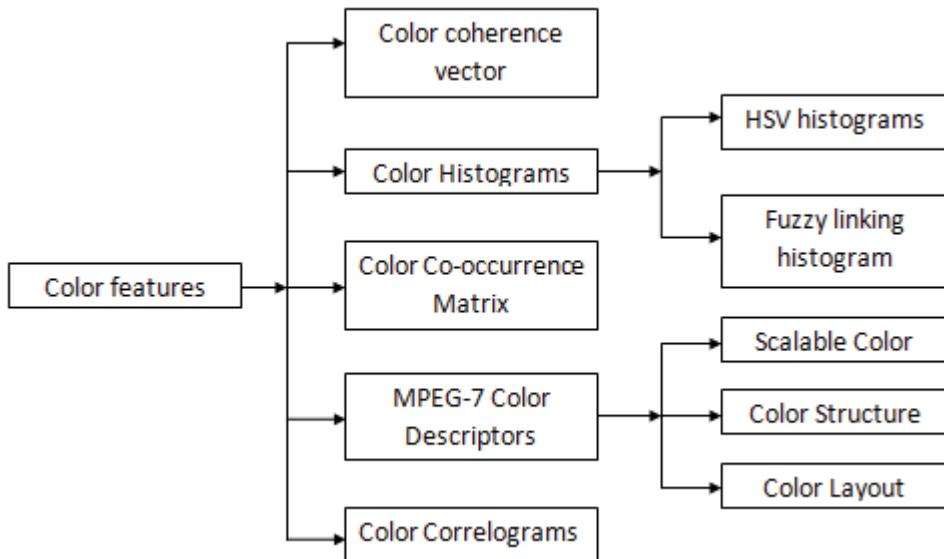


Figure 2.2: Color description techniques

#### 2.4.1.1 Color Space

A wide range of colors can be created using the primary colors (cyan (C), magenta (M), yellow (Y), and black (K)). Those colors define a space to represent other colors. The resulting color space provides a unique representation for every possible color created by combining those primary colors.

The most used color spaces are RGB, CIE L\*a\*b\*, CIE L\*u\*v\*, HSV (or HSL, HSB) and opponent color space. There is no agreement on which is the best.

However, one of the most important characteristics of an appropriate color space for image retrieval is its uniformity. "Uniformity means that two color pairs that are equal in similarity distance in a color space are perceived as equal by viewers. In other words, the measured proximity between two colors indicated by human perception must be directly related to the measured similarity distance between the two colors in a color space" [15].

The RGB space is a widely used color space for image storing and display. It is represented by three color components red, green, and blue. This space is a device-dependent space and perceptually non-uniform. This means that the perceptual difference between two RGB values does not correspond to the actual RGB space value difference.

The HSV color space has three components: hue, saturation and value. Hue is the angle of color (Figure 2.3(a)). For example, red, green and blue are just three particular angles, separated by  $120^\circ$ . Yellow is just the opposite of blue, i.e.  $240^\circ - 180^\circ = 60^\circ$ . Orange is located between Red ( $0^\circ$ ) and Yellow ( $60^\circ$ ), so Orange =  $30^\circ$ . Value determines the brightness of color (Figure 2.3(b)). All colors that match the criteria  $R + G + B = Z$  have all the same brightness value  $Z$ . Saturation measures how much a specific color is saturated by gray (Figure 2.3(c)).

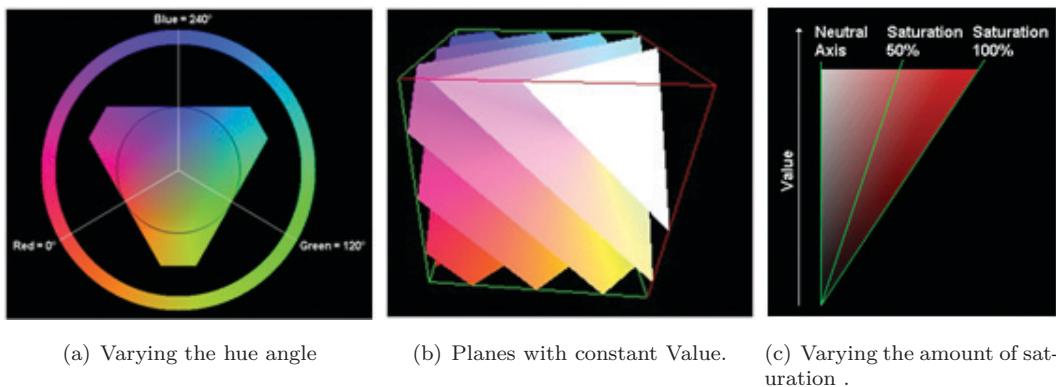


Figure 2.3: Relating HSV to RGB color spaces

The HSV color space overcomes the limitations of the RGB color space because it reflects human vision quite accurately and uses only one of its components (hue) to describe color in an image. The other two components (i.e., saturation and value) are significant only when describing black, white, gray, and the various shades of colors.

The CIE color space were specified by the international commission on illumination to match human color perception of objects. The CIE L\*a\*b\* color space had its color components as the combinations of red and yellow, red and blue, green and yellow, and green and blue. For determining the exact color combination, three components were defined to this color space,  $L^*$  the lightness coordinate,  $a^*$  the red/green coordinate and  $b^*$  the yellow/blue coordinate.

For the purpose of CBIR, the HSV and CIE LUV color spaces are the most used as they coincide better with human vision.

#### 2.4.1.2 Color Histograms

Histogram is a global statistical descriptor that represents the distribution of colors in an image. In color image processing methods, histograms usually carries the statistical information of the three components of the used color space. Swain and Ballard (1991) [17] demonstrated that color histograms of multicolored objects provide a robust efficient cue for indexing into a large database of models. They also introduced a similarity measure called histogram intersection and developed the histogram backprojection algorithm for solving the object location problem.

To build a histogram-based retrieval system, a suitable perceptually uniform color space such as HSV, CIELAB or CIELUV is required along with a feature representation such as classic or fuzzy histograms and a similarity metric such as the histogram Intersection method.

Sural, S. Gang Qian Pramanik, S. (2002) [18] analyzed the properties of the HSV (Hue, Saturation and Value) color space with emphasis on the visual perception of the variation in Hue, Saturation and Intensity values of an image pixel. Hue was given more importance as it captures intrinsic information about the color of objects or surfaces in a scene. They generated HSV histogram and compared it with the RGB histogram. Their results showed that the HSV histogram retained a uniform color transition than RGB histogram.

In constructing HSV histogram, each component of the HSV color space was quantized into a certain number of regions (figure 2.4). An implementation of HSV histogram [16] divided the hue into eight regions, whereas saturation and value were divided into four each. This created a  $(8 \times 4 \times 4) = 128$  bins histogram.



Figure 2.4: Quantization of HSV color space components

The advantages of using color histograms include:

1. Robust, since color histograms are rotation and scale invariant.
2. Histograms are straightforward to implement.
3. Fast. The histogram computation has  $O(M^2)$  complexity for an  $M \times M$  image, while a histogram comparison has  $O(n)$ , where  $n$  is the number of histogram bins, or quantization levels, of the colors used.
4. Low storage requirements. The color histogram size is much smaller than the size of the image itself.

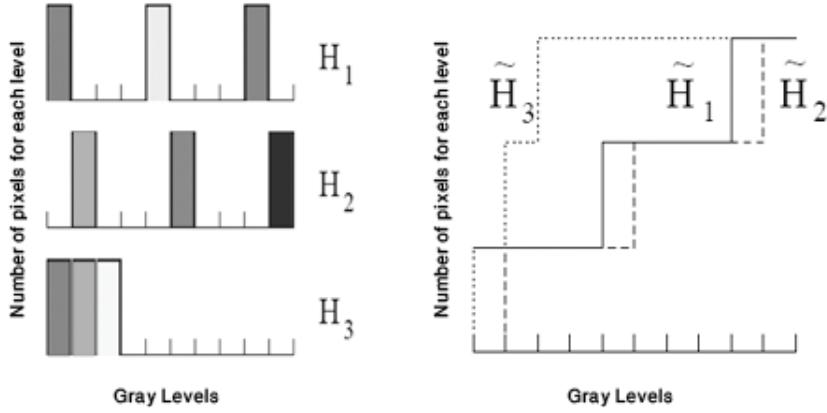


Figure 2.5: Generating the cumulative histogram.

A limitation of color histograms was that a change of illumination will shift the histogram bins which will cause difficulties in matching. Markus Stricker and Markus Orengo (1995) [19] described the cumulative color histograms (figure 2.5) and showed that the cumulative histogram is more robust to color indexing than traditional color histogram techniques. In this technique, the color composition of an image was viewed as discrete form of probability distribution or a cumulative histogram. The cumulative color histogram  $\tilde{H}(M) = (\tilde{h}_{c1}, \tilde{h}_{c2}, \dots, \tilde{h}_{cn})$  of image M that have a range of color  $\{c_1, c_2, \dots, c_n\}$  is defined as:

$$\tilde{h}_{c_j} = \sum_{c_i \leq c_j} h_{c_i} \quad (2.1)$$

Another limitation of color histograms was that perceptually similar colors may be quantized into different bins, so images which are similar to each other but have small differences in scene or contain noise will produce histograms with dissimilar adjacent bins and have large distances corresponding to each other. This problem is called the perceptually similar colors problem [20]. K. Konstantinidis and A. Gasteratos and I. Andreadis (2005) [21] solved the problem by using a small number of bins produced by a fuzzy linking histogram constructed from the L\* a\* b\* color space. They gave more weights to the a\* and b\* components than L\* component as they provided the color information of an image.

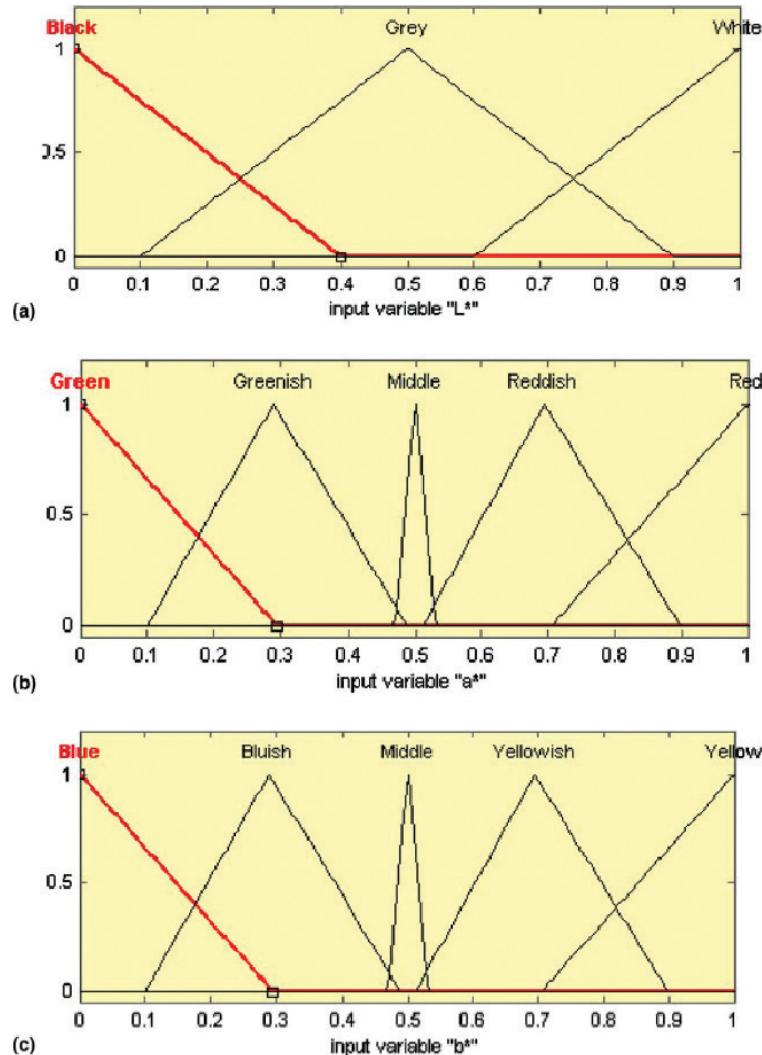


Figure 2.6: Generated fuzzy regions for the  $L^*$ ,  $a^*$  and  $b^*$  color components.

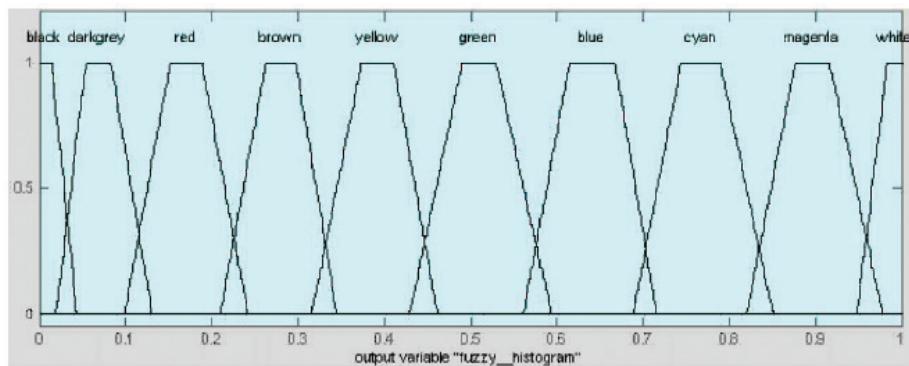


Figure 2.7: Membership functions of the output of the fuzzy system.

They selected the L\* a\* b\* color space because it was a perceptually uniform color space which approximates the way humans perceive color. They carried a number of tests and concluded that the a\* and b\* components should be subdivided into five regions representing green, greenish, the middle of the component, reddish and red for a., blue, bluish, the middle of the component, yellowish and yellow for b\*, whereas L\* should be subdivided into only three regions: dark dim and bright areas (figure 2.6). The input fuzzification was carried by using triangular shaped built-in membership functions (MF) representing the regions for the three input components (L\*, a\*, b\*). The idea of using triangular shaped regions was that a certain fully saturated color decrease in saturation as it goes to the neighboring color, thus drawing a triangular.

For each pixel in the image, the L\*, a\* and b\* values were computed. Figure 2.6 is used to calculate each component participation to the fuzzy regions. For example if the L\* component has a value 0 then it has a fully participation (=1) to the first fuzzy region of L. regions. If it has the value 0.2 then it has a partial participation to the first fuzzy region ( $\approx 0.49$ ) and a partial participation to the second fuzzy region ( $\approx 0.22$ ). All other regions have a participation of 0. This operation is applied to other components. The resulting participation values are the result of the fuzzification phase. A set of 27 rules map the participation values to the output bins using a largest of maximum defuzzification algorithm. For example a rule that said If (L\* is black), (a\* is amiddle) and (b\* is bmiddle) then (fuzzyhist is black) has the participation values to the first fuzzy region of L\*, third fuzzy region of a\* and third fuzzy region of b\* summed together. This is done for all other rules. The rule giving the largest value has its output bin incremented.

Their approach generated a 10 bins fuzzy histogram using an output 10 equally divided membership funcsions (figure 2.7). The bins of the histogram were : (1) black,(2) dark grey, (3) red, (4) brown, (5) yellow, (6) green, (7) blue, (8) cyan, (9) magenta and (10) white. Figure 2.8 shows a query image and its resulted fuzzy

linked histogram. It can be shown that bins 1, 6 and 7 are highly activated because black shadows are the dominant colors in the image. The similarity metric used was the histogram intersection.

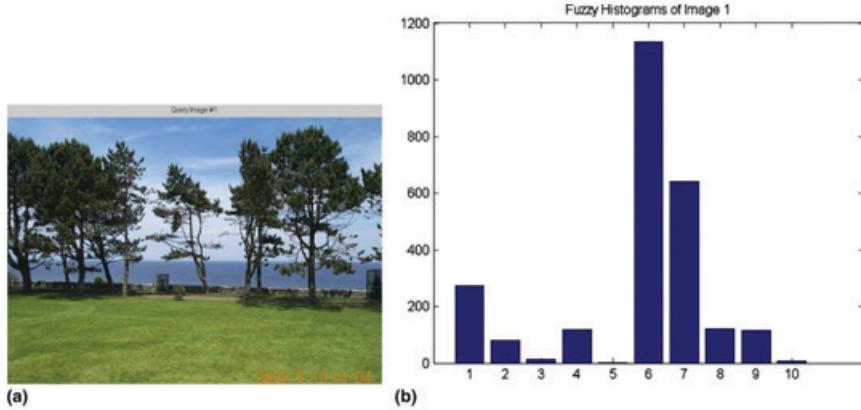


Figure 2.8: (a) Query image 1 and (b) its fuzzy linked histogram.

Further Limitations include:

1. No spatial information of the color distribution is included, so the problem of two completely different images having similar histograms may arise.
2. Immune to lighting variations.

#### 2.4.1.3 Color Moments

The color moments were described by Markus Stricker and Markus Orengo (1995) [19] where they used the first order (mean), the second (variance) and the third order (skewness) calculated over the HSV color space. Their results showed that the use of color moments outperforms the cumulative color histograms and traditional color histogram techniques.

#### 2.4.1.4 Color Coherence Vector

When an image is transformed into a histogram, all spatial information is discarded (Stricker M., Dimai A. 1996 [22]). Indexing using color histograms have significant limitations of the lack of spatial information. To solve this problem, Greg Pass Ramin Zabih, Justin Miller (1996) [23] described a new method for comparing

pairs of images that combine color histograms with spatial information. They classified each pixel in a given color bucket as coherent or incoherent based on whether or not it was part of a large similarly-colored region. A Color Coherence Vector (CCV) stored the number of coherent versus incoherent pixels with each color. They argued that their algorithm can provide histogram refinement and effective feature extraction.

#### 2.4.1.5 Color Correlograms

T. Ojala, M. Rautiainen, E. Matinmikko, M. Aittola (2001) [24] studied content based retrieval of images using color correlograms computed in HSV color space. They explored different quantizations of the HSV color space, and tried to make the correlogram more sensitive to changes in color. Results showed that the retrieval of HSV correlograms outperformed that of RGB correlograms. They concluded that HSV color space had better correspondence with human visual perception of color.

R. Barcellos and L. Lorenzi, R. Oliani, A. Gonzaga,(2005) [25] studied color autocorrelograms using HSV color space. Their results showed that the retrieval robustness to illumination condition changes.

#### 2.4.1.6 Color Co-occurrence Matrix

A. Vadivel , Shamik Sural , A.K. Majnzumdar (2007) [26] presented an approach for representing color and intensity of pixel neighborhoods in an image using a co-occurrence matrix. They analyzed the properties of the HSV color space, and suggested suitable weight functions for estimating relative contribution of color and gray levels of an image pixel. They constructed an Integrated Color and Intensity Co-occurrence Matrix (ICICM) and argued that the ICICM matrix could provide higher recall and precision compared to other methods.

#### 2.4.1.7 MPEG-7 Color Descriptors

The MPEG-7 standard [9] provided Multimedia Description Schemes (DSs) for describing and annotating visual content. The MPEG-7 visual content Descriptors

were designed mainly for describing low-level visual features. In addition, they describe higher-level visual features such as regions, segments and objects. They also produce complex descriptions by integrating together multiple descriptors and DSs and by declaring relationships among the description components. The MPEG-7 standard color descriptors includes:

- Scalable Color Descriptor (SCD)
- Color Structure Descriptor (CSD)
- Color Layout Descriptor (CLD)

The SCD was a color histogram in the HSV color space encoded by the Haar transform for storage efficiency. The size of the histogram depends on the HSV quantization Method. This descriptor started by extracting an HSV histogram with 256 bins where the hue (H) component is quantized into 16 bins and each of the saturation (S) and value (V) is quantized into 4 bins. The process of calculating the Haar transform consisted of a loop which iterates  $I = \log_2(n)$  times, where n is length of the histogram vector. This resulted in a set of 16 low pass coefficients and 249 high-pass coefficients.

The CSD was a generalization of the color histogram. It provided a color feature that captured both color content (color histogram) and the distribution structure of this content. Its main applications were image-toimage matching and mainly for still-image retrieval, where an image might consist of either a single rectangular frame or arbitrarily shaped, possibly disconnected, regions. The extraction method integrated the color structure information into the descriptor by investigating colors in a structuring element of 8x8 pixels that slide over the image.

The CLD provided a very compact and resolution invariant representation. It captured the spatial distribution of colors in an image. It proceeded by dividing the image into  $64 \times 64$  pixels blocks. For each  $8 \times 8$  block the average of Y, Cb and Cr was calculated. Finally, the 8 block was passed to a 2-D Discrete Cosine Transform

(DCT). Applications of this descriptor included Sketch-based image retrieval and content filtering.

These approaches compressed the features to provide a very compact representation of color features which make an advantage for speeding up the image retrieval process. Unfortunately the compression of these descriptors degrades the description accuracy of the features hence, the retrieval accuracy.

#### 2.4.2 Texture Feature Representation

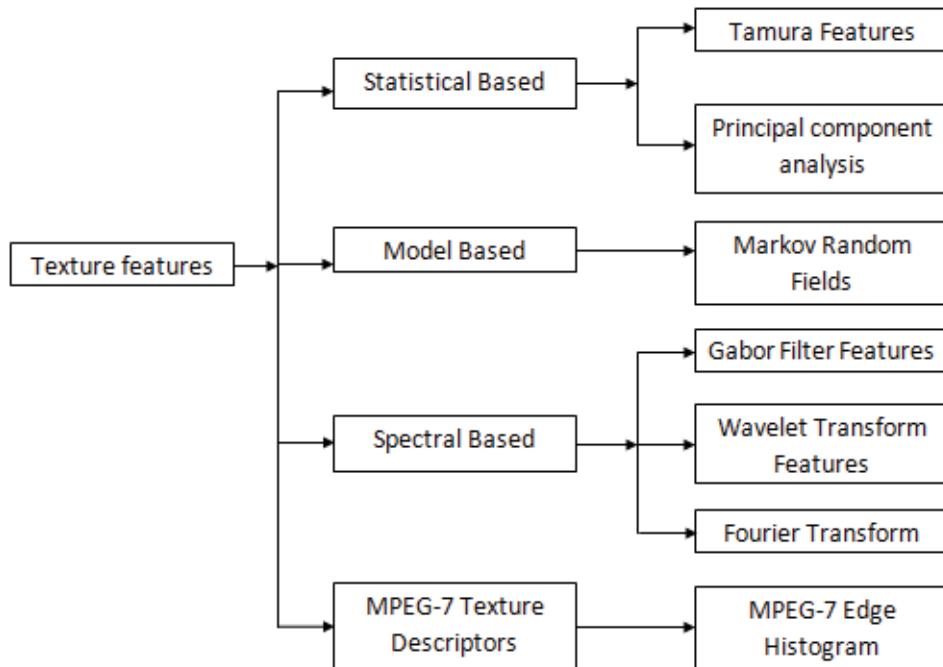


Figure 2.9: Texture description techniques.

There is no formal definition of texture. "Texture refers to the visual patterns that have properties of homogeneity that do not result from the presence of only a single color or intensity. It is an innate property of virtually all surfaces, including clouds, trees, bricks, hair, and fabric" [16].

Textures are a powerful discriminating feature but it is more complex than color. An image that has the same texture of another image can have the same

spatial arrangement of colors but may be not the same colors. Several models had been proposed to represent texture features. Figure 2.9 shows the current techniques used to describe texture features.

#### 2.4.2.1 Statistical Based Methods

These methods represent texture by the statistical distribution of the image intensity.

- **Tamura Features.** These features were originally proposed by Tamura H., Mori S., Yamawaki T. (1978) [27] and were designed according to studies of human visual perception of texture. They proposed six features, Coarseness, Contrast, Directionality, Line-likeness, Regularity and Roughness. Coarseness is a measure of the granularity of the texture and related to distances of notable variations of grey levels. Contrast measures how grey levels vary in an image and to what extent their distribution is biased to black or white. Directionality measures the Frequency distribution of oriented local edges against their directional angles.
- **Principal component analysis (PCA).** Guang-Ho Cha (2005) [28], illustrated the potential of kernel PCA for dimensionality reduction and feature extraction in content-based image retrieval. The basic idea was to first map the input space into a feature space via a nonlinear map and then compute the principal components in that feature space. They used Gaussian kernels to compute the principal components in the feature space of an image data set. Results showed that kernel PCA performs better than linear PCA in content-based image retrievals.

#### 2.4.2.2 Model Based Methods

- **Markov Random Fields (MRF).** Vacha, P. Haindl, M (2008) [29] proposed textural features that were invariant to illumination spectrum and extremely robust to illumination direction. The features required only a single training image per texture and no knowledge of illumination direction or spectrum.

Results showed the superiority of the illumination invariant features toward the local binary patterns, steerable pyramid and Gabor textural features.

#### 2.4.2.3 Spectral Based Methods

- **Gabor Filter Features.** Anil K. Jain, Farshid Farrokhnia (1991) [30] presented a texture segmentation algorithm inspired by the multi-channel filtering theory for visual information processing. The channels were characterized by a bank of Gabor filters that uniformly covered the spatial-frequency domain. They also proposed a systematic filter selection scheme based on reconstruction of the input image from the filtered images.
- **Wavelet Transform Features.** K. Muneeswaran , L. Ganesan , S. Arumugam , K. Ruba Soundar (2008) [31] introduced a method to extract the features by combining the texture discriminant features of spatial and spectral distribution of image attributes. A comparison was made with the popular Gaussian and Gabor wavelets based methods for segmenting the image. The segmented outputs and the classification efficiency of the proposed method were found to be better.
- **Fourier Transform.** Abdelhamid Abdesselam (2009) [32] described a Fourier-based technique for characterizing image textures. The performance of the technique was compared with several Fourier- and wavelet-based methods. Results showed that the new method outperformed several existing Fourier-based and wavelet-based methods.

#### 2.4.2.4 MPEG-7 Edge Histogram

The MPEG-7 edge histogram descriptor [9] described the spatial distribution of five types of edges. These edges were four directional edges (vertical, horizontal, 45°, and 135°) and one non-directional edge. Edges may be used for retrieving images with similar semantic meaning. By combining edge histogram descriptor with other Descriptors such as the color histogram descriptor, retrieval performance might be improved.

This descriptor worked by (Figure 2.10) first partitioning the image into 16 ( $4 \times 4$ ) blocks. The image partitioning always yielded 16 equal-size sub-images regardless of the size of the original image. A 5-bin histogram for edge distribution was generated from each sub image that corresponded to the directional and non-directional edge types. Since there were 16 sub-images in the image, a total of  $5 \times 16 = 80$  histogram bins was required.

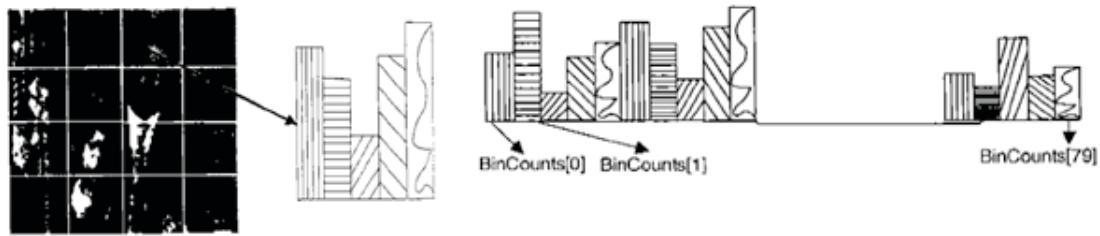


Figure 2.10: Constructing 80 bins edge histogram.

### 2.4.3 Shape Feature Representation

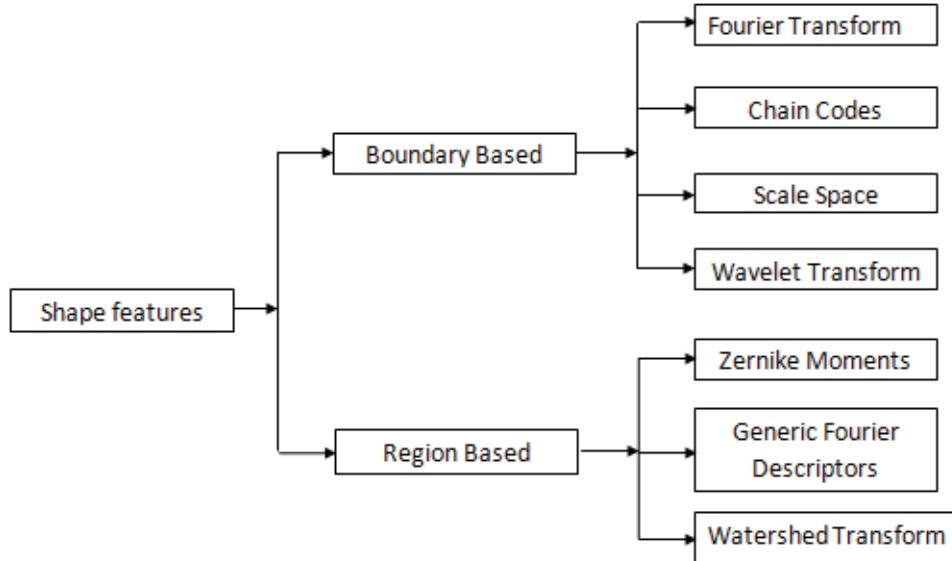


Figure 2.11: Shape description techniques

Shape features represent the geometric information of image objects or regions. The shape description methods were classified by the MPEG7 standard into either boundary-based or region-based. Region-based methods segment an image

into a number of regions and capture the distribution of region pixels. Contour-based Methods describe the shape properties of an object contour. Shape features should be invariant to geometric deformations and still identify objects even in the presence of noise and occlusion. This section survey some of the shape features extraction and description techniques that are commonly used in image retrieval applications. Figure 2.11 shows the current techniques used to describe texture features.

#### 2.4.3.1 Boundary-based Methods

- **Fourier Descriptors.** D. Zhang and G. Lu. (2001) [33] built a Java retrieval framework to compare shape retrieval using FDs derived from four shape signatures. these methods were the centroid distance, cumulative angle, complex coordinates and curvature Fourier description methods. Their results showed that shape retrieval using FDs derived from centroid distance signature was better than the other three signatures.
- **Chain Codes.** Y. K. Liu, W. Wei, P. J. Wang, and B. Zalik (2005) [34] introduced three new vertex chain codes. The first new code was introduced without increasing the average bits per code. The second introduced a variable-length vertex chain code. The third proposed a new compressed chain code based on the Huffman method. The efficiencies of the new vertex chain codes were then compared theoretically and practically against the most popular chain codes. Results showed that the new compressed chain code was the most efficient.
- **Scale Space.** A. C. Jalba, M. H. F. Wilkinson, and J. B. T. M. Roerdink (2006) [35] presented a multi-scale morphological method for the purpose of shape-based object recognition. They defined a connected operator similar to the morphological hat-transform. two scale-space representations were built using the curvature function as the underlying 1-D signal. Results showed that their method outperformed other compared methods.

- **Wavelet Descriptors.** El-Hadi Zahzah (2008) [36] proposed a method for 2D partial shape recognition under affine transform using the discrete dyadic wavelet transform. They showed how to choose the starting point on the contour using the orientation of the natural axis. The method was tested on a database of 5000 fish species, and the results were satisfactory.

#### 2.4.3.2 Region-based Methods

- **Grid Descriptors.** A. Sajjanhar, and G. Lu (1997) [37] presented an Enhanced Generic Fourier descriptor (EGFD) for object-based image retrieval. The Generic Fourier Transform (GFD) was derived by applying 2-D Fourier transform on a polar raster sampled shape image. For example, figure 2.12 shows two patterns having different orientations (figure 2.12(a) and 2.12(b)) with their resulting cartesian fourier transform. The frequencies also have diffrent alignment; resulting in matching difficulties. The corresponding polar fourier transforms (figure 2.12(c) and 2.12(d)) have the corresponding frequencies matched; allowing straightforward matching. The proposed EGFD improves GFD significantly by solving GFD's low retrieval performance on severely skewed and stretched shapes. Also an optimized Major Axis Algorithm (MAA) was proposed for finding major axis of generic shapes. Their results showed that The EGFD outperformed GFD and ZMD significantly on both perspective transform test and general distortion test.
- **Zernike Moments.** Dengsheng Zhang and Guojun Lu (2002) [38] studied and compared three region shape descriptors: Zernike moment descriptors (ZMD), grid descriptors (GD) and geometric moments descriptors (GMD). The strengths and limitations of these methods were analyzed and clarified. They implemented a Java retrieval framework to test the retrieval performance. Experiments on standard shape databases showed that ZMD was the most suitable for shape retrieval in terms of computation complexity, compact representation, robustness, hierarchical coarse to fine representation and retrieval performance.

- **Generic Fourier Descriptors.** Zhang DS, LuG, (2002) [39] presented an Enhanced Generic Fourier descriptor (EGFD) for object-based image retrieval. The Generic Fourier Transform (GFT) was derived by applying 2-D Fourier transform on a polar raster sampled shape image. For example, figure shows two patterns having different orientations (figure 2.12(a) and 2.12(b)) with their resulting cartesian fourier transform. The frequencies also have different alignment; resulting in matching difficulties. The corresponding polar fourier transforms (figure 2.12(c) and 2.12(d)) have the corresponding frequencies matched; allowing straightforward matching. The proposed EGFD improves GFT significantly by solving GFT's low retrieval performance on severely skewed and stretched shapes. Also an optimized Major Axis Algorithm (MAA) was proposed for finding major axis of generic shapes. Their results showed that The EGFD outperformed GFT and ZMD significantly on both perspective transform test and general distortion test.

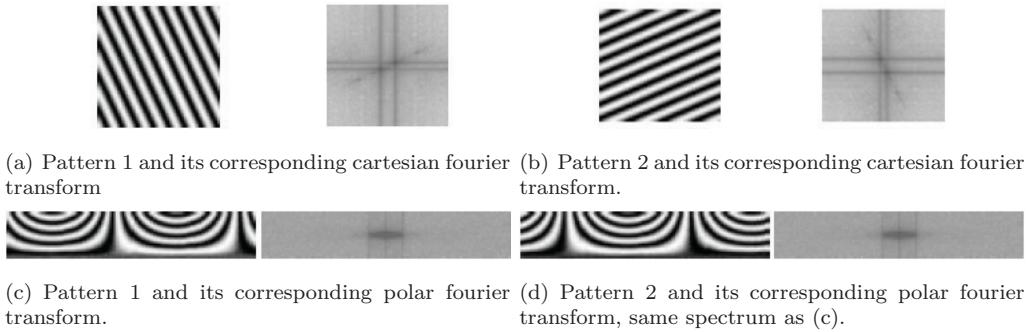
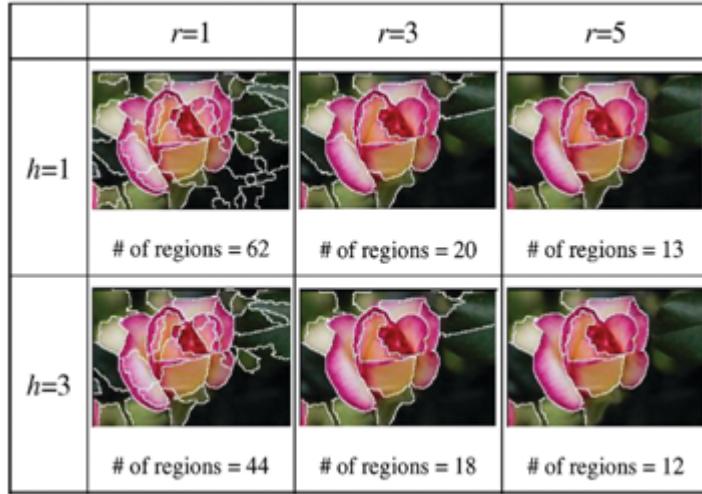


Figure 2.12: Obtaining The same frequency spectrum of a pattern invariant to rotation at a specific scale.

- **Watershed Transform.** Chiang, Cheng-Chieh and Hung, Yi-Ping and Yang, Hsuan and Lee, Greg C. (2009) [40] used the watershed transform in developing a region based image retrieval system that used the segmented regions to represent the units of an image. The watershed transform resulted in a very large number of small regions, this result made the extracted regions hardly useful. The system overcame the over segmentation problem by using two scaling parameters  $r$  and  $h$ . Parameter  $r$  was the size of the structur-

Figure 2.13: Watershed segmentation controlled by parameters  $r$  and  $h$ .

ing element of dilation operators whose application eliminates local minima of size less than  $r$  pixels. Parameter  $h$  was the height of elevation used to remove the local minima with low contrast. These two parameters were used to control the coarseness of the segmentation results. As  $r$  or  $h$  increased, the number of regions generated decreased (figure 2.13). During their evaluation, they used  $r = 1$  and  $h = 3$ . Problems with this system were that it didn't provide a generalized scheme for searching huge image repositories.

## 2.5 Salient Feature Representation

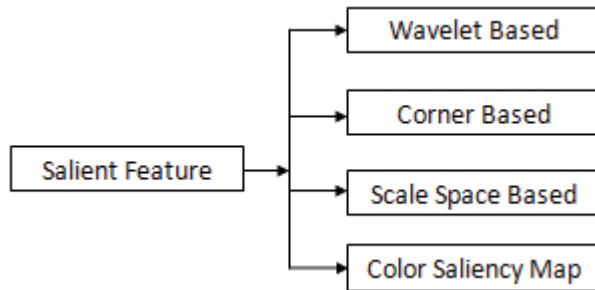


Figure 2.14: Salient description techniques.

Salient features are types of weak segmentation that captures the most salient locations or salient regions in an image. These salient locations are stored as a

reduced description of the image content and can be used for indexing and matching purposes. Global features related to color or texture usually lack the ability to capture parts of an image having different characteristics [41]. Therefore, local computation of image information is necessary.

"Strong segmentation is a division of the image data into regions in such a way that region  $T$  contains the pixels of the silhouette of object  $O$  in the real world and nothing else, specified by:  $T = O$ " [3]. Strong segmentation has many difficulties to be achieved. These difficulties can be circumvented by applying weak segmentation techniques where segmentation is targeted by signs or significant properties of the image. "Weak segmentation is a grouping of image data in conspicuous regions  $T$  internally homogeneous according to some criterion, hopefully with  $T \subset O$ " [3]. Such signs may be helpful for segmenting the image into the most discriminative salient regions.

Schiele [42] observed that certain feature locations of objects under test are more discriminating or salient than others. These discriminated locations can be used to identify an object in a scene. Salient features are types of weak segmentation that captures the most salient locations or salient regions in an image. These salient locations are stored as a reduced description of the image content and can be used for indexing and matching purposes. Salient features were used by Schmid and Mohrw [43] to build an image retrieval application that extract feature descriptors around detected salient locations.

Figure 2.14 shows several methods that had been developed to detect the most salient locations in an image such as corner detection techniques, wavelet based techniques, scale space or pyramid techniques and color saliency map.

### 2.5.1 Wavelet Based

N.Sebe , Q.Tian, E.Loupia, M.S.Lew, T.S.Huang (2001) [41] presented a wavelet-based salient point extraction algorithm. They showed that extracting color and texture information arround detected salient locations provided significantly improved results as compared to global feature approaches.

Wavelet transform is a multiresolution representation that represents image variations at different scales. The wavelet-based salient point extraction algorithm [44] tracked wavelt coefficients along several scales to detect the most salient locations (figure 2.15). A set of coefficients at a specific scale  $2^{j+1}$  were computed with

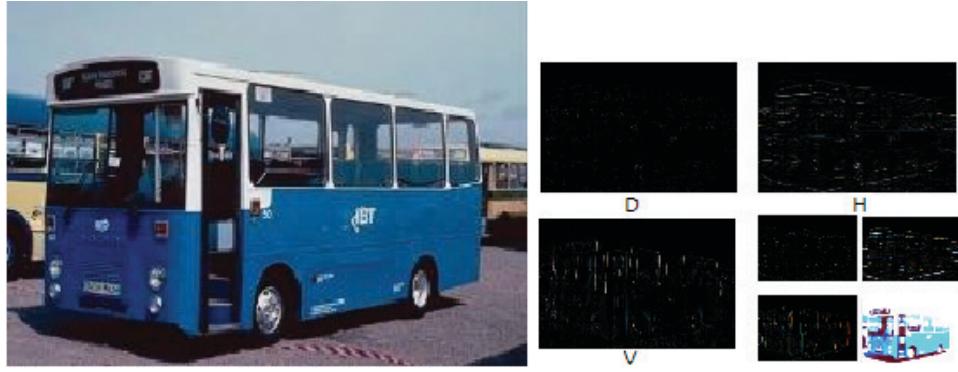


Figure 2.15: An image and its diagonal (D). horizontal (H) and vertical (V) wavelet coefficients.

the same points as the coefficients  $W_{2^j}f(n)$  of the next down scale. This set of coefficients was called the children  $C(W_{2^j}f(n))$  of the coefficients  $C(W_{2^{j+1}}f(n))$ . For each wavelet coefficient in the most down scale  $2^j$ , the highest coefficient from its children in the next scale  $2^{j+1}$  was chosen. This process continued till the finer resolution. The salient points were those having the highest coefficients along the scales. A saliency measure was used as the sum of absolute values of wavelet coefficients in the track (figure 2.16) and was defined as:

$$\text{saliency} = \sum_{k=1}^{-1} |C^k(W_{2^j}f(n))|, 0 \leq n \leq 2^j N, -\log_2 N \leq j \leq -1 \quad (2.2)$$

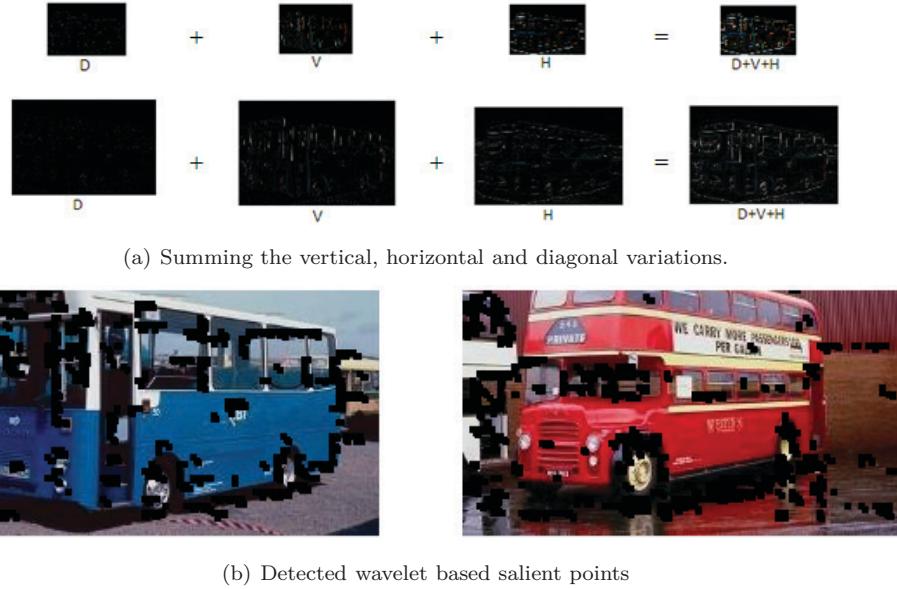


Figure 2.16: Detection of salient points using the wavelet transform.

Where  $N$  is the length of a single image dimension. The Haar wavelet was used as the simplest wavelet function for fast calculation.

### 2.5.2 Corner Based

A corner is an intersection of two lines or edges. Also a point can be a corner if it has different edge directions in its local neighborhood. Several corner detection algorithms have been developed to locate corners as salient points in images. Also regions can be defined around these points to provide short description of the image.

Dongxiang Zhou, Yun-hui Liut, and Xuanping Cai (June 2004) [45] proposed an improved Smallest Univalue Segment Assimilating Nucleus (SUSAN) corner detector and compared its performance with Harris and SUSAN corner detectors. The SUSAN detector used a circular mask over the pixel to be tested (the nucleus). The mask had a region  $M$ , a pixel within the mask is given by  $\vec{m} \in M$ . So the nucleus was at  $\vec{m}$ . The algorithm compared every pixel within the mask to the nucleus using the function:

$$c(\vec{m}) = \exp \left\{ \frac{-(I(\vec{m}) - I(\vec{m}_0))^6}{t} \right\} \quad (2.3)$$

$t$  determines the radius of the mask. The defined area of the susan was calculted as follows:

$$n(M) = \sum_{\vec{m} \in M} c(\vec{m}) \quad (2.4)$$

If  $g$  is the geometric threshold for the image, then the SUSAN response at a pixel is given by:

$$R(M) = \begin{cases} g - n(M) & \text{if } n(M) < g \\ 0 & \text{Otherwise,} \end{cases} \quad (2.5)$$

The improved SUSAN detector adopted an adaptive multi-threshold strategy based on local brightness rather than one threshold for the whole image, and divided the circular mask area of SUSAN corner detector into two or more parts. The number of pixels were calculated in each part. The exact positions of image corners were those with local minimum number of pixels.

### 2.5.3 Scale Space Based

David G Lowe, (2004) [46] proposed the Scale Invarient Feature Transform (SIFT) descriptor. This descriptor was a feature extraction scheme for matching differing images of an object or scene. The extracted features were invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint. This approach had four major stages of computation:

1. **Scale-space extrema detection:** Gaussian filters with different variance were applied to each image at different scales. The resulted Gaussian images were then subtracted to produce a difference of Gaussian pyramid. Detected interest points represented maxima and minima of the difference-of-Gaussian images.

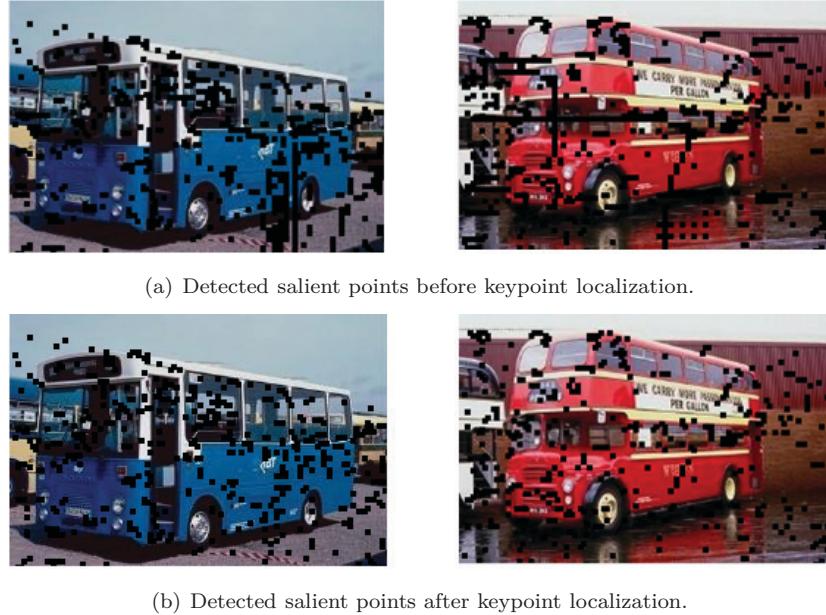


Figure 2.17: Rejecting detected points having low contrast (sensitive to noise) or have a small ratio of principal curvature.

2. **Keypoint localization:** this process reject detected points having low contrast (sensitive to noise) or small ratio of principal curvature (figure 2.17). A threshold was made on contrast and of principal curvature to reject noise or unstable points.
3. **Orientation assignment:** gradients (figure 2.18(b)) were used to assign a specific orientation to each keyposint. After assignment the keypoint descriptor could be generated relative to the assigned orientation, scale and location, so overcoming these limitations.
4. **Keypoint descriptor:** the gradient magnitude and orientation was computed in an  $8 \times 8$  region around each keypoint location. The gradient magnitude was weighted by the a Gaussian window (figure 2.18(c)) to give more importance to points near the interest keypoint than those far from it. Orientation hisograms with 8 orientations were generated (figure 2.18(d)), summarizing the content over  $4 \times 4$  subregions around the keypoint. The length value of an orientation was the sum of gradient magnitudes near that orient-

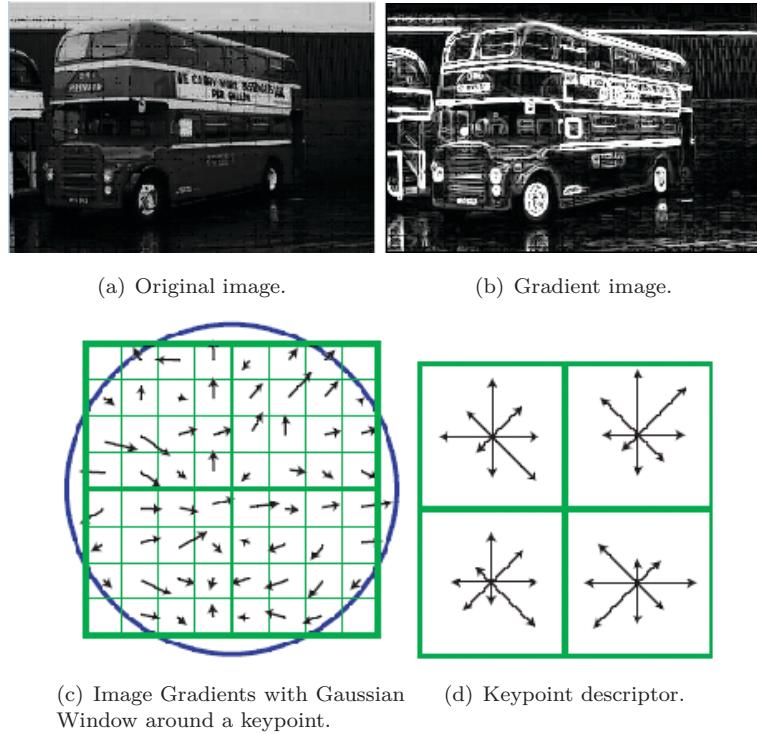


Figure 2.18: Keypoint description using gradient magnitudes and orientations.

tation within the region. So a  $2 \times 2$  descriptor array were computed for each keypoint.

Kadir, T. and Zisserman, A. and Brady, M. (2004) [47] proposed another approach that used entropy measures to identify regions of saliency within a broad range and sequences of images. The entropy represents the amount of information in a certain region so they selected those scales that had the maximum amount of information. The approach selected the most optimal scale around interest points by choosing those scales that have the maxima or peak in entropy for increasing scales. A circular function with increasing radius was used as the scaling function. The entropy was calculated for the enclosing pixels at each scale and weighted by the difference of local gray level histograms as a measure of self-dissimilarity in scale-space.

Jonathon S. Hare and Paul H. Lewis, (2004) [1] had extended the two previous approaches by using peaks in a difference-of- Gaussian pyramid to select scale-

invariant salient regions. They had shown that peaks in the difference-of-Gaussian pyramid are robust to a range of transformations and perform better than other approaches of finding salient regions based on image entropy. They also showed the concept of using salient regions for content based retrieval.

#### 2.5.4 Color Saliency Map

Joost van de Weijer, Theo Gevers, and Andrew D. Bagdano (January 2006) [48] described the color saliency boosting method where color distinctiveness is explicitly integrated in the design of salient point detectors. The method could be incorporated into existing detectors which were mostly focused on shape distinctiveness. Their results showed that color saliency boosting substantially increases the information content of the detected points.

#### 2.5.5 Composite Feature Representation

Composite feature descriptors also known as integrated feature descriptors compose more than one feature to better describe image content. Several techniques had been developed that integrate more than one feature to enhance the retrieval accuracy of CBIR systems.

Qasim Iqbal and J. K. Aggarwal (2002) [49] combined structure, color and texture features for efficient image retrieval. Structure was extracted by the application of perceptual grouping principles. Color analysis was performed by mapping all pixels in an image into a fixed color palette that uses linguistic tags to describe color content. Texture analysis was done using a bank of even-symmetric Gabor filters. Their results showed good classification in the broader image classes and limited success in sub-classification.

A. Vadivel, Shamik Sural and A. K. Majumdar. (2004) [50] proposed a soft-decision approach from the HSV color space for modeling combined human visual

perception of color and texture in a single feature vector called COLTEX. They developed a web-based system for demonstrating their work and for performing user queries. Their results showed that the proposed approach provided better recall and precision.

S. A. Chatzichristos and Y. S. Boutalis (2008) [10] proposed a composite feature descriptor that combined color and texture in a single quantized histogram. The proposed feature descriptor was named Fuzzy Color and Texture Histogram (FCTH). The fuzzy color and texture histogram was a combination of three fuzzy systems. Each system output was the input to the next system. The approach worked by initially segmenting the image into 1600 blocks as a compromise between image detail and performance. Each block had a minimum size of 4 pixels and passed through the three fuzzy systems to determine its final bin in a 192-bin quantized histogram. Their results showed better retrieval accuracy than MPEG-7 color and texture descriptors.

Chuen-Horng Lin, Rong-Tai Chen, Yung-Kuan Chan. (2009) [51] proposed three image features for image retrieval. Moreover, a feature selection technique was developed to maximize the detection rate and simplify the computation of image retrieval. The first and second image features were based on color and texture features, respectively called Color Co-occurrence Matrix (CCM) and Difference Between Pixels of Scan Pattern (DBPSP). The third image feature was based on color distribution, called Color Histogram for K-Mean (CHKM). Their results indicated better accuracy. A limitation of this approach is the use of fixed window mask for the determination of the CCM and DBPSP features which is inadequate for varying scales of objects.

## 2.6 Similarity Measurements Techniques

Content-based image retrieval measures the visual similarity between a query image and database images. The retrieval result is an images list ranked by their similarities to the query image. This section provides a review of some commonly used similarity measures.

### 2.6.1 Histogram Comparison Methods (Metrics)

The histogram comparison methods provide a similarity measure for matching images based on their extracted histogram. This section presents the most used histogram comparison methods met in literature for the purpose of image retrieval. In the next mathematical expressions,  $H_Q$  represents the query histogram  $H_C$  represents the histogram to be compared, and  $i$  represents the number of bins.

#### 2.6.1.1 Bhattacharyya distance

This distance [52] measured the statistical separability between spectral classes. It was a probabilistic distance measure that provided an estimate of the probability of correct classification and overpassed zero histogram entries. It is given by:

$$B(H_Q, H_C) = -\ln \sum_i \sqrt{H_Q(i) \times H_C(i)} \quad (2.6)$$

#### 2.6.1.2 Divergence factor

This measure computed the color distribution compactness with respect to the histograms of images. It is given by:

$$D(H_Q, H_C) = \sum_i \left[ (H_Q(i) - H_C(i)) \ln \frac{H_Q(i)}{H_C(i)} \right] \quad (2.7)$$

#### 2.6.1.3 Euclidean distance

It is one of the oldest distance measures that were used for image retrieval. it was given by:

$$L(H_Q, H_C) = \sqrt{\sum_i (H_Q(i) - H_C(i))^2} \quad (2.8)$$

#### 2.6.1.4 Matusita distance

This distance provided a probability of correct classification. It was given by:

$$M(H_Q, H_C) = \sqrt{\sum_i \left( \sqrt{H_Q(i)} - \sqrt{H_C(i)} \right)^2} \quad (2.9)$$

#### 2.6.1.5 The histogram intersection

This measure [17] was robust with respect to changes in image resolution, histogram size, occlusion, depth, and viewing point. It can be described by:

$$H(H_Q, H_C) = \frac{\sum_i \min(H_Q(i) - H_C(i))}{\min(\sum_i H_Q(i) - \sum_i H_C(i))} \quad (2.10)$$

### 2.6.2 Object Matching

Serge Belongie, Jitendra Malik and Jan Puzicha (2002) [53] presented an approach to measure similarity between shapes. The measurement of similarity was preceded by 1) solving for correspondences between points on the two shapes, 2) using the correspondences to estimate an aligning transform. The dissimilarity between two shapes was computed as a sum of matching errors between corresponding points together with a term measuring the magnitude of the aligning transform. Their results were presented for silhouettes, trademarks, handwritten digits, and the COIL data set.

### 2.6.3 Region based Matching

Jia Li, James Z. Wang, Gio Wiederhold (2000) [54] presented the Integrated Region Matching (IRM) method as a similarity measure for region-based image matching. The IRM measure incorporated properties of all regions in images by a region-matching scheme. the similarity approach reduced the influence of inaccurate segmentation,. The IRM had been implemented as a part of the SIMPLICITY

image retrieval system. Their results showed that the approach achieved more accurate retrieval.

### 2.6.3.1 Salient Feature Matching

Chiou-Ting Hsu and Ming-Chou Shih (2002) [55] presented a content-based image retrieval technique based on interest points matching and geometric hashing. They estimated points with significant luminance variations as interest points. The matching was invariant to global and local geometric transforms. They formulated a matching criterion using weighted voting technique to incorporate the spatial interrelationship. Results indicated satisfactory retrieval in the case of partial matching and geometric transformation.

Mikolajczyk, K. and Tuytelaars, T. and Schmid, C. and Zisserman, A. and Matas, J. and Schaffalitzky, F. and Kadir, T. and Gool, L. Van (2005) [56] provided a comparison of affine region detectors and used repeatability rate as a similarity measure for matching salient points or regions of two images. Repeatability is a measure of how many points or regions in one image are repeated in another image. For example suppose there is a point  $X$  that have two projection matrices  $P_1$  and  $P_2$ . The projection of  $X$  on an image  $I_1$  is given by  $p_1 = P_1 \cdot X$  and on image  $I_2$  is given by  $p_2 = P_2 \cdot X$ . A detected salient point  $p_1$  in image  $I_1$  is called repeated if point  $p_2$  is detected in the other image  $I_2$ . There must be a relation between the two points, such relation can be described by a Planar homography  $p_2 = H \cdot p_1$ . The repeatability rate for an image  $I_1$  toward another image  $I_2$  is defined as "The percentage of points that are repeated with respect to the total number of detected points" [1].

Minakshi Banerjee, MalayK.Kundu, PradiptaMaji (2009) [57] presented an image retrieval scheme using visually significant point features. The clusters of points around significant curvature regions (high, medium, and weak type) were extracted using a fuzzy set theoretic approach. The performance of the system was eval-

ated using different sets of examples from a general purpose image database. They showed the robustness of the system when images undergo different transformations.

#### 2.6.4 Learning-Based Matching

Hong Chang, Dit-Yan Yeung (2007) [58] proposed a kernel approach to improve the retrieval performance of CBIR systems by learning a distance metric based on pairwise constraints between images as supervisory information instead of manually choosing a distance function in advance. They performed experiments on two real-world image databases and showed that their method outperformed other distance learning methods due to its higher flexibility in metric learning.

#### 2.6.5 Similarity Measures Fusion

Miguel Arevalillo-Herráez, Juan Domingo, Francesc J. Ferri (2008) [59] presented a technique which allowed combining a set of distance functions into a composite measure. The technique had been compared to individual distance measures and other normalization approaches. Their results showed that the composite measure reached acceptable levels of performance and accuracy.

### 2.7 Examples of some CBIR Systems

#### 2.7.1 QBIC (Query by Image Content)

This system was developed by IBM almaden research center, san jose, CA. (1993) [6]. The system used color, texture and shape features to describe images. Color features were the 3D average color vector of an object or the whole image in RGB, YIQ, Lab, and Munsell color space and the 256-dimensional RGB color histogram. Texture features were modified versions of the coarseness, contrast, and directionality features proposed by Tamura. The shape features consisted of shape area, circularity, eccentricity, major axis orientation and a set of algebraic

moment invariants. QBIC implemented an image retrieval method based on a rough user sketch. For matching, the weighted Euclidean distance was used. QBIC applied multidimensional indexing to enhance the performance. Retrieval results presented by their order of similarity.

### 2.7.2 SIMBA (Search Images by Appearance)

This system was developed by the institute for pattern recognition and image processing, freiburg university, germany (2001) [60]. The system used a histogram of a function of transformed local pixel patterns of image taken over all rigid motions. By combining different functions and different color layers, a multidimensional histogram was built. A fuzzy version of the histogram was developed to get rid of the discontinuities of the histogram operator at bin boundaries. The user was presented with a selectable number (from 1 to 25) of random images from the database. The user selected one of these and the query would start. The system used histogram intersection as well as X<sub>2</sub> and L<sub>1</sub> and L<sub>2</sub> distances for matching. No indexing data structure was used. The results of a query were presented in decreasing order of match.

### 2.7.3 SIMPLICITY

This system was developed in the stanford university (2001) [61]. The system classified images into global semantic classes, such as textured or nontextured, graph or photograph. A series of statistical image classification methods was developed, including the graph-photograph, textured-nontextured classifiers. The application of advanced wavelets had been explored in feature extraction. Moreover, the system used a new developed image region segmentation algorithm using wavelet-based feature extraction and the k-means statistical clustering algorithm. For similarity, a new measure for the overall similarity between images was developed, called the Integrated Region Matching (IRM) measure.

### **2.7.4 SMURF (Similarity-based Multimedia Retrieval Framework)**

This system was developed by the Center for geometry, imaging, and virtual environments, institute of information and computing sciences, utrecht university, the netherlands (2002) [62]. The system incorporated color, texture, and shape features and primarily oriented toward shape-based retrieval. The search started by selecting an image randomly. Matching of two polylines was done using the weighted point set matching with a pseudometric, proportional transportation distance. In SMURF, the search was done with an optimal approximate k-nearest neighbor search in the m-dimensional "vantage space". Results were shown in decreasing order of match.

### **2.7.5 MIRROR (MPEG-7 Image Retrieval Refinement based On Relevance feedback)**

This system was developed by the department of electronic engineering, the city university of Hong Kong (2005) [63]. The system was developed for evaluating MPEG-7 visual descriptors and developing new retrieval algorithms. The system core was based on MPEG-7 Experimentation Mode (XM) with web-based user interface for query by image example. A new merged color palette approach for MPEG-7 Dominant Color Descriptor (DCD) similarity measure was developed. MIRROR consisted of three main modules. The feature Extraction module used in extracting the descriptor from images. The similarity measure module used in matching and ranking images. The relevance feedback module used in receiving user feedback on retrieved results.

### **2.7.6 Img(Rummager) An Interactive Content Based Image Retrieval System**

This system was developed by the Chalkis Institute of Technology, Greece (2009) [64]. This system contained a number of current state of the art descriptors.

The system could execute an image search query, either on XML-based index, or directly on a folder containing images. It extracts the features in real time. It also supports hybrid search of images and keywords. Several color and texture descriptors were implemented in this system. Some of the color descriptors include MPEG-7 Scalable Color Descriptor (SCD), MPEG-7 Color Layout Descriptor (CLD), MPEG-7 Dominant Color Descriptor (DCD), Fuzzy 8/10/24 HSV linking, fuzzy 10 LAB linking, auto correlograms and spatial color layout. Some of the texture descriptors include co-occurrence matrix on several histograms, laws texture, Wang's wavelet, MPEG-7 Edge Histogram Descriptor (EHD) and Tamura histogram. Also image descriptors could be saved and indexed as XML files.

# Chapter 3

## Saliency Image Retrieval Approach

### 3.1 Introduction

This chapter discusses a new proposed technique that forms a local descriptor for salient extracted regions. Moreover, the chapter describes a modification of the greedy graph matching algorithm [65] for similarity measurement. A flow chart that describes the approach is shown in figure 3.1. The approach begins with the extraction of the most stable salient regions that have the most complexity and information. The extraction of scale affine salient regions was proposed by ([66], [47]). The work is further extended by [1], where the peaks in the difference-of-Gaussian pyramid were used as starting points for the extraction of salient regions. The difference-of-Gaussian approximates the scale normalized Laplacian-of-Gaussian [46]. A comparison of affine region detectors was performed by [56] and showed that the most stable interest points can be produced from the minima and maxima of the difference-of-Gaussian pyramid compared to a range of other operators. The process is described in section 3.2.

After extracting salient regions, the 192 bin fuzzy color and texture histogram is used as a local descriptor for each region. This descriptor was chosen by experiments with various descriptors. The process is described in section 3.3.

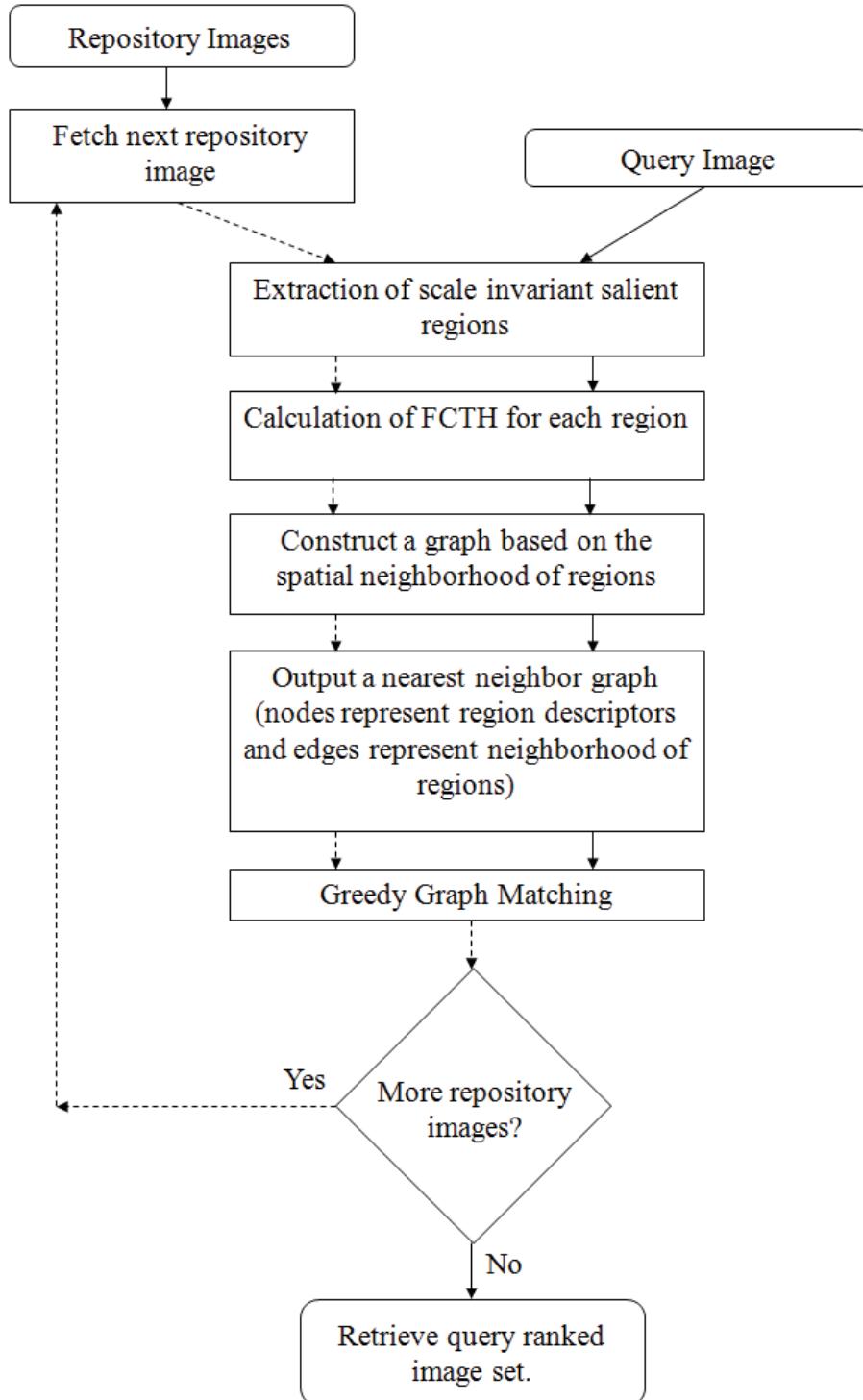


Figure 3.1: A flowchart showing the proposed approach.

A query image needs to be compared with images dataset to determine the most relevant images. The idea of repeatability rate was adopted by ([1], [67], and [68]) as a similarity measurement for ranking images where points in one image are related to points in a second image by a planar homography. This approach ignores the spatial relationships between points. This research have investigated the spatial relationship between extracted salient regions using graph matching algorithms where a Voronoi diagram [69] is constructed for each image using regions as sites. A graph is constructed from the generated Voronoi diagram in which each region and its associated descriptor is a node in the graph. Two nodes in the graph have an edge if their associated nodes are neighbors in the Voronoi diagram. The process is described in section 3.4.

A similarity measurement based on a proposed modified scoring function of the greedy graph matching algorithm adopted from [65] is described in section 3.5.

## 3.2 Salient Feature Extraction

An image salient regions extraction scheme was proposed by Kadir and Brady [47], Where regions were extracted by maximizing entropy versus scale. This approach used entropy measures to identify regions of saliency within a broad range and sequences of images. The entropy represents the amount of information

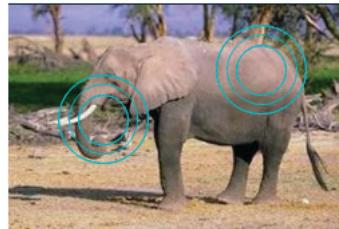


Figure 3.2: Calculating the saliency interest regions at different scales.

in a certain region so they selected those scales that had the maximum amount of information. The approach scaned the image pixel by pixel. At each pixel, a circular function (figure 3.2) with increasing radius was used as the scaling

function. The most optimal scale was selected that have the maxima or peak in entropy for increasing scales. The entropy was calculated for the enclosing pixels at each scale and weighted by the difference of local gray level histograms as a measure of self-dissimilarity in scale-space. The Kadir and Brady approach proposed a saliency metric  $y$  as a function of scale  $s$  and position  $\vec{x}$ :

$$Y_D(\vec{s}, \vec{x}) \triangleq H_D(\vec{s}, \vec{x}) \times W_D(\vec{s}, \vec{x}) \quad (3.1)$$

$$H_D(\vec{s}, \vec{x}) \triangleq \int_{i \in D} P_D \log_2 P_D(s, \vec{x}) \cdot di \quad (3.2)$$

Where  $P_D(s, \vec{x})$  is the probability density as a function of scale  $s$ , position  $\vec{x}$  and descriptor value  $i$  which takes on values in  $D$  that is the set of all descriptor values. The weighting function  $W_D(\vec{s}, \vec{x})$  is defined by:

$$W_D(\vec{s}, \vec{x}) \triangleq s \cdot \int_{i \in D} \left| \frac{\delta}{\delta x} P_D(s, \vec{x}) \right| \cdot di \quad (3.3)$$

The vector of scales at which entropy is peaked is defined by:

$$\vec{S} \triangleq \left\{ s : \frac{\partial^2 H_D(s, \vec{x})}{\partial s^2} < 0 \right\} \quad (3.4)$$

A certain saliency threshold is defined for the whole image. Scales having values above the threshold were selected as the most salient regions in the image.

The weighting of the measured entropy by a self-dissimilarity measure was done because a situation may occur that the most salient scale around each pixel is that of higher entropy alone (figure 3.3). When selecting the most salient regions for allover the image, they will be those with the most high entropy. In figure 3.3 according to this situation the first pixel has the most salient region of value 7 and the second pixel has the most salient region of value 14. So the second region is more salient than the first. It can be shown that the difference between entropies of variable scales around each pixel show a smooth transition around the second

pixel and a rough transition around the first. As mentioned saliency reflects complexity and the first salient region is more complex than the second region. By weighting the entropy by a measure of self-dissimilarity between successive scales, the first region is marked more salient than the second region.

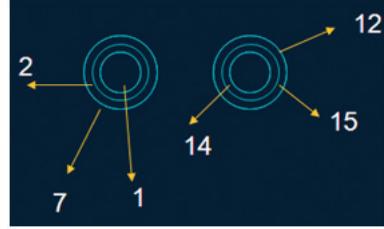


Figure 3.3: The need of weighting the entropy by a measure of self-dissimilarity between successive scales.

The problem with the previous approach was that entropy was being very sensitive to noise, which produced instability in detected salient regions. The proposed saliency retrieval approach extracts the most stable salient regions based on the technique of Jonathon and Lewis [1]. In this technique each image is repeatedly convolved, with Gaussians to produce a set of Gaussian blurred images. The resulted Gaussian images were then subtracted to produce a difference of Gaussian pyramid (figure 3.4(a)). Detected interest points represented maxima and minima

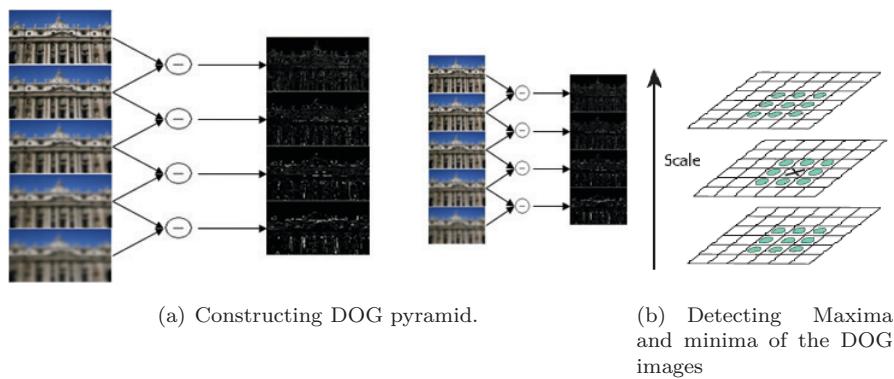


Figure 3.4: Gaussin Images are subtracted to produce DOG images.

of the difference-of-Gaussian images. To get these points, each pixel (marked with X) in the difference-of-Gaussian images was compared to its 26 neighbors (figure

3.4(b)) in 3x3 regions at the current, next and previous adjacent scales (marked with circles).

Peaks detected in the difference-of-Gaussian pyramid are used as starting points for the detection of salient regions; this has the effect of generating stable salient regions even in the presence of geometric deformations. Figure 3.5 shows two images taken for the same building but with different view point. Figure 3.5(b) shows occlusion with other objects. The stability of the extracted salient regions for the two images at different scales can be shown.

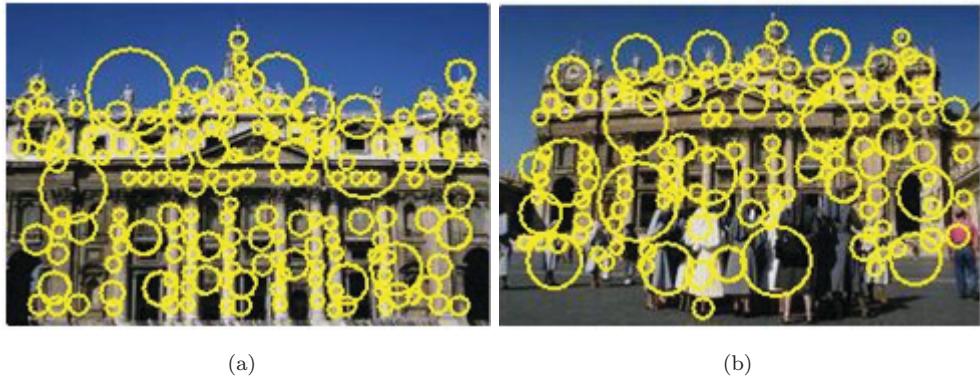


Figure 3.5: Stability of extracted salient regions.

### 3.3 Local Description

For choosing appropriate descriptor for the extracted regions all parameters of what is being salient and non-salient within a region should be captured in the descriptor. Many descriptors have been developed that extract one image feature and other approaches combine several natures of visual features to improve description. The need is to extract a Local invariant descriptor that is robust with respect to occlusion and geometrical transformations.

S. A. Chatzichristos and Y. S. Boutalis [10] proposed a composite feature descriptor called the Fuzzy Color and Texture Histogram (FCTH). This descriptor

combined color and texture in a single histogram. The effectiveness of this method was verified by experimental work (section 3.6) compared against other descriptors.

A 192-bin histogram is generated for each extracted salient region that works as its local descriptor. Each image is described by a set of salient regions distributed on its layout and each region will be associated with a 192-bin histogram as a local region descriptor. This set of regions works as a reduced description of the entire image.

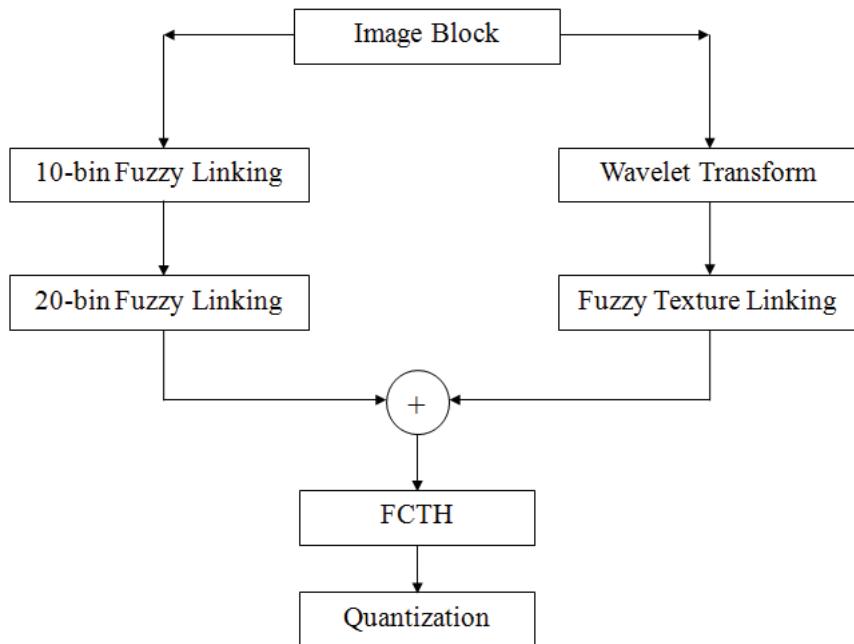


Figure 3.6: Combination of fuzzy systems to generate a single color and texture histogram.

The fuzzy color and texture histogram was a combination of three fuzzy systems. Each system output was the input to the next system (figure 3.6). The approach worked by initially segmenting the image into 1600 blocks as a compromise between image detail and performance. Each block had a minimum size of  $4 \times 4$  pixels and passed through the three fuzzy systems to determine its final bin in a 192-bin quantized histogram.

The first fuzzy system undertakes the extraction of a 10-bin fuzzy-linking histogram [21]. This histogram was formed from the three channels of HSV color

space as input. Each bin in the output histogram represented a specific color as follows (0) Black, (1) Gray, (2) White, (3) Red, (4) Orange, (5) Yellow, (6) Green, (7) Cyan, (8) Blue and (9) Magenta. Each channel of the HSV color space were separated in a set of fuzzy regions that form the fuzzification phase of the first fuzzy system. The limits of these regions were deduced from the vertical edges of images that represent the channels H (Hue), S (Saturation) and V (Value). For example, figure 3.7 shows the vertical edges of the channel H. These edges were detected by using a coordinate logic filters (CLF) edge detection algorithm and were used for determining the position of 8 membership regions of the hue channel.

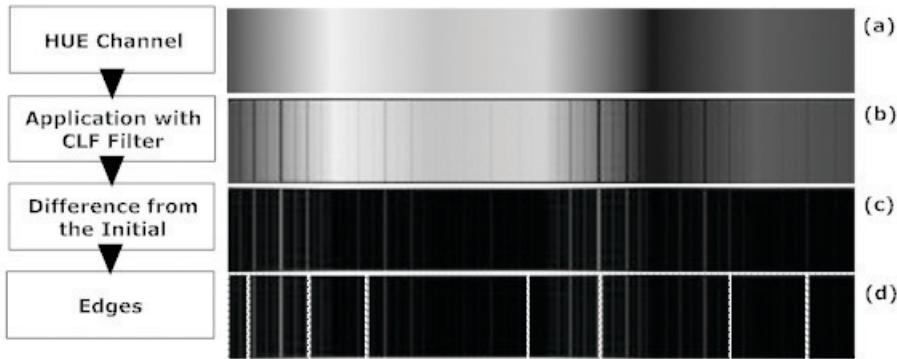


Figure 3.7: Edges extraction on hue (H) channel.

In this fuzzy system the Hue channel was separated into 8 fuzzy areas (figure 3.8(a)). These regions correspond to (0) Red to Orange, (1) Orange, (2) Yellow, (3) Green, (4) Cyan, (5) Blue, (6) Magenta and (7) Blue to Red. The saturation channel was divided into 2 regions (figure 3.8(b)). These two regions defined the shade of color based on white. The first region defined that the color was a shade of gray. The second defined if the color was a clear color from the hue channel. The value channel was divided into 3 fuzzy regions (figure 3.8(c)). The first region defined black irrespective of the two other channels. The second region combined with saturation regions defined gray. The third was for other colors.

For each segmented block, the mean color was calculated and the H, S and V values were computed. Figure 3.8(a) is used to calculate the H component partic-

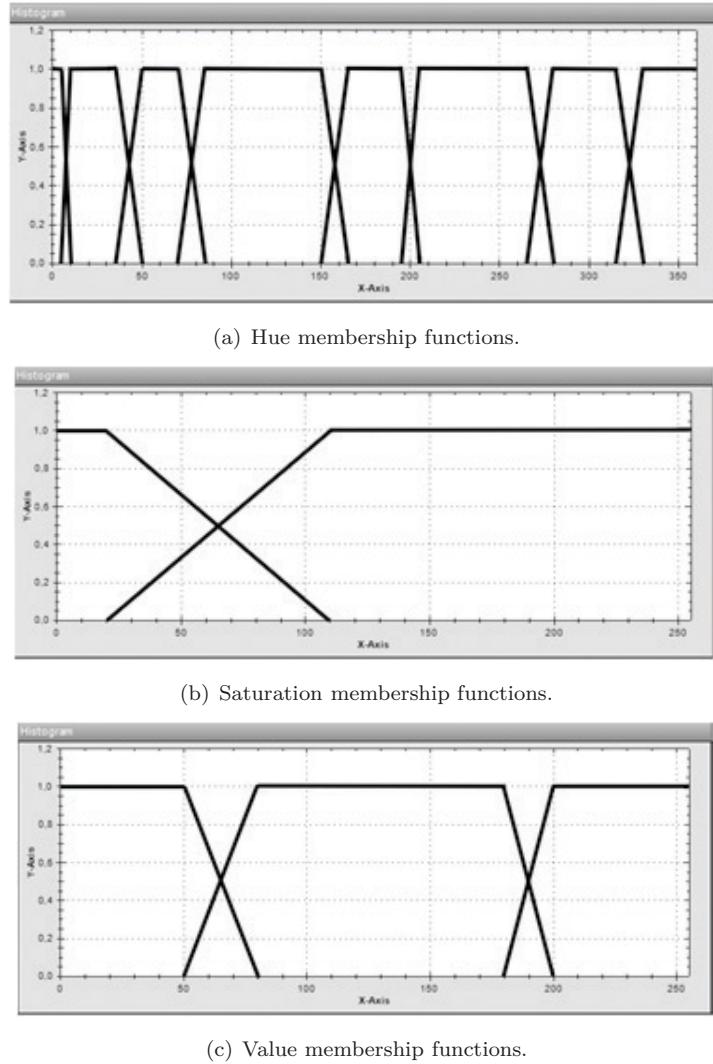


Figure 3.8: Membership functions used to generate the 10-bin fuzzy linking histogram.

ipation to the fuzzy regions. For example if the H component has a value 0 then it has a fully participation of 1 to the first fuzzy region of H (the red to orange region). If it has the value 40 then it has a partial participation to the second fuzzy region (orange region) of  $\approx 0.47$  and a partial participation to the third fuzzy region (yellow region) of  $\approx 0.22$ . All other regions have a participation of 0. This operation is applied to other components to determine their participation. The resulting participation values are the result of the fuzzification phase. A set of 20 rules map the participation values to the output bins using a largest of maximum defuzzification algorithm. For example a rule that said If H is Orange, S is bright and V is black

then output is black has the participation values to the second fuzzy region of H, second fuzzy region of S and first fuzzy region of V summed together. This is done for all other rules. The rule giving the largest value has its output bin incremented.

The output 10-bin histogram worked as input to a second fuzzy system that separates each bin color in 3 hues. This system shaped a 24 bins histogram as an output. Each bin in this histogram represented a specific color as follows : (0) Black, (1) Grey, (2) White, (3) Dark Red, (4) Red, (5) Light Red, (6) Dark Orange, (7) Orange, (8) Light Orange, (9) Dark Yellow, (10) Yellow, (11) Light Yellow, (12) Dark Green, (13) Green, (14) Light Green, (15) Dark Cyan, (16) Cyan, (17) Light Cyan, (18) Dark Blue, (19) Blue, (20) Light Blue, (21) Dark Magenta, (22) Magenta, (23) Light Magenta. The values of Saturation and Value from each block as well as the bin values of the 10-bin histogram worked as inputs to this 24-bin fuzzy linking histogram. Channel S and V were divided into 2 regions (figure 3.9).

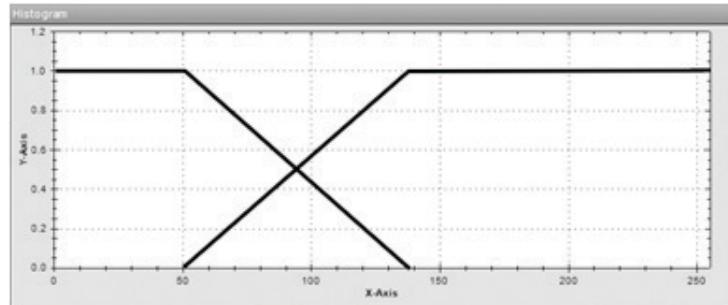


Figure 3.9: Membership functions of channels S (Saturation) and V (Value) used to generate the 24-bin fuzzy linking histogram.

For incorporating texture information, the Haar wavelet transform was applied to the region block and the HH, LH and HL features that represent energies in high frequency bands of wavelet transforms were used. These features were the square root of the second order moment of wavelet coefficients in high frequency bands. Figure 3.10 shows an example for the extraction of each feature.

Using the minimum block size  $4 \times 4$ , a one-level wavelet transform decomposed each block into four frequency bands (Figure 3.10). Each band contains  $2 \times 2$  coeffi-

cients. For example the coefficients in the HL band were  $\{C_{k,l}, C_{k,l+1}, C_{k+1,l}, C_{k+1,l+1}\}$ . One feature was then computed as:

$$f = \left( \frac{1}{4} \sum_{i=0}^1 \sum_{j=0}^1 C_{k+i,l+j}^2 \right)^{\frac{1}{2}} \quad (3.5)$$

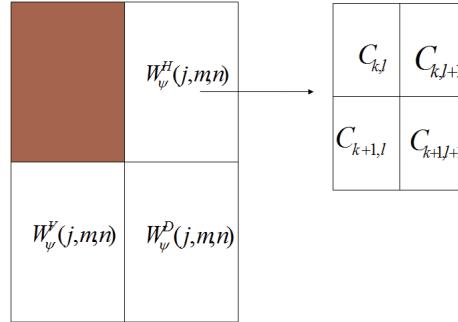
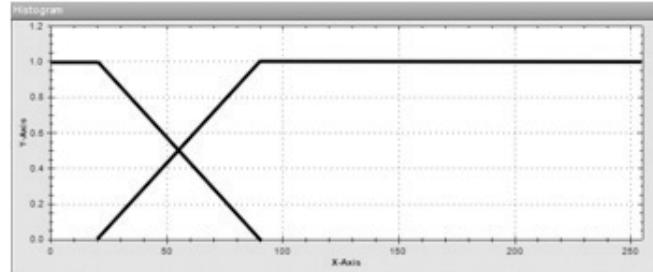
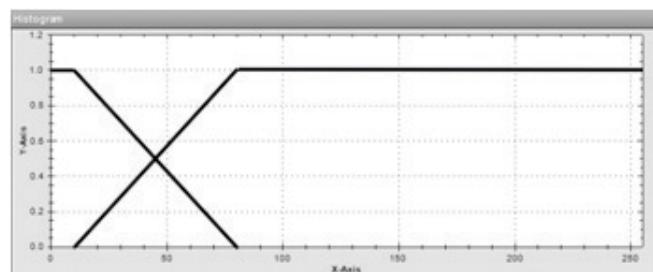


Figure 3.10: Extraction of three texture features that correspond to energies in high frequency bands of wavelet transforms.



(a) The membership functions of  $f_{HL}$  and  $f_{LH}$ .



(b) The membership functions of  $f_{HH}$ .

Figure 3.11: Membership functions used to generate 8-bin texture histogram.

This procedure was applied to the other HH, LH bands to extract three features  $f_{HL}$ ,  $f_{LH}$  and  $f_{HH}$  for each block. These 3 features were used as input to a third fuzzy system which shapes a histogram of 8 areas as output. These areas corre-

sponded to (0) Low Energy Linear area, (1) Low Energy Horizontal activation, (2) Low Energy Vertical activation, (3) Low Energy Horizontal and Vertical activation, (4) High Energy Linear area, (5) High Energy Horizontal activation, (6) High Energy Vertical activation, (7) High Energy Horizontal and Vertical activation.

The  $f_{HL}$  and  $f_{LH}$  were divided into 2 fuzzy areas as shown in figure 3.11(a) and  $f_{HH}$  was divided into 2 fuzzy areas as shown in figure 3.11(b) A set of 8 rules were applied to map the 3 inputs to a bin in the texture histogram.

The 192-bin Fuzzy color and texture histogram resulted from the combination of the 24-bin fuzzy linking histogram and the 8-bin texture histogram. The combination was done by multiplying each bin of the 8-bin histogram by the bins of the 24-bin linking histograms. So the resulted histogram was a total of  $24 \times 8 = 192$  bins. A quantization of this histogram was done to reduce the size of the histogram for efficient matching.

## 3.4 Region Spatial Relationships

To determine the spatial relationships between extracted regions, the idea of Voronoi diagram from computational geometry has been incorporated [69]. Given a set of points, the Voronoi diagram partition the plane so that each point is associated with the region of the plane containing points closest to it as shown in Figure 3.12.

Let  $S$  denote a set of  $n$  points (called sites) in the plane. For two distinct sites  $p, q \in S$ , "the dominance of  $p$  over  $q$  is defined as the subset of the plane being close to  $p$  as to  $q$ " [69]. Formally,

$$dom(p, q) = \{x \in R^2 | \delta(x, p) < \delta(x, q)\} \quad (3.6)$$

Where  $\delta$  is the Euclidean distance function. Obviously  $dom(p, q)$  is a closed half

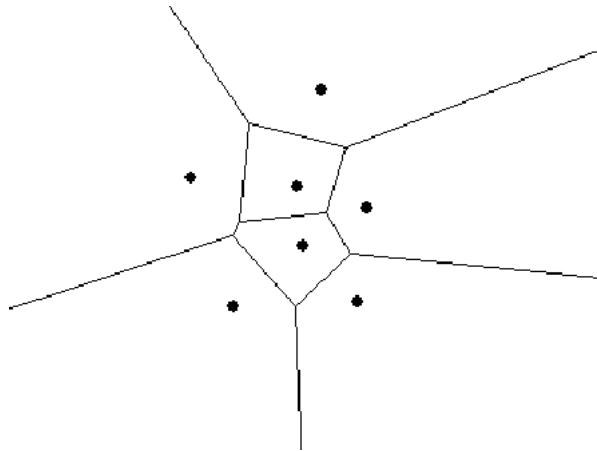


Figure 3.12: A Voronoi diagram of 7 points in a plane.

plane bounded by the perpendicular bisector of  $p$  and  $q$ . This bisector separates all points of plane close to  $p$  from those points close to  $q$ . "The region of a site  $q \in S$  is the portion of the plane containing all of the dominances of  $p$  over the remaining sites in  $S$ " [69]. Formally,

$$reg(p) = \bigcap_{q \in S - \{p\}} dom(p, q) \quad (3.7)$$

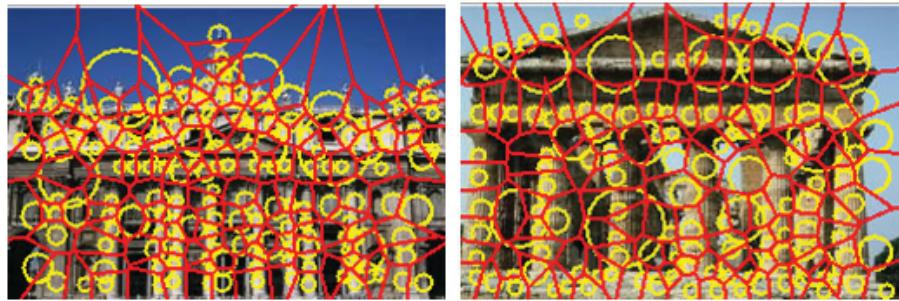


Figure 3.13: Resulting Voronoi diagram for two images.

Regions are convex polygons, because they result from the intersection of half planes. The Voronoi diagram is the combination of the resulting regions. Delaunay triangulation goes hand in hand with Voronoi diagram because each edge of the resulted triangle connects two sites in the plane if and only if their Voronoi regions share a common edge. So for a site, neighboring sites can be traversed using direct edges from Delaunay triangulation.

A Voronoi diagram is generated for each image for determining the spatial relationship of the extracted salient regions. The center point of each salient region is used as a site in the Voronoi diagram. This determines the spatial relationships of regions. Figure 3.13 shows the generated voronoi diagram for two images by using extracted regions center points as sites.

## 3.5 Similarity Measurement

A graph is a very powerful representation to represent structural properties and can be used to provide a representation of an image for the purpose of matching. To use graph matching as a similarity measure, Dasigi and Jawahar [65] indicated that the need to "(1) Find a representation that best discriminates two graphical structures. (2) Design a matching process that provides efficient matching of these representations." The proposed approach dealt with the first issue where salient regions, each with FCTH descriptor have been extracted and a Voronoi diagram describing the spatial relations between regions have been constructed. The approach needs to deal with the second issue and design an efficient matching process.

Current solutions to the graph matching problem are computationally expensive since they address all mappings of the node sets. This approach uses an approximate greedy solution to graph matching called the nearest neighbor graph (NN graph) [65] which can be constructed by connecting edges from each node to its neighborhood nodes in a fixed neighborhood threshold . First, a nearest neighbor graph needs to be constructed for each image. Nodes in the graph are sites (salient regions) in the Voronoi diagram and two nodes in the graph have a connecting edge if and only if they are neighbors in the Voronoi diagram or have a connected triangle edge in Delaunay triangulation.

So two nodes  $v_1, v_2$  have an edge in a nearest neighbor graph if,

$$dist(v_1, v_2) \leq \tau \quad (3.8)$$

Where  $dist(v_1, v_2)$  is the Euclidean distance between the detected regions. The most optimal value chosen for  $\tau$  is one. This value models the sufficient number of edges that make the graph representation discriminative [65]. So a region has edges only to its direct neighbors in the Voronoi diagram, so the graph evolves from the Voronoi diagram and model the spatial neighborhood of regions.

The matching process works as follows, given two graphs  $G_1 : V_1, E_1$  and  $G_2 : V_2, E_2$ , for each node  $v_i$  in  $G_1$  a list of best matched nodes is associated according to a matching function  $m(v_i)$ . In this work the matching function compares regions descriptors. The local descriptor was chosen to be the fuzzy color and texture histogram where the Tanimoto Coefficient T [19] was used as the matching function.

$$m(v_i) = T_{ij} = t(x_i, x_j) = \frac{x_i^T x_j}{x_i^T x_i + x_j^T x_j - x_i^T x_j} \quad (3.9)$$

Where  $x^T$  is the transpose of vector  $x$ . The Tanimoto Coefficient takes the value of 1 for absolute convergence and tends to zero for maximum deviation.

Given the node  $v_i$  in  $G_1$  and its associated best matched list of nodes. The need is to find the best matched pair of nodes:  $v_i, v_j$  among matched nodes in the best matched list. The nearest neighbor graph matching technique in [65] choose the best matched pair  $v_i, v_j$  that had the lowest mean of distances between neighbors of node  $v_i$  and neighbors of each node  $v_j$  in the best matched list. This produced a non accurate rank because of neighboring nodes that may have large matching distances. A different similarity function is proposed which compare the neighborhood nodes of node  $v_i$  with the neighborhood nodes of node  $v_j$  and assign the rank of pair  $r(v_i, v_j)$ ,

$$r(v_i, v_j) = \frac{\text{num\_of\_matched\_neighbors\_of\_}(v_i, v_j)}{\text{total\_num\_of\_neighbors\_of\_}v_i} \quad (3.10)$$

The chosen best matched pair is the one with the highest rank  $r(v_i, v_j)$ . The

rank of best matched pairs  $bestR(v_i)$  is calculated for each node in  $G_1$ . A final similarity rank is assigned as the mean rank,

$$similarity\_rank = \frac{\sum_i bestR(v_i)}{num\_of\_nodes\_in\_G_1} \quad (3.11)$$

For a given query image, a nearest neighbor graph is constructed for each image in the data set and the proposed final similarity rank is determined by applying the greedy nearest neighbor graph matching algorithm.

## 3.6 Experimental Work

### 3.6.1 Standard Datasets

Many authors used their own image sets and retrieval performance measures. This made a comparison of retrieval algorithms virtually impossible. This framework uses two standard databases the Wang1000 and the Uncompressed Color Image Database (UCID) that were created especially for the purpose of evaluating content based image retrieval techniques.

Each standard database comes with a set of standard queries and their associated relevant image results. This set of relevant images is called the ground truth and has to be defined for the database to carry out evaluations.

#### 3.6.1.1 Wang 1000

The Wang database [61] contains 10 categories. Each category contains 100 images that varies from complex structures to smooth regions. The database images have several variations such as occlusion between objects, changes of illumination and geometric deformations. Table 3.1 shows the database categories.

The MIRROR image retrieval system [9] separated the Wang database into 20 queries each with a proposed ground truth.

ID	Category Name
1	Africa people and villages
2	Beach
3	Building
4	Buses
5	Dinosaurs
6	Elephants
7	Flowers
8	Horses
9	Mountains and glaciers
10	Food

Table 3.1: Wang database categories.

### 3.6.1.2 UCID

The UCID dataset [70] contains 1338 uncompressed TIFF images. These images represent a variety of topics including natural scenes, man-made objects, and both indoors and outdoors. The images were taken by a camera, setting its exposure, contrast, color balance, etc. to automatic as most average users would do.

162 standard queries with their associated ground truth were defined on the image set.

### 3.6.2 Performance Evaluation

Several measures were proposed to measure the performance of a CBIR system. Precision and recall were popular ones. "Precision and recall are metrics to evaluate the ranking of the images returned by the system for one category" [71]. There is a tradeoff between the precision and recall. The precision curve continue to decrease or become stationary, and the best precision curve is one which has the least decrement. This means that for a certain number of queries carried over the system, most of the images retrieved by the system are relevant to the query. The recall curve continue to increase and the best recall curve is one which has the fastest increment. This means that for a certain number of queries carried over the system, most of the retrieved images are relevant, and few of them are lost.

For a query  $q$  having a defined ground truth  $R(q)$  and a retrieved result set  $Q(q)$ . The precision of the retrieval is defined as the fraction of the retrieved images that are indeed relevant to the query" [15].

$$precision = \frac{|Q(q) \cap R(q)|}{|Q(q)|} \quad (3.12)$$

"The recall is the fraction of the relevant images that is returned by the query" [15].

$$recall = \frac{|Q(q) \cap R(q)|}{|R(q)|} \quad (3.13)$$

Another measurement is the precision/recall curve. For each number of images returned by the system, we define a couple (precision, recall), this curve is the set of all these couples. This curve begins from the top left, continues to decrease regularly and terminates at the bottom right. The best precision/recall curve is one that decreases slowly, indicating that a lot of relevant images are returned by the system and few of them are lost. The size of the category has no effect, allowing different categories to be compared by this curve.

To evaluate a system over all the categories, Top N performance measurements [67] can be used. When submitting a query  $q$  to a CBIR system, the system returns  $N$  image results sorted by similarity to the query image, where  $N$  is the number of top similar images. We denote  $PR_N$  as the precision of the top  $N$  returned sorted results. The aim of the user after submitting a query is to search for the most relevant images  $R(q)$ . the precision  $PR_j, j = 1, 2, \dots, N$  of the top  $N$  results of a query  $q$  is defined as :

$$PR_N(q_i) = \sum_{i=1}^N \frac{\psi(p_k, R(q))}{N}, \psi(x, y) = \begin{cases} 1 & \text{if } x \in y \\ 0 & \text{if } x \notin y \end{cases} \quad (3.14)$$

So the average precision for all queries performed on a CBIR system for a certain  $N$  number of returned results is defined as:

$$PR_N = \sum_{i=1}^{Total\_Query\_Count} \frac{PR_N(q_i)}{Total\_Query\_Count} \quad (3.15)$$

Similarly, the recall  $RE_j, j = 1, 2, \dots, N$  of the top  $N$  results of a query  $q$  is defined as:

$$RE_N(q_i) = \sum_{i=1}^N \frac{\psi(p_k, R(q))}{\|R(q)\|} \quad (3.16)$$

And the average recall for all queries is defined as:

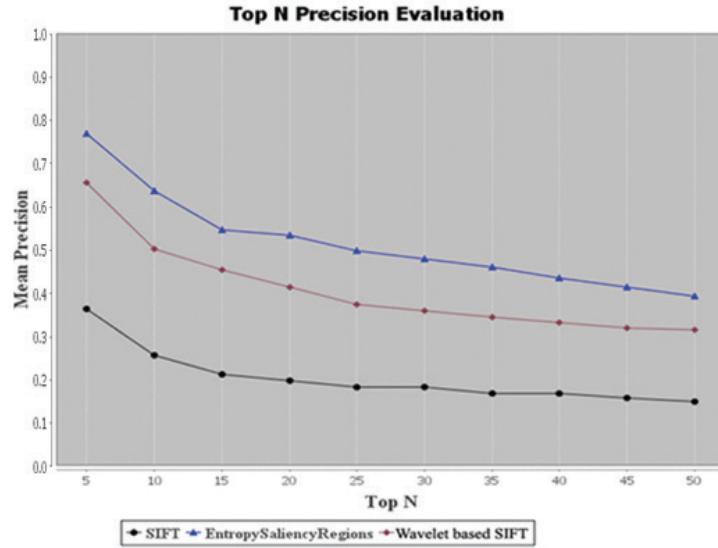
$$RE_N = \sum_{i=1}^{Total\_Query\_Count} \frac{RE_N(q_i)}{Total\_Query\_Count} \quad (3.17)$$

### 3.6.3 Results

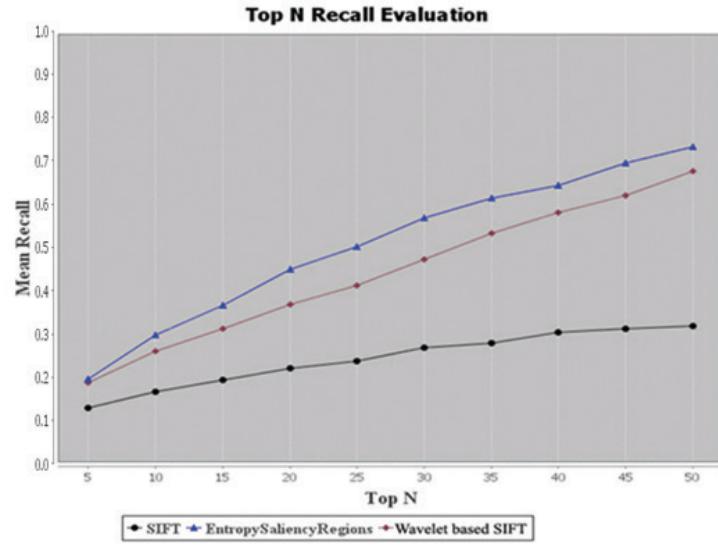
Salient features descriptors are evaluated toward each other. The techniques included in the evaluation are the SIFT descriptor, Wavelet based SIFT and the scale invariant salient regions descriptor. The wavelet based SIFT is proposed during the research work. This approach integrated wavelet based salient points and SIFT descriptor. The wavelet based salient points are used instead of the keypoints detected by the difference of Gaussian pyramid. The approach is implemented and evaluated toward other saliency description techniques. The implementation of the SIFT descriptor is based on the OSCAR open source system [72]. The repeatability rate was used as the similarity measure to provide ranking of the retrieved results.

Figure 3.14 shows the mean precision and mean recall top  $N$  comparisons of salient feature descriptors evaluated over Wang database.  $N$  ranges from 5 to 50 step 5. The scale invariant salient regions descriptor provides higher precision and recall than other approaches due to its capability to capture the most salient scale invariant regions. The wavelet based SIFT is shown to have degraded precision and recall. This is because The wavelet transform provides high response at locations which have high frequencies or changes. These locations are usually indicate edges.

So the detected wavelet based salient points are more targeted to edges, ignoring several image details. The SIFT descriptor has the lowest precision and recall due to the use of fixed window to describe the salient regions around keypoints. This



(a) Precision comparisions of salient feature descriptors.

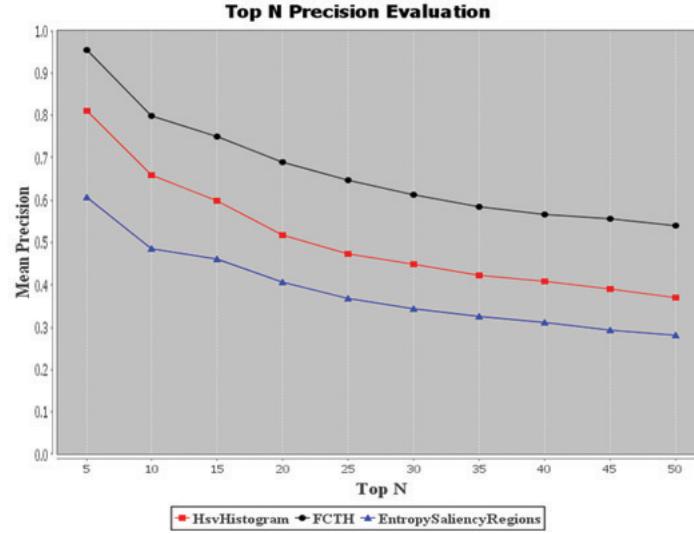


(b) Recall comparisions of salient feature descriptors.

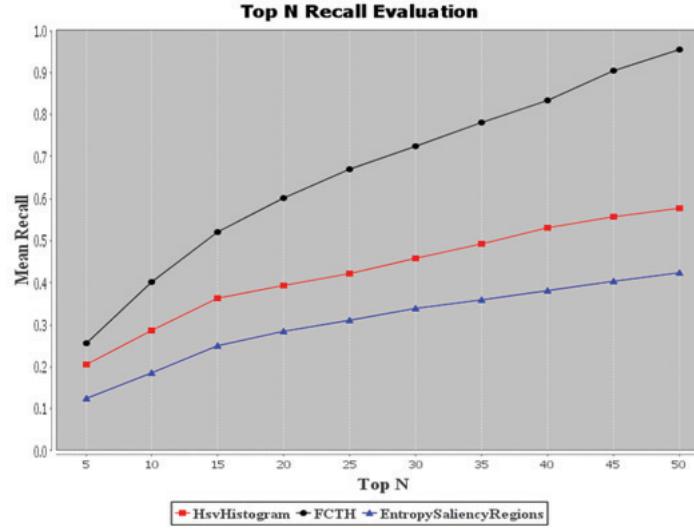
Figure 3.14: Precision and and recall comparisions of salient feature descriptors performed on Wang database.

limits the SIFT capability for image retrieval due to several image deformations. The SIFT descriptor a good candidate for applications such as camera calibration.

Figure 3.15 shows the mean precision and mean recall comparisons of the FCTH composite feature descriptor, the HSV histogram descriptor and the entropy scale invariant salient regions descriptor evaluated over Wang database. Comparisons



(a) Precision comparisions of FCTH, HsvHistogram and entropy scale invariant salient regions descriptors.



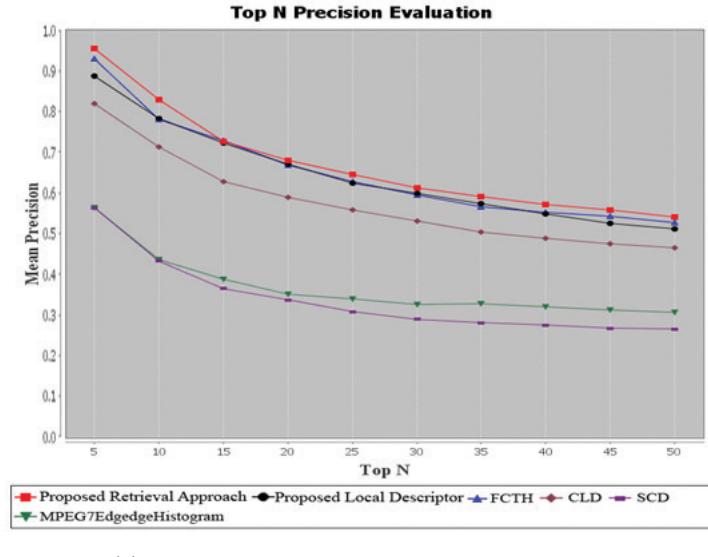
(b) Recall comparisions of FCTH, HsvHistogram and entropy scale invariant salient regions descriptors.

Figure 3.15: Precision and and recall comparisions of FCTH, HsvHistogram and entropy scale invariant salient regions descriptors performed on Wang database.

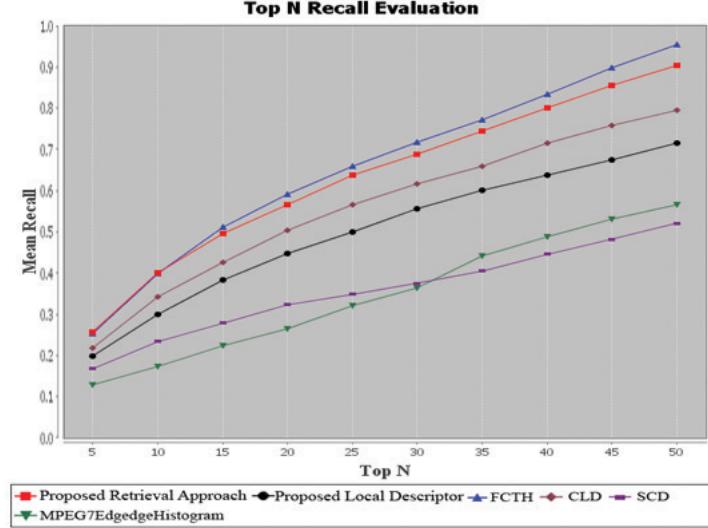
show that the FCTH provides higher precision and recall than other approaches. This is due to the integration of color and texture features that provides better description for the image content and better retrieval accuracy than other tech-

niques. Image features need to be itegrated to provide better retrieval accuracy.

Figure 3.16 shows the top N mean precision and recall comparison of the proposed approach compared to other approaches evaluated over the Wang database. Figure 3.17 shows the same comparison evaluated over the UCID database. The



(a) Mean precision evaluated over Wang database.



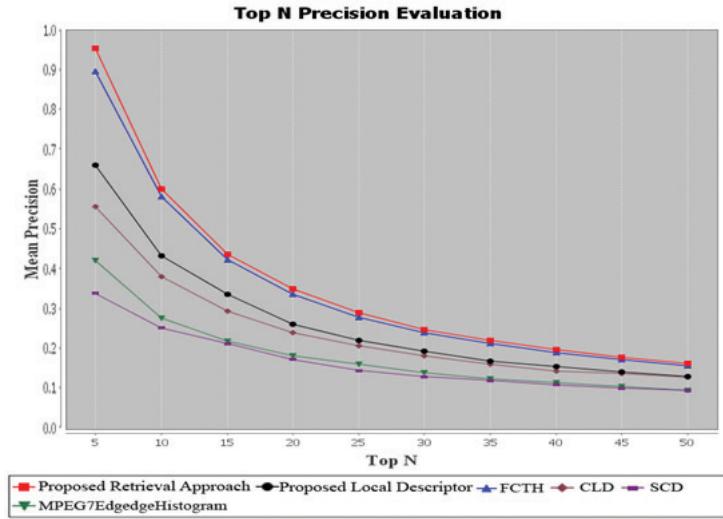
(b) Mean recall evaluated over Wang database.

Figure 3.16: Comparison of the mean precision and recall of the top N results for the proposed approach toward other approaches evaluated over Wang database.

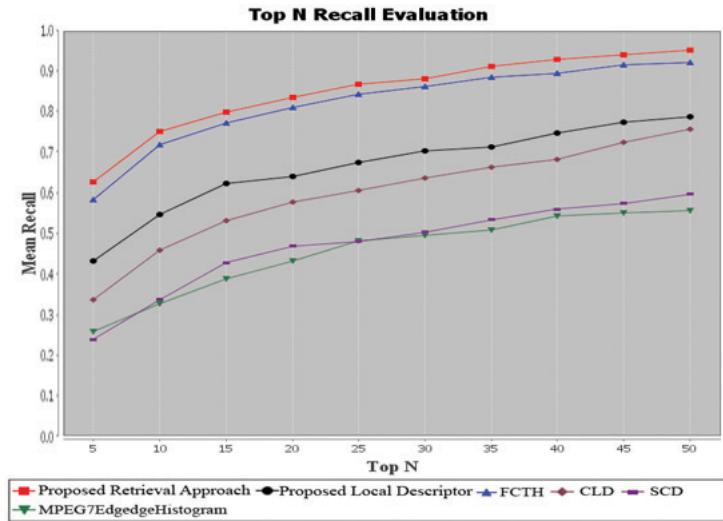
comparison was made between the proposed retrieval approach that combines salient regions with graph matching approach as local descriptor and FCTH [10]

as global descriptor, the salient regions with graph matching approach local descriptor alone, the FCTH global descriptor and the following MPEG7 descriptors [9]:

Color descriptors: Scalable Color Descriptor (SCD), Color Layout Descriptor (CLD). Texture Descriptor: Edge Histogram Descriptor (EHD). Also the results have been compared to the fuzzy color and texture histogram (FCTH) composite descriptor .



(a) Mean precision evaluated over UCID database.



(b) Mean recall evaluated over UCID database.

Figure 3.17: Comparison of the mean precision and recall of the top N results for the proposed approach toward other approaches evaluated over UCID database.

Figure 3.16 and 3.17 show that the salient regions with graph matching local descriptor technique alone is more accurate for queries having complex spatial structure and less accurate for queries having large smoothed regions. When the salient regions with graph matching local descriptor technique is combined with the FCTH global descriptor approach, the combination provides higher precision than other approaches (figure 3.16(a) and 3.17(a)). Alternatively the proposed approach has degraded accuracy for images having large smoothed region with less or no complexity in its structures because very few number of salient regions will be extracted. This results in a degraded recall of retrieved images (figure 3.16(b) and 3.17(b)).

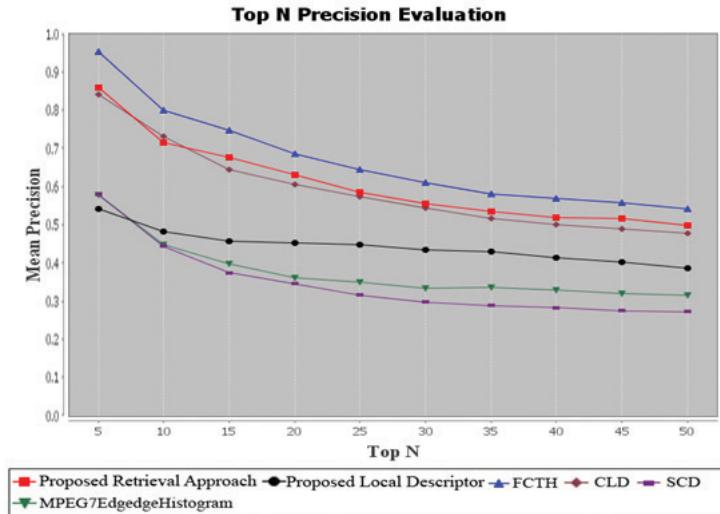


Figure 3.18: Drop of precision of proposed retrieval approach due to using the actual NN graph matching technique evaluated over Wang database.

Figure 3.18 shows the results of using the actual NN graph matching technique. The figure shows a drop of the precision of the proposed approach toward other approaches; this is because the mean of distances between neighbors was used to choose the best matched pair of two nodes. If two nodes in the neighbors have large distance; the mean will be influenced by this distance even if the other neighbors have very small distances. The proposed modification of this graph matching technique enhance the results as shown in figure 3.16(a) where equation (3.11) was used to assign a rank for each pair of nodes and choose the best pair.

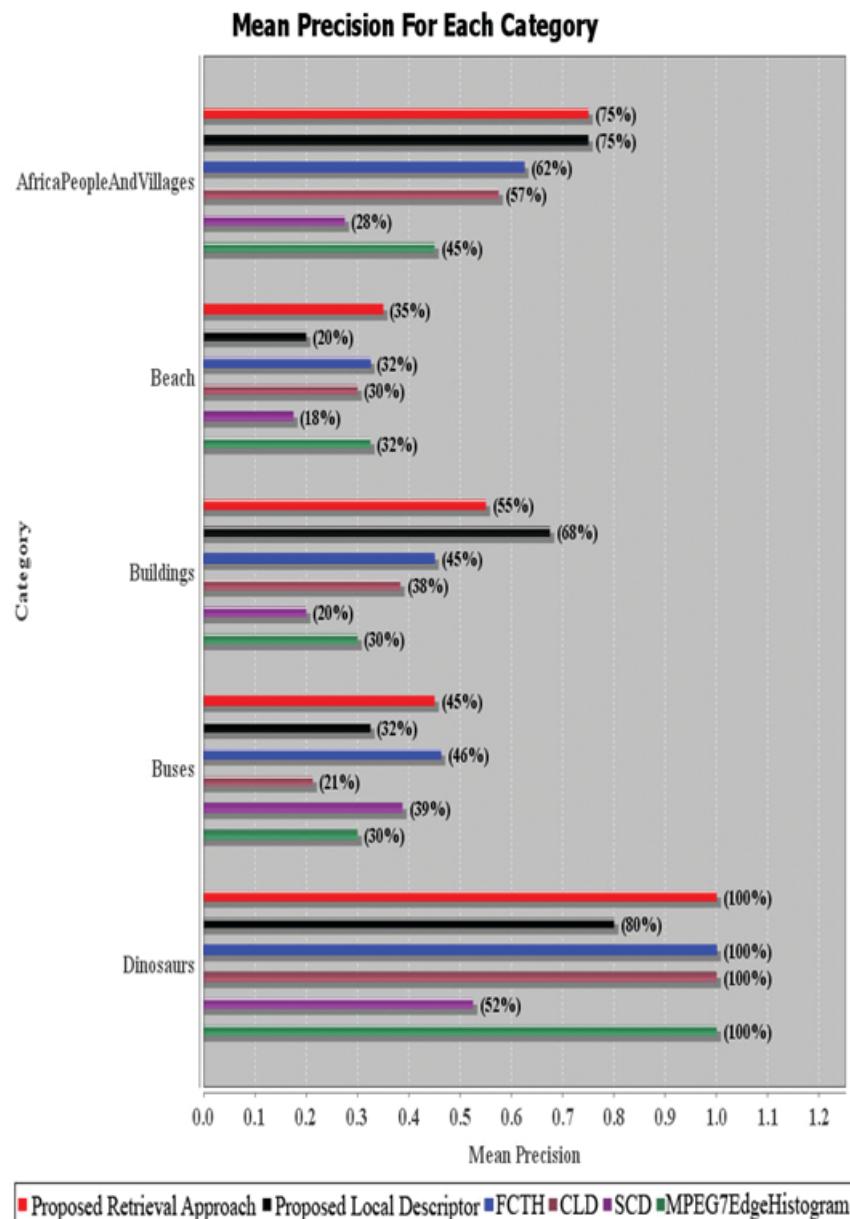


Figure 3.19: The mean precision for first 5 categories of Wang database.

Figure 3.19 and 3.20 show the mean precision for the ten categories of the Wang database. For each category the bars are arranged from top to down

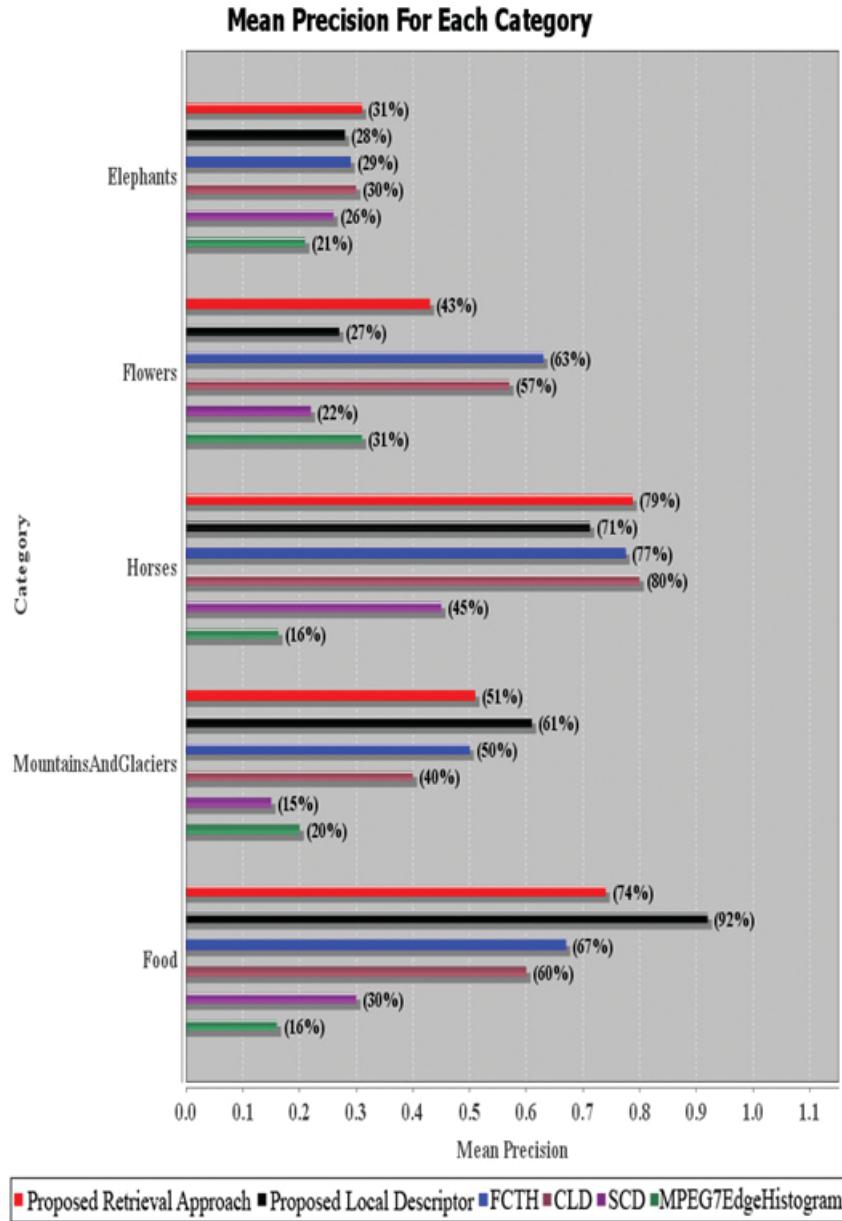


Figure 3.20: The mean precision for last 5 categories of Wang database.

with respect to the compared approaches. The figures show that the proposed retrieval approach provides higher precision for categories that contain images having complex details such as africaPeopleAndVillages, beach, buildings, elephants, mountainsAndGlaciers and food. This is due to the use of salient regions that capture the most scale invariant complex regions in image. Moreover, the pro-

posed approach has degraded accuracy for categories having smooth details such as buses, flowers, Dinosaurs and horses.

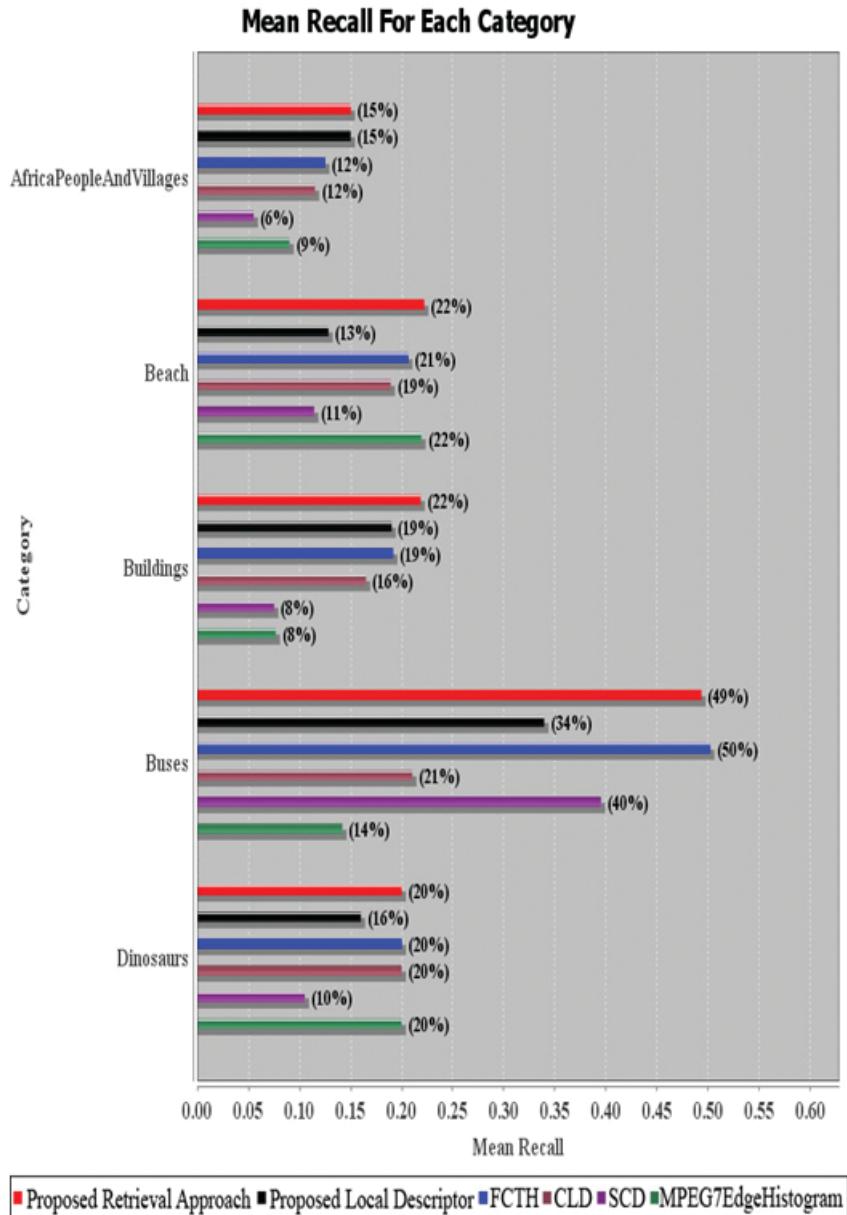


Figure 3.21: The mean recall for first 5 categories of Wang database.

Figure 3.21 and 3.22 show the mean recall for the ten categories of the Wang database. The figures show the drop of recall for categories containing images that have smooth details such as buses, Dinosaurs, flowers and horses. Although the

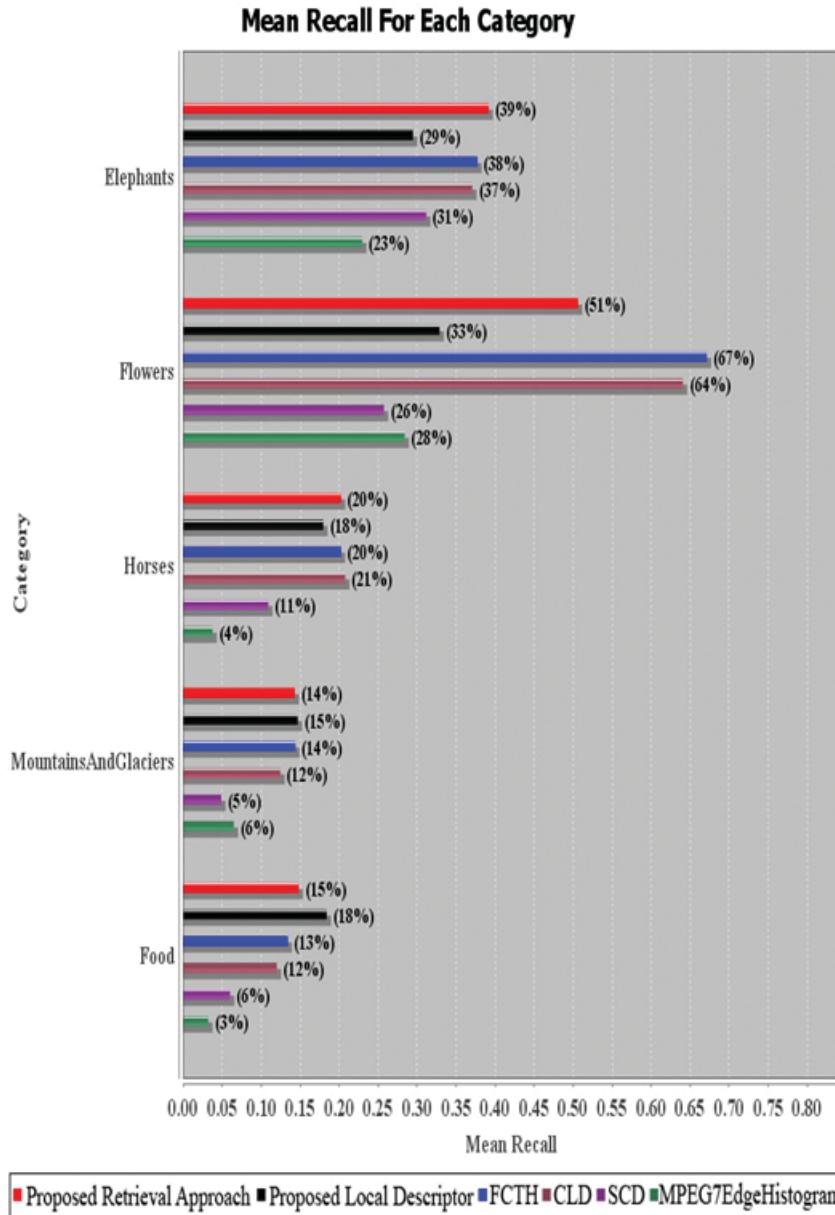


Figure 3.22: The mean recall for last 5 categories of Wang database.

recall is better for the remaining categories, a few number of salient regions will be extracted in images having smooth details. Thus degrading the recall accuracy.

# Chapter 4

## Principal Regions Detection for Efficient Retrieval

### 4.1 Introduction

This chapter addresses region-based image retrieval and proposes a new region-based image retrieval approach called Principal Regions Image Retrieval (PRIR). This approach provides a local description of an image. It segments an image to the most general principal regions each of which is characterized by a fuzzy feature histogram reflecting the color and texture properties. A nearest neighbor graph is generated for the segmented regions and a greedy graph matching algorithm with a modified scoring function is applied to determine the image rank. The approach provides efficient region based segmentation and treats the over segmentation problem. The proposed approach combines local and global description to enhance the retrieval results. Figure 4.1 shows a block diagram of the proposed approach.

The new proposed technique provides an efficient retrieval for large collections of images. The technique begins with performing morphological operations on each image of the dataset. The morphological operations [73] smooth the contours of objects and break narrow isthmuses and eliminate thin protrusions. In addition, they fuse narrow breaks and fill gaps in the contour and eliminate small holes. The process is described in section 4.2.

After performing morphological operations, an HSV color quantization [25] is applied to eliminate the differences of color shades existed between similar regions thus, fusing the neighboring regions having the same color attributes. The process is described in section 4.3.

To control the number of general principal regions, firstly, region labeling [74] is performed, secondly the regions are sorted based on their size, and finally, a specific resolution is selected to combine small regions with larger ones based on the color hue criteria. The process is described in section 4.4. After extracting the principal regions, we use the 192 bin fuzzy color and texture histogram composite feature descriptor [10] as a local descriptor for each region. Section 4.5 describes the operation. We have investigated the spatial relationship between the detected principal regions by constructing a nearest neighbor graph [65] in which each region and its associated descriptor is a node in the graph, then, a greedy graph matching algorithm with a modified scoring function [75] is applied to determine the final image rank. Section 4.6 describes the operation.

Finally Section 4.7 provides the experimental results of an image retrieval system that implements the proposed retrieval approach and shows the effectiveness of our approach compared to other approaches.

## 4.2 Morphological Operations

Morphological operations [73] create an output image of the same size of input image by applying a structuring element. Dilation and erosion are the most primitive morphological operations. Another two important operations that are built upon dilation and erosion are the morphological opening and closing. Opening generally smoothes the contour of object, breaks narrow isthmuses, and eliminates thin protrusions. Closing tends to smooth sections of contours but also opposed

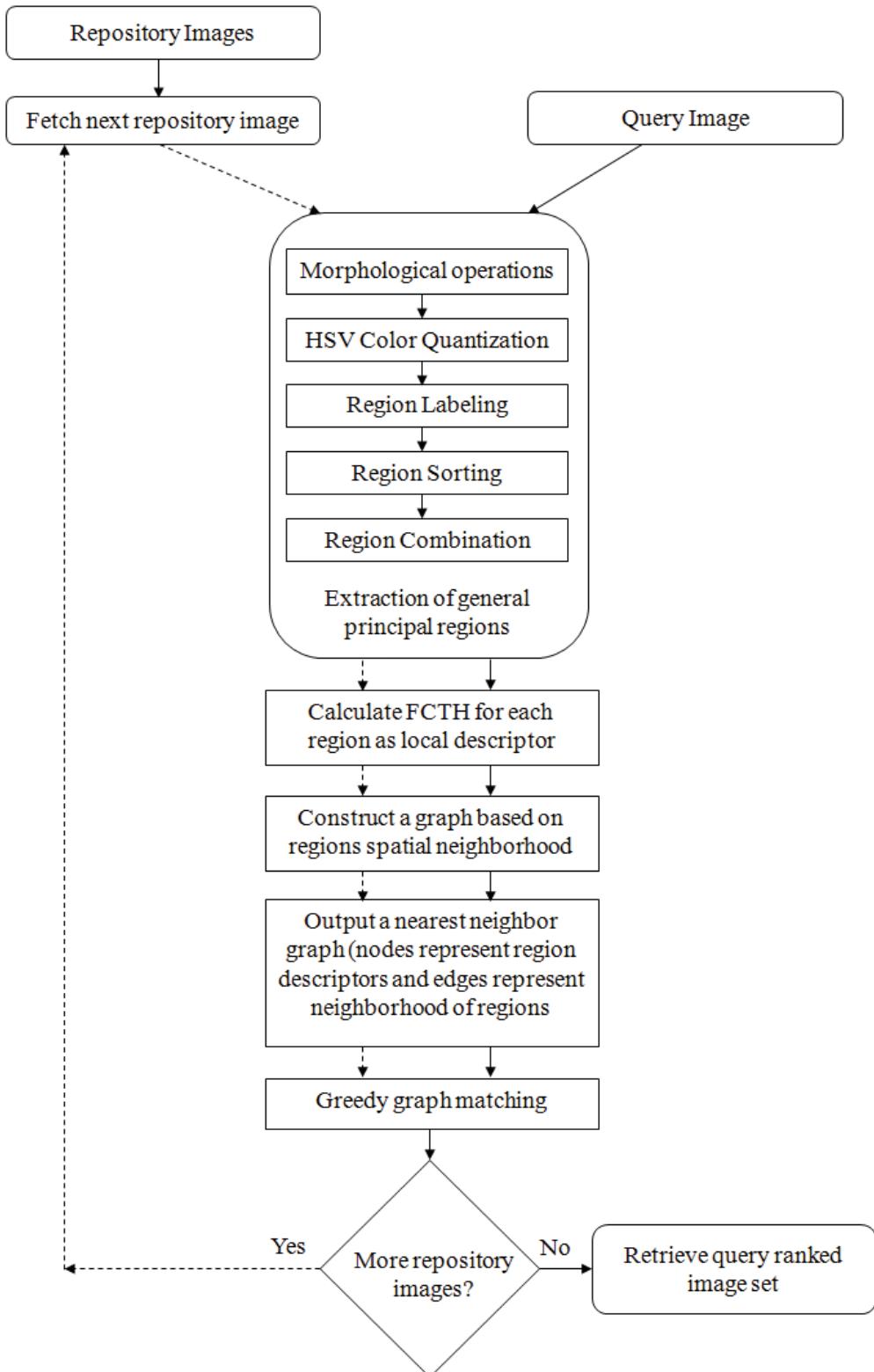


Figure 4.1: A flowchart showing the proposed approach.

to opening it generally eliminates small holes, and fills gaps in the contour.

### 4.2.1 Dilation

Dilation adds pixels to the boundaries of objects in an image (figure 4.2). Pixels beyond the image border are assigned the minimum value afforded by the data type. For binary images; these pixels are assumed to be set to 0. For grayscale images, the minimum value for uint8 images is 0. Given an image  $I \in E^N$  and a structuring element  $S \in E^N$ , the dilation operator on sets I and S are defined by:

$$I \oplus S = \{c \in E^N | \exists i \in I, \exists s \in S, c = i + s\} \quad (4.1)$$



Figure 4.2: The effect of dilate morphological operation using a  $9 \times 9$  structuring element of value 1.

### 4.2.2 Erosion

Erosion removes pixels on object boundaries (figure 4.3). Pixels beyond the image border are assigned the maximum value afforded by the data type. For binary images these pixels are assumed to be set to 1. For grayscale images, the maximum value for uint8 images is 255. The Erosion operator on sets I and S are defined by:

$$I \ominus S = \{x \in E^N | \forall s \in S, x - s \in I\} \quad (4.2)$$



Figure 4.3: The effect of erode morphological operation using a  $9 \times 9$  structuring element of value 1.

#### 4.2.3 Morphological Opening



Figure 4.4: The effect of morphological opening operation using a  $9 \times 9$  structuring element of value 1.

Morphological opening of an image is erosion followed by dilation, using the same structuring element for both operations (figure 4.4). Opening generally smoothes the contour object, breaks narrow isthmuses, and eliminates thin protrusions. Morphological opening is expressed by :

$$I \circ S = (I \Theta S) \oplus S \quad (4.3)$$

$$(I \circ S) \circ S = I \circ S \quad (4.4)$$

#### 4.2.4 Morphological Closing

Morphological closing of an image consists of dilation followed by erosion with the same structuring element (figure 4.5). Closing also tends to smooth sections of contours but also opposed to opening it generally combines narrow breaks and long thin gulfs, eliminates small holes, and fills gaps in the contour. Morphological opening is expressed by :

$$I \bullet S = (I \oplus S) \Theta S \quad (4.5)$$

$$(I \bullet S) \bullet S = I \bullet S \quad (4.6)$$

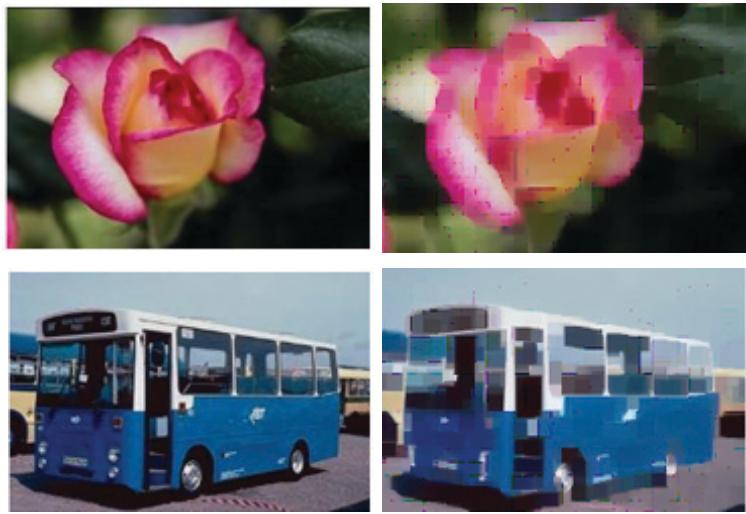


Figure 4.5: The effect of morphological closing operation using a  $9 \times 9$  structuring element of value 1.

#### 4.2.5 Morphological Opening followed by a Closing

In this work, for each image, we applied an opening followed by a closing (figure 4.6) [10] using a  $9 \times 9$  structuring element of ones. This has the effect of removing

dark spots and stem marks and attenuating bright and dark artifacts.

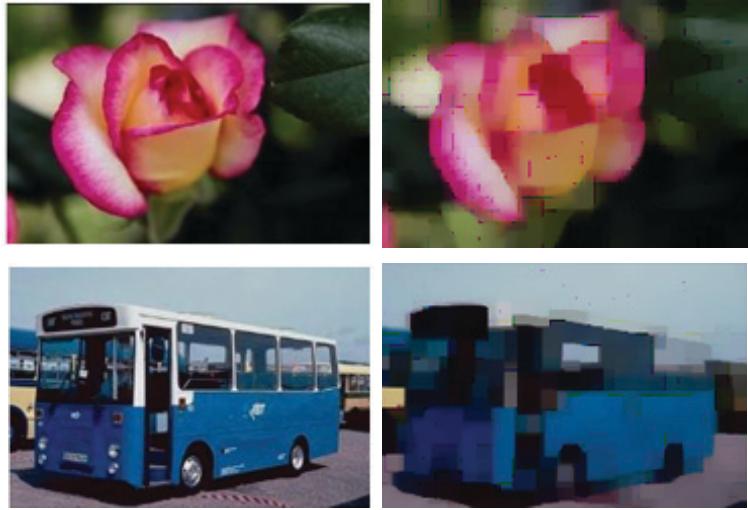


Figure 4.6: The effect of morphological opening followed by Closing using a  $9 \times 9$  structuring element of value 1.

### 4.3 HSV Color Quantization

HSV is a widely adopted space in image and video retrieval. The main advantages of this color space [74] are that chrominance (H, S) and luminance (V) components are decoupled. The RGB model is suited for image color generation, whereas the HSV model is suited for image color description. The HSV histogram is based on the HSV color. The HSV histogram partition the HSV color Space, where each partition represents a bin in the histogram and contains closely related colors with different shades.

In this framework, HSV color quantization [25] is applied to remove the differences of color shades existed between similar regions thus, fusing the neighboring regions having the same color attributes (figure 4.7). The framework uses an implementation of HSV histogram that divides hue into sixteen regions, whereas saturation and value are divided into eight each. The three color components then are linked, thus creating a  $(16 \times 8 \times 8)$  histogram of 1024 bins.

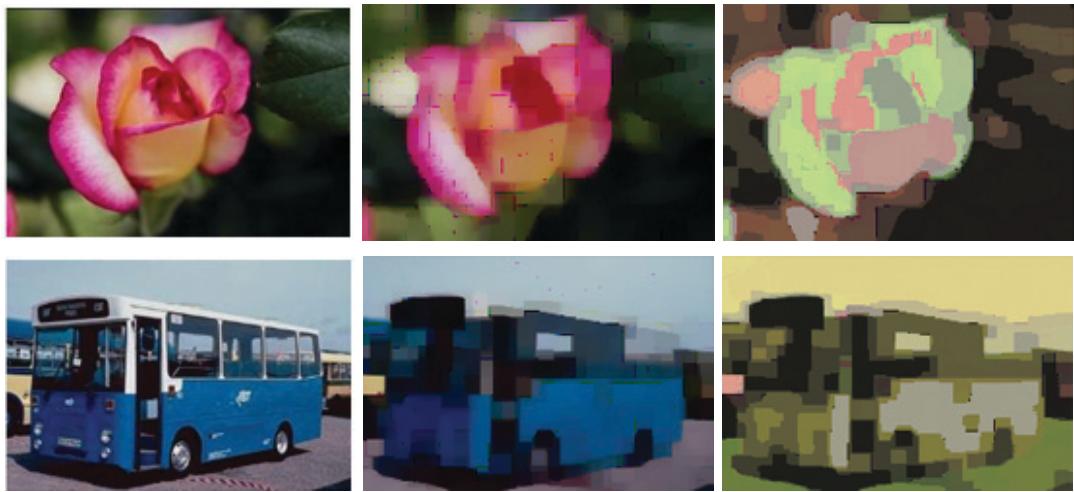


Figure 4.7: Applying HSV color Quantization after morphological operations.

## 4.4 Over-Segmentation Handling

The over-segmentation problem results from the existence of variations and noise in the color level values. The morphological operations and HSV color quantization reduce the effect of over-segmentation but don't solve this problem completely (figure 4.9).

In order to overcome this problem completely, the number of segmented regions has to be controlled. Firstly, region labeling is performed, secondly the regions are sorted based on their size, and finally, a specific number of general principal regions is selected to combine small regions with larger ones based on the hue criteria. The following subsections will explain each step in more details.



Figure 4.8: The over-segmentation problem.

#### 4.4.1 Region Labeling

Segmentation divides an image into a number of homogeneous regions. Region labeling is the operation of giving a unique identification (ID) to these regions. This stage gives each segmented region a unique ID.

#### 4.4.2 Region Sorting

Sorting is the process of ordering a collection of entities based on specific criteria. The list of segmented regions forms the collection to be sorted and the criteria of sort is the size (number of pixels) of each region. Regions are sorted in a descending order (figure 4.9).

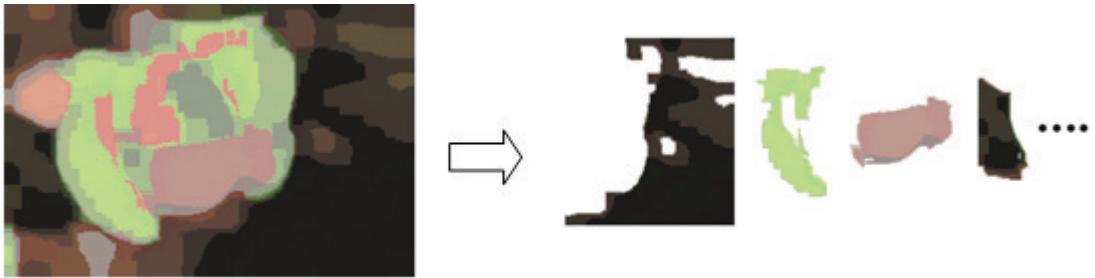


Figure 4.9: Sorting image segmented regions based on size.

#### 4.4.3 Region Combination

Region combination is the process of combining neighboring regions based on some criteria. This work chooses the hue criteria for region combination as the hue represents the pure color value irrespective of its shades and saturation. A region is combined to one of its neighbors having the smallest difference in hue. The number of segmented regions can be controlled by identifying a set of large size regions and combining other small ones to them. The identified set of regions is chosen as the top elements in the segmented regions sorted list. The number of regions selected varies from smooth images to complex ones (multi-resolution). Figure 4.10 shows the segmentation results of a set of database images. The number of principal regions is chosen to be 10, 20 and 30. It can be shown that

we can have a good approximation between smoothness and complexity using 20 principal regions.

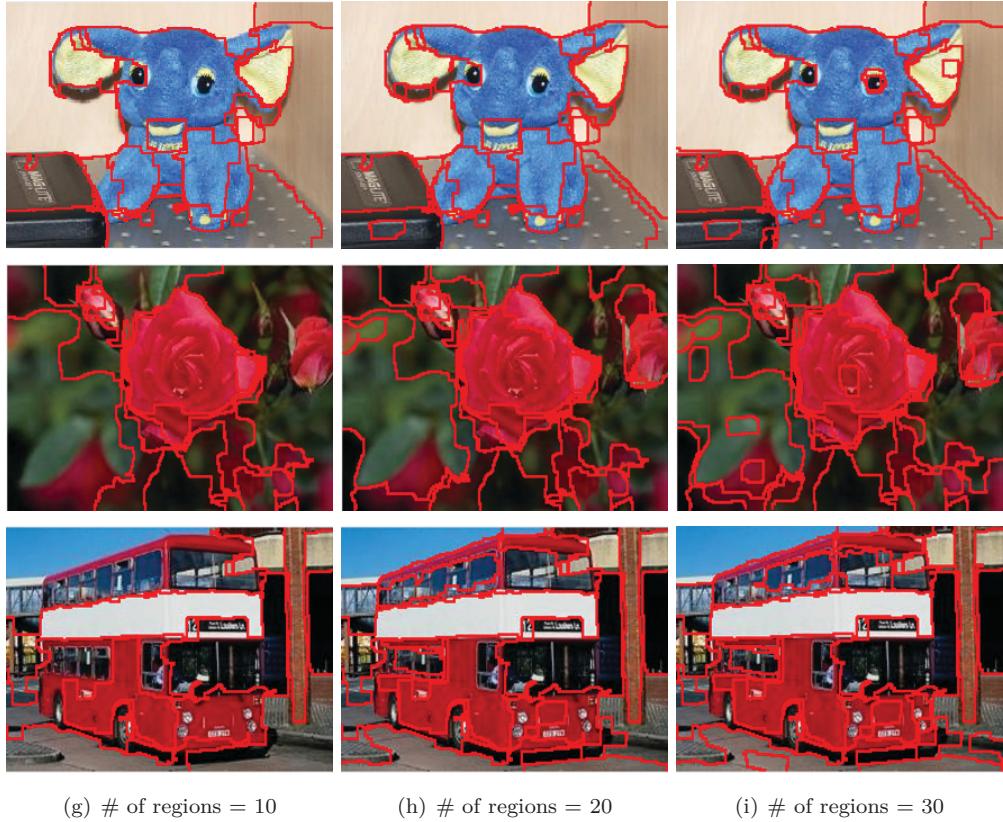


Figure 4.10: Controlling the number of segmented regions by combining small regions to large regions.

## 4.5 Description of Segmented Regions

Local description of image segmented regions provides a summary description of image content. An appropriate descriptor must provide a local invariant description and be robust with respect to occlusion and geometrical transformations. In this framework, the 192 bin fuzzy color and texture histogram [10] is used as a local descriptor.

A 192-bin histogram is generated for each segmented region and works as its local descriptor. Each image is described by a set of segmented regions distributed

on its layout and each region is associated with a 192-bin histogram as a local region descriptor. This set of regions works as a reduced description of the entire image.

## 4.6 Spatial Graph Generation

Spatial graph investigates the neighboring relationships between the non overlapped extracted segmented regions. The nearest neighbor graph [65] (NN graph) is a spatial graph constructed by connecting edges from each node to its neighboring nodes in a fixed neighborhood threshold  $\tau$ .

So two nodes have an edge in a nearest neighbor graph if,

$$dist(v_1, v_2) \leq \tau \quad (4.7)$$

This graph is generated using the neighborhood of pixels between segmented regions (figure 4.11). Each pixel has an 8 connected neighbors, a region is a neighbor to another, if at least one pixel from this region is 8-connected to a pixel from the other region. Each region and its associated descriptor is a node in the graph. Edges in the spatial graph represent the neighboring relationships between segmented regions. After generating the NN graph, the greedy nearest neighbor

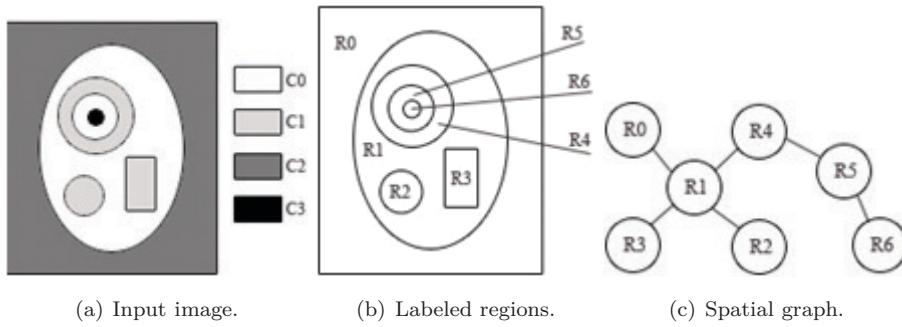


Figure 4.11: Spatial graph generation, regions and their associated descriptors are nodes in the graph.

graph matching algorithm with the proposed modified scoring function [75] is used as a similarity measure as explained in the previous chapter.

For a given query image, a NN graph is constructed for each image in the data set and the proposed final similarity rank is determined by applying the greedy nearest neighbor graph matching algorithm.

## 4.7 Experimental Work

The experiments are performed on the 1000 image Wang database [61] and the 1338 Uncompressed Color Image Database (UCID) [70] standard datasets.

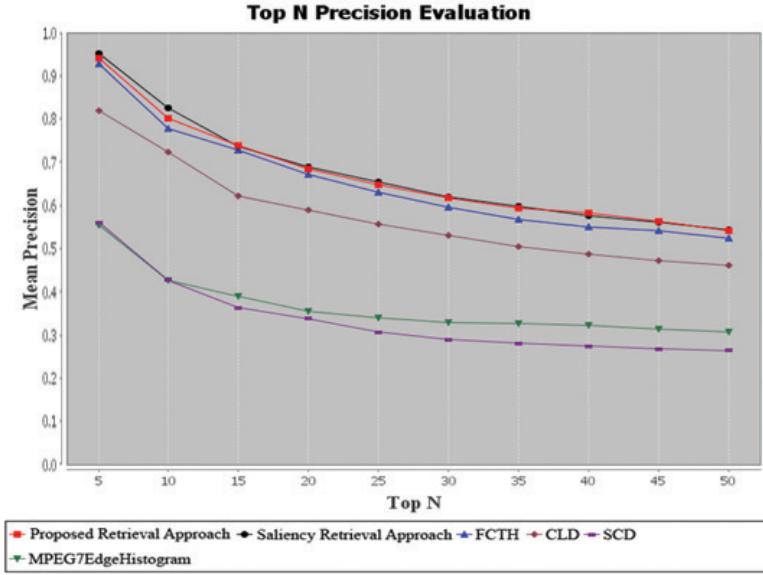
The segmentation stage of the proposed approach uses 20 general principal regions. This number of regions gives better segmentation results as shown in figure 4.10 and provide a compromise between images that have large details and others having small details. Images having segmented regions above 20 will have their regions sorted by size and combined to the largest 20 regions based on the hue criteria.

The proposed approach combines local region description with global FCTH description. This integration captures general and local features of images. The results of the proposed retrieval approach have been compared with the results of the following techniques:

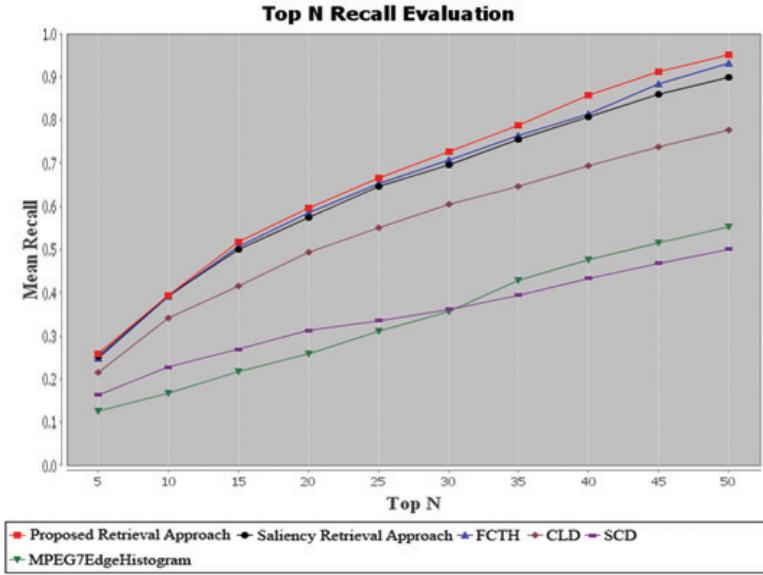
MPEG7 Color descriptors [9]: Scalable Color Descriptor (SCD) and Color Layout Descriptor (CLD). MPEG7 Texture Descriptor [9]: Edge Histogram Descriptor (EHD). Composite descriptors: the saliency retrieval approach [75] and FCTH descriptor.

Figure 4.12 shows the top N precision and recall comparison evaluated over the Wang database. Figure 4.13 shows the same comparison evaluated over the UCID database. N ranges from 5 to 50 step 5. Figure 4.12(a) shows that the proposed approach provides more precision than FCTH, close to the saliency approach and better than other approaches over Wang database. The saliency retrieval approach

describes the most complex regions in images. This allows the retrieval of the most



(a) Top N precision comparison.

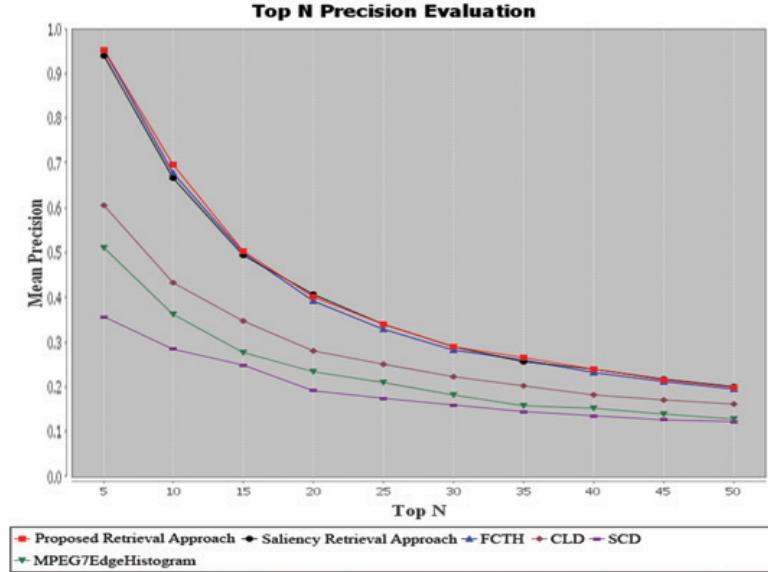


(b) Top N recall comparison.

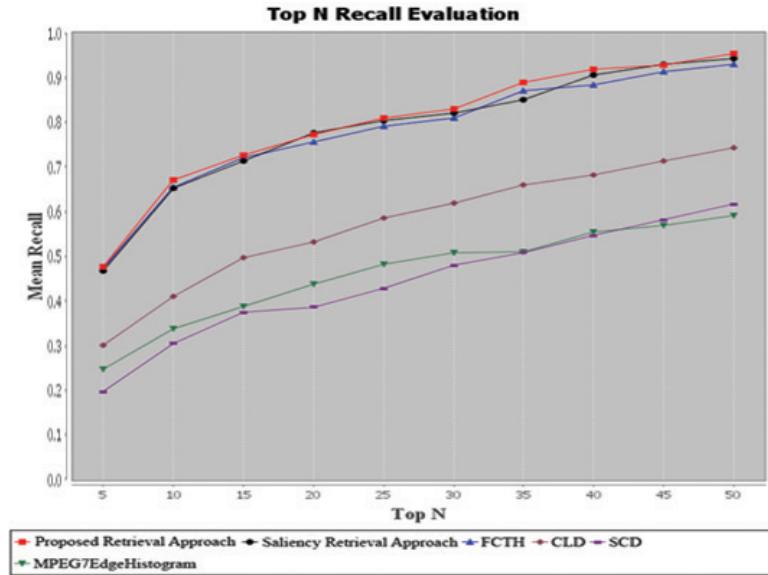
Figure 4.12: Top N precision and recall comparisons evaluated over Wang database.

precise images having the same salient attributes of a query image. This approach is efficient for images having complex spatial structure as in wang database where query relevant images have large variations and object dissimilarity. Alternatively the saliency approach has degraded accuracy for images having large smoothed regions with less or no complexity in its structures because very few number of

salient regions will be extracted. The saliency approach focuses on the most precise



(a) Top N precision comparison.



(b) Top N recall comparison.

Figure 4.13: Top N precision and recall comparisons evaluated over UCID database.

details that enhance precision while ignoring smooth details that lowers the recall. The proposed retrieval approach works as an extension to the saliency approach that enhances the precision for images having large smoothed regions and provides better recall than other approaches as shown in figure 4.12(b).

Figure 4.13(a) and 4.13(b) shows that the proposed approach provides better precision and recall against other approaches when evaluated over the UCID database where query relevant images are quite similar and has small geometric and photometric variations.

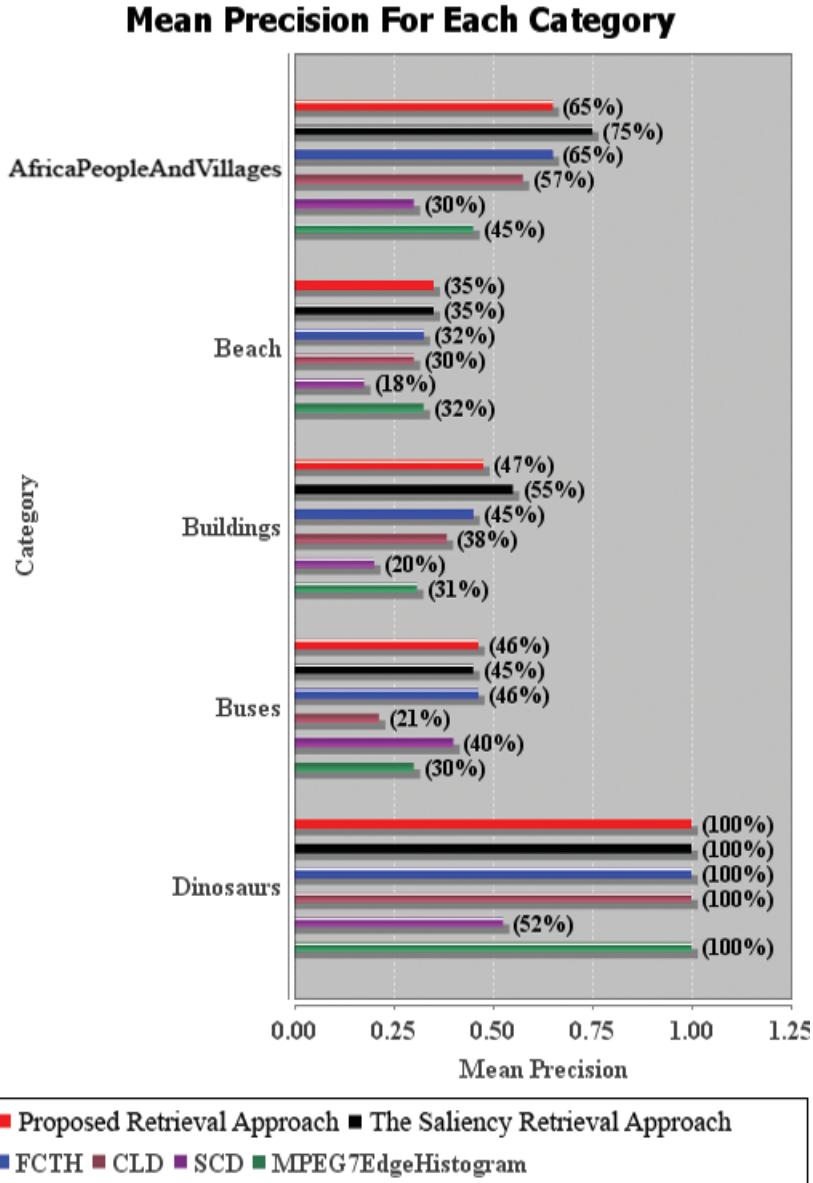


Figure 4.14: The mean precision for first 5 categories of Wang database.

Figure 4.14 and 4.15 show the the mean precision for the ten categories of the Wang database. For each category the bars are arranged from top to down with respect to the compared approaches. The figures show that the proposed re-

trieval approach provides higher precision for categories that contain images having smooth details such as buses, flowers and dinosaurs. This is due to segmenting the

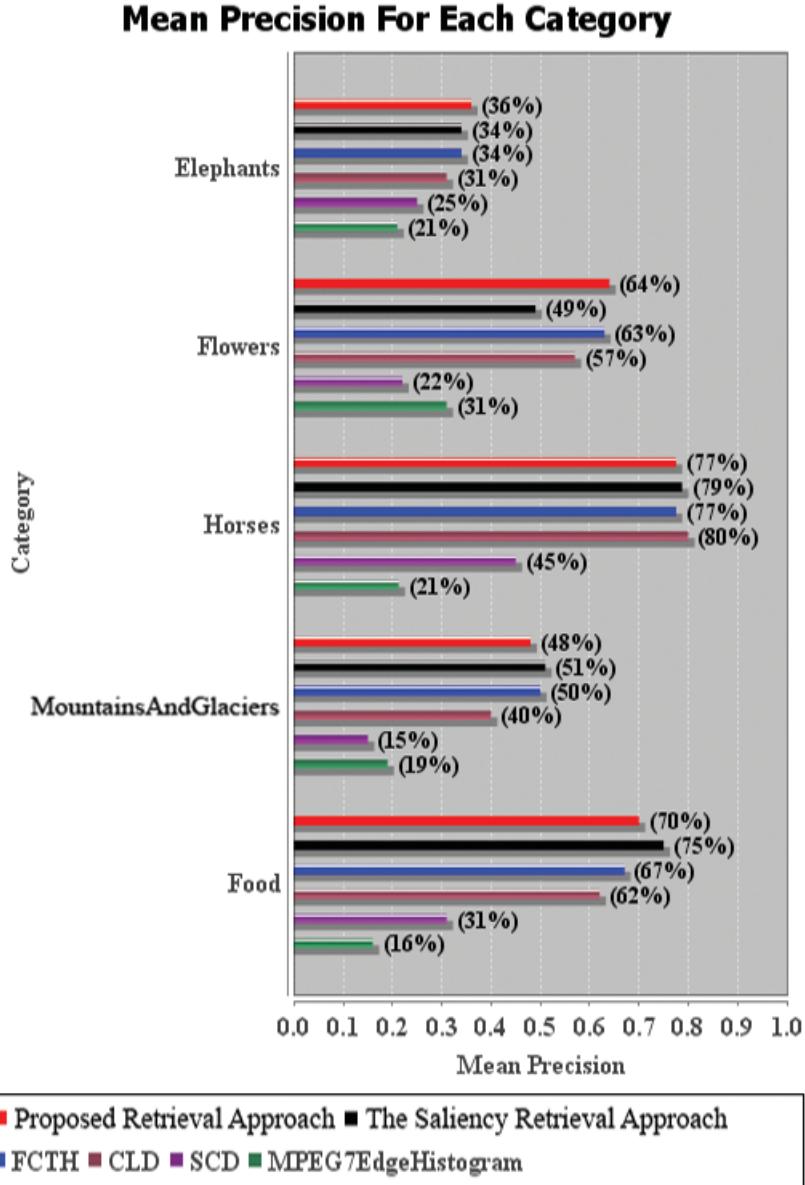


Figure 4.15: The mean precision for last 5 categories of Wang database.

image into the general principal regions, thus capturing the features of the smooth regions. Moreover, the proposed approach has precision close to the saliency approach for categories having complex details such as africaPeopleAndVillages, beach, buildings, elephants, mountainsAndGlaciers, food and horses.

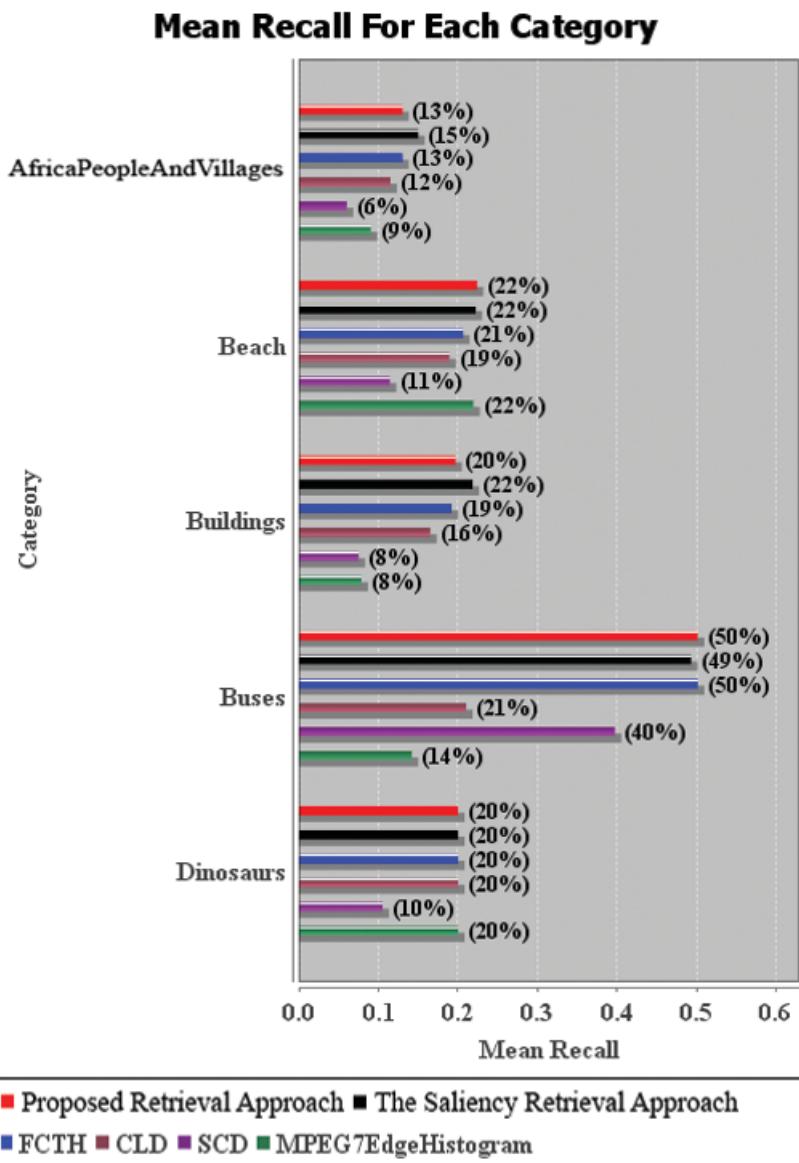


Figure 4.16: The mean recall for first 5 categories of Wang database.

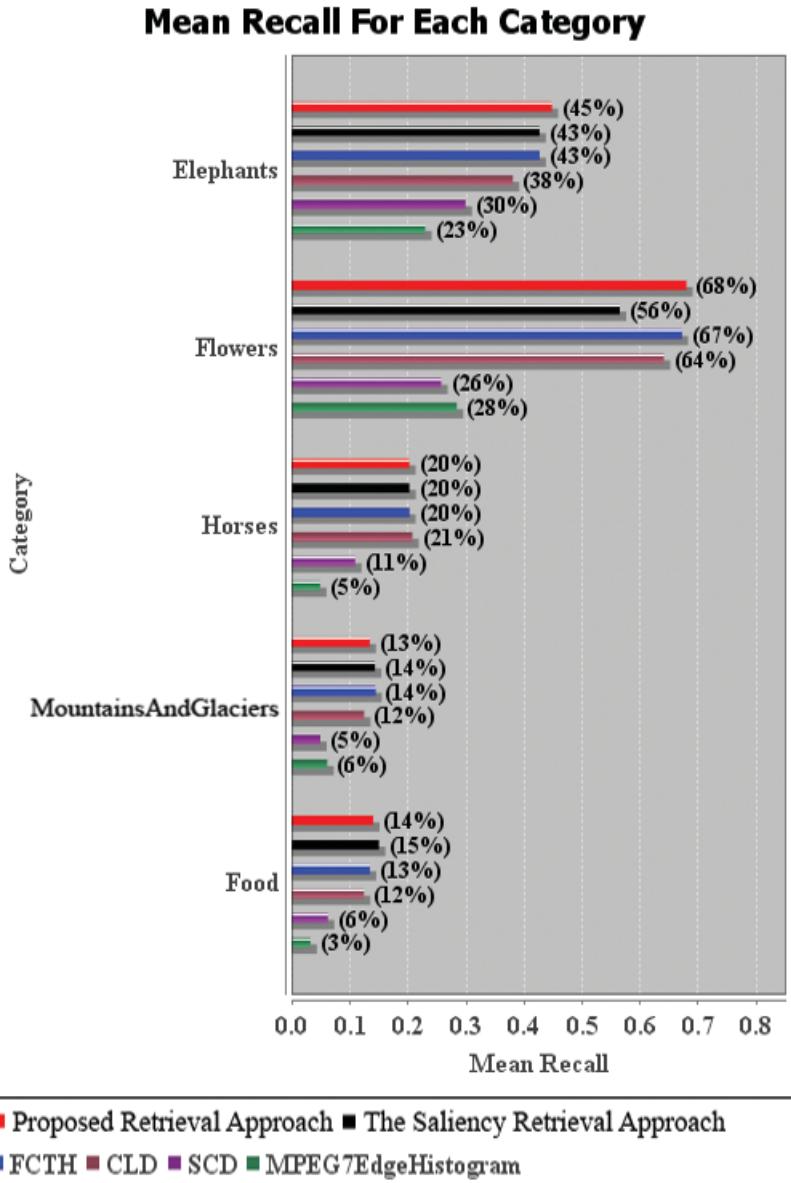


Figure 4.17: The mean recall for last 5 categories of Wang database.

Figure 4.16 and 4.17 show the the mean recall for the ten categories of the Wang database. The figures show the improvement of recall of the proposed approach for categories such as dinosaurs, buses, beach ,flowers, elephants and horses. The proposed approach provides a recall close to the saliency approach for categories such as africaPeopleAndVillages, mountainsAndGlaciers, food and buildings category. These categoris contain images having large complexity in spatial structure, so the saliency approach is better in this situation.

## Chapter 5

# Conclusion and Future Work

Through this study, several improvements in visual image search were provided that increased the quality and accuracy. The study compared current techniques of visual content description and proposed two image retrieval approaches, the salient image retrieval approach and the principal region detection retrieval approach.

The investigated techniques were implemented and used in a CBIR system that use images retrieved from Google keyword based search engine to further enhance the results. According to the research work, Image features need to be integrated to provide more accurate description of image content and better image retrieval accuracy.

The salient features capture locations or regions in an image that have complexity in spatial structure. These locations provide a description of image details. The proposed salient image retrieval approach allows the retrieval of the most precise images having the same salient attributes of a query image. This approach extracted image salient regions and constructed a graph by detecting the spatial relationships between these regions. A modification was performed on the scoring function of the greedy graph matching algorithm that worked as a similarity measure to detect the final image rank. Global image features describe the whole image. Alternatively, local features describe parts or regions of an image. The

advantages of integrating global and local features together have been utilized for better retrieval efficiency. The proposed retrieval approach combines the the fuzzy color and texture histogram (FCTH) global descriptor to enhance the retrieval accuracy. The experimental results showed that this combination provides more accurate results than MPEG7 color and texture descriptors and FCTH composite descriptor. This approach is efficient for images having complex spatial structure as it captures the most complex regions. It has a reduced accuracy for images having large smoothed regions because a few salient regions will be detected.

The proposed general principal regions retrieval approach provides an approximation between smoothness and complexity for images having large occlusion and complexity in structure. A generalized multi-resolution region-based segmentation scheme was presented. This scheme was integrated with color and texture features to provide efficient retrieval in large collection of images. The new segmentation scheme provides efficient segmentation and treats problems existed in previous region based segmentation algorithms such as over segmentation. The approach works by segmenting an image to the most general principal regions that work as local descriptors. A spatial graph is generated by detecting the spatial relationships between regions. A greedy graph matching algorithm with a modified scoring function worked as a similarity measure to detect the final image rank. Results showed that this approach provided precision close to the saliency approach for images having large geometric variations. For images with small variations, this approach provides more accurate results than the saliency approach and other approaches.

The proposed approaches exhibit higher reliability and achieve the goaled robustness. Moreover, they succeeded to provide more accurate results under several real variations of image collections. They can be used as efficient retrieval approaches in personal and web image collections where image descriptors can be saved as XML files for fast indexing.

## Future Work

Automatic image annotation is an extension area for the research work. This area develops computer systems that can assign metadata in the form of tags or captions to an image automatically. the proposed retrieval approaches can help organize and locate images of interest from an image collection.

Story illustration is an area that develops computer systems that provide automated story picturing. In this area both annotations and visual content play an important role to match the semantic keywords extracted from the story. the proposed retrieval approaches can be utilized to provide visual content similarity.

The approaches can provide image signatures that can be used to categorize all pixels in an image into one of several classes. So, they can be used in classification and categorization applications.

A relevance feedback stage can be added to further improve the accuracy of the proposed retrieval approaches. This stage incorporate the user to indicate the relevant and the not relevant results of a query, in order to enhance the returned results next time.

# References

- [1] Jonathon S. Hare and Paul H. Lewis. Salient regions for query by image content. *Image and Video Retrieval: Third International Conference, CIVR, Dublin, Ireland ISSN 0302-9743*, pages 317–325, Jul. 21-23, 2004.
- [2] Prakash K.S.S. and Sundaram R.M.D. Combining novel features for content based image retrieval. *Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. 14th International Workshop on*, pages 373–376, Jun. 27-30, 2007.
- [3] S. Santini A. Gupta A. M. W. Smeulders, M. Worring and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, Dec. 2000.
- [4] J. R. Smith and S.-F. Chang. Visualseek: A fully automated content-based query system. *Proc. 4th ACM Int'l Conf. on Multimedia*, pages 87–98, 1996.
- [5] Stan Sclaroff, Leonid Taycher, and Marco La Cascia. Imagerover: A content-based image browser for the world wide web. In *CAIVL '97: Proceedings of the 1997 Workshop on Content-Based Access of Image and Video Libraries (CBAIVL '97)*, page 2, 1997.
- [6] Myron Flickner Jim Hafner Wayne Niblack Dragutin Petkovic Christos Faloutsos, Ron Barber and William Equitz. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 3(3/4):231–262, 1994.
- [7] Virage Inc. Vir image engine. [www.virage.com/products/image\\_vir.html](http://www.virage.com/products/image_vir.html), 2001.

- [8] Kwok-Wai Cheung Ka-Man Wong and Lai-Man Po. Mirror: an interactive content based image retrieval system. *Proceedings of IEEE International Symposium on Circuit and Systems*, 2:1541–1544, May 23-26, 2005.
- [9] Jose M. Martinez. Mpeg-7 overview. <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>.
- [10] S. A. Chatzichristos and Y. S. Boutalis. Fcth - fuzzy color and texture histogram - a low level feature for accurate image retrieval. *9th International Workshop on Image Analysis for Multimedia Interactive Services, IEEE Computer Society, Klagenfurt, Austria*, pages 191–196, 2008.
- [11] T.-S. Lai. Chroma: a photographic image retrieval system. *PhD thesis, School of Computing, Engineering and Technology, University of Sunderland, UK*, 2000.
- [12] van den Berg J. Subject retrieval in pictorial information systems. In *Proceedings of the 18th International Congress of Historical Sciences. Round Table 34. Electronic Filing, Registration, and Communication of Visual Historical Data*, pages 21–29, 1995.
- [13] Virginia E. Ogle and Michael Stonebraker. Chabot: Retrieval from a relational database of images. *IEEE Computer*, 28(9):40–48, Sep. 1995.
- [14] Content-based image retrieval. <http://en.wikipedia.org/wiki/CBIR>.
- [15] Zhang H. LongF. and Dagan Feng D. Fundamentals of content-based image retrieval,in multimedia information retrieval and management – technological fundamentals and applications. *Springer-Verlag*, pages 1–26, 2003.
- [16] Yu jin Zhang. *Semantic-Based Visual Information Retrieval*. IRM Press, 2007.
- [17] Michael J. Swain and Dana H. Ballard. Color indexing. *International Journal of Computer Vision*, 7:11–32, 1991.
- [18] Shamik Sural, Gang Qian, and Sakti Pramanik. Segmentation and histogram generation using the hsv color space for image retrieval. *Proceedings of the*

- 6th ACM international conference on Image and video retrieval, Amsterdam, The Netherlands, 2:II–589– II–592, 2002.
- [19] Stricker M. and Orengo M. Similarity of color images. In *Proc. of the SPIE Conference*, 2420:381 – 392, 1995.
- [20] G. Lu and J. Phillips. Using perceptually weighted histograms for color-based image retrieval. In *Proc. Of International Conference on Signal Processing*, pages 1150 – 1153, 1998.
- [21] K. Konstantinidis, A. Gasteratos, and I. Andreadis. Image retrieval based on fuzzy color histogram processing. *Optics Communications*, 248(4-6):375 – 386, 2005.
- [22] Stricker M. and Dimai A. Spectral covariance and fuzzy regions for image indexing. In *Machine Vision and Applications*, 10:66 – 73, 1997.
- [23] Greg Pass Ramin Zabih and Justin Miller. Comparing images using color coherence vectors. *International Multimedia Conference Proceedings of the fourth ACM international conference on Multimedia. Boston, Massachusetts, United States*, pages 65 – 73, 1997.
- [24] T. Ojala, M. Rautiainen, E. Matinmikko, and M. Aittola. Semantic image retrieval with hsv correlograms. *Scandinavian Conference on Image Analysis, 2001, Bergen, Norway*, pages 621 – 627, 2001.
- [25] R Barcellos, R. S. Oliani, L. T. Lorenzi, and A. Gonzaga. Content based image retrieval using color autocorrelograms in hsv color space. In *Proceedings Of Xviii Brazilian Symposium on Computer Graphics and Image Processing*, page B5: 153 ff, 2005.
- [26] A. Vadivel, Shamik Sural, and A. K. Majumdar. An integrated color and intensity co-occurrence matrix. *Pattern Recogn. Lett.*, 28(8):974 – 983, 2007.

- [27] Tamura H., Mori S., and Yamawaki T. Textural features corresponding to visual perception. In *IEEE Transactions on Systems, Man and Cybernetics*, 8:460 – 472, 1978.
- [28] Guang-Ho Cha. Kernel principal component analysis for content based image retrieval. *Pacific-Asia conference on advances in knowledge discovery and data mining*, 3518:844–849, 2005.
- [29] Pavel V'acha and Michal Haindl. Illumination invariants based on markov random fields. *International Conference on Pattern Recognition (ICPR) 2008*, pages 1–4, 2008.
- [30] Anil K. Jain and Farshid Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recogn.*, 24(12):1167–1186, 1991.
- [31] K. Muneeswaran, L. Ganesan, S. Arumugam, and K. Ruba Soundar. Texture image segmentation using combined features from spatial and spectral distribution. *Pattern Recognition Letters*, 29(7):905–917, May 2008.
- [32] Abdelhamid Abdesselam. Texture image retrieval using fourier transform. *international conference on communication, computer and power (icccp'09)*, 2009.
- [33] D. Zhang and G. Lu. A comparative study on shape retrieval using fourier descriptors with different shape signatures. *Proc. of International Conference on Intelligent Multimedia and Distance Education (ICIMADE01)*, pages 1–9, Jun. 2001.
- [34] Yong Kui Liu, Wei Wei, Peng Jie Wang, and Borut alik. Compressed vertex chain codes. *Pattern Recognition*, 40(11):2908–2913, 2007.
- [35] Jalba A.C., Wilkinson M.H.F., and Roerdink J.B.T.M. Shape representation and recognition through morphological curvature scale spaces. *IEEE Trans. Image Processing*, 15(2):331–341, 2006.

- [36] El hadi Zahzah. Wavelet transform for partial shape recognition using sub-matrix matching. *VISAPP 2008 - International Conference on Computer Vision Theory and Applications*, pages 513–517, 2008.
- [37] Atul Sajjanhar and Guojun Lu. A grid based shape indexing and retrieval metho. *Special Issue of Australian Computer Journal on Multimedia Storage and Archiving Systems*, 29(2):131–140, Nov. 1997.
- [38] Dengsheng Zhang and Guojun Lu. A comparative study of three region shape descriptors. *Digital Image Computing Techniques and Applications*, pages 21–22, Jan. 2002.
- [39] Dengsheng Zhang and Guojun Lu. Enhanced generic fourier descriptors for object-based image retrieval. *in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 4:3668–3671, 2002.
- [40] Cheng-Chieh Chiang, Yi-Ping Hung, Hsuan Yang, and Greg C. Lee. Region-based image retrieval using color-size features of watershed regions. *J. Vis. Comun. Image Represent.*, 20(3):167–177, 2009.
- [41] N.Sebe, Q.Tian, E.Loupia, M.S. Lew, and T.S.Huang. Salient points for content-based retrieval. *British Machine Vision Conference (BMVC'01)*, pages 401–410, 2001.
- [42] Bernt Schiele and James L. Crowley. Object recognition using multidimensional receptive field histograms and its robustness to view point changes, 1996.
- [43] Cordelia Schmid, Roger Mohr, and Christian Bauckhage. Evaluation of interest point detectors. *Int. J. Comput. Vision*, 37(2):151–172, 2000.
- [44] E. Loupias, N. Sebe, S. Bres, and J.-M. Jolion. Wavelet-based salient points for image retrieval. *In International Conference on Image Processing*, pages 518–521, Nov. 1999.

- [45] Dongxiang Zhou, Yun hui Liut, and Xuanping Cai. An efficient and robust comer detection algorithm. *Proceedings of the 5th World Congress on Intelligent Control and Automation*, June 2004.
- [46] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.
- [47] T. Kadir, A. Zisserman, and M. Brady. An affine invariant salient region detector. In *ECCV04*, pages Vol I: 228–241, 2004.
- [48] Joost van de Weijer, Theo Gevers, and Andrew D. Bagdanov. Boosting color saliency in image feature detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(1):150–156, 2006.
- [49] Qasim Iqbal and J.K. Aggarwal. Combining structure, color and texture for image retrieval: A performance evaluation. *16th International Conference on Pattern Recognition (ICPR)*, 2:438–443, 2002.
- [50] A. Vadivel, M. Mohan, Shamik Sural, and A. K. Majumdar. A color-texture histogram from the hsv color space for video shot detection. In *Proc. of Indian Conference on Vision, Graphics and Image Processing*, pages 435–440, 2004.
- [51] Chuen-Horng Lin, Rong-Tai Chen, and Yung-Kuan Chan. A smart content-based image retrieval system based on color and texture feature. *Image Vision Comput.*, 27(6):658–665, 2009.
- [52] Keinosuke Fukunaga. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press., 1990.
- [53] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.
- [54] Jia Li, James Z. Wang, and Gio Wiederhold. Irm: Integrated region matching for image retrieval. *International Multimedia Conference Proceedings of the eighth ACM international conference on Multimedia*, pages 147–156, 2000.

- [55] Chiou ting Hsu and Ming chou Shih. Content-based image retrieval by interest points matching and geometric hashing. *Proc. Of SPIE Photonics Asia Conference*, 4925:80–90, Oct. 2002.
- [56] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *Int. J. Comput. Vision*, 65(1-2):43–72, 2005.
- [57] Minakshi Banerjee, Malay K. Kundu, and Pradipta Maji. Content-based image retrieval using visually significant point features. *Fuzzy Sets Syst.*, 160(23):3323–3341, 2009.
- [58] Hong Chang and Dit-Yan Yeung. Kernel-based distance metric learning for content-based image retrieval. *Image Vision Computing*, 25(5):695–703, 2007.
- [59] Miguel Arevalillo-Herrez, Juan Domingo, and Francesc J. Ferri. Combining similarity measures in content-based image retrieval. *Pattern Recognition Letters*, 29(16):2174–2181, 2008.
- [60] Sven Siggelkow and Hans Burkhardt. Fast invariant feature extraction for image retrieval. *State-of-the-Art in Content-Based Image and Video Retrieval*, pages 43–68, 2001.
- [61] James Z. Wang, Jia Li, and Gio Wiederhold. Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):947–963, 2001.
- [62] Jules Vleugels and Remco C. Veltkamp. Efficient image retrieval through vantage objects. *VISUAL '99: Proceedings of the Third International Conference on Visual Information and Information Systems*, pages 575–584, 1999.
- [63] Ka-Man Wong, Kwok-Wai Cheung, and Lai-Man Po. Mirror: an interactive content based image retrieval system. *Proceedings of IEEE International Symposium on Circuit and Systems*, 2:1541–1544 , 23–26, May 2005.

- [64] *Img(Rummager): An Interactive Content Based Image Retrieval System*, 2009.
- [65] Praveen Dasigi and C.V. Jawahar. Efficient graph-based image matching for recognition and retrieval. In *Proceedings of National Conference on Computer Vision, Pattern recognition, Image Processing and Graphics*, 2008.
- [66] Timor Kadir and Michael Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.
- [67] J. Stottinger, N. Sebe, T. Gevers, and A. Hanbury. Colour interest points for image retrieval. *Vision Winter Workshop*, 2007.
- [68] N. Sebe, Q. Tian, E. Loupias, M. S. Lew, and T. S. Huang. Evaluation of salient point techniques. *Image and Vision Computing*, 21(13-14):1087 – 1095, 2003.
- [69] Franz Aurenhammer. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, 23(3):345–405, 1991.
- [70] G. Schaefer and M. Stich. Ucid -an uncompressed colour image database. In *Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia*, pages 472–480, 2004.
- [71] Sylvie Philipp-Foliguet, Julien Gony, and Philippe-Henri Gosselin. Frebir: An image retrieval system based on fuzzy region matching. *Comput. Vis. Image Underst.*, 113(6):693–707, 2009.
- [72] Oscar software framework. <http://leanxcam.origo.ethz.ch/wiki/oscar-software-framework-manualdescription>.
- [73] M. Kowalczyk, P. Koza, P. Kupidura, and J. Marciniak. Application of mathematical morphology operations for simplification and improvement of correlation of images in close-range photogrammetry. In *ISPRS08*, page B5: 153 ff, 2008.

- [74] Ji quan Ma. Content-based image retrieval with hsv color space and texture features. *Web Information Systems and Mining, International Conference on*, 0:61–63, 2009.
- [75] Mohamed A. Helala, Mazen M. Selim, and Hala H. Zayed. An image retrieval approach based on composite features and graph matching. *Computer and Electrical Engineering, International Conference on*, 1:466–473, 2009.

# Publications

Mohamed A. Helala, Mazen M. Selim, and Hala H. Zayed. An image retrieval approach based on composite features and graph matching. Computer and Electrical Engineering, International Conference on, 1:466–473, 2009.

Mohamed A. Helala, Mazen M. Selim, and Hala H. Zayed. An Image Retrieval Approach Based On Principal Regions Detection. Submitted to Egyptian Informatics Journal.