CONTENT BASED IMAGE RETRIEVAL BASED ON COMBINATION FEATURES

MEKCHONE BOUNTHANH

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF ENGINEERING IN COMPUTER ENGINEERING

INTERNATIONAL COLLEGE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2013

KMITL-2013-IC-M-006-005

Thesis Title	Content-Based Image Retrieval System Based on Combination Features
Student	Mr. Mekchone Bounthanh
Student I.D	54601151
Degree	Master of Engineering
Program	Computer Engineering
Year	2013
Thesis Advisor	Assoc. Prof. Dr. Boonwat Attachoo
Thesis Co-Advisor	Prof. Dr. Kasuhiko Hamamoto

ABSTRACT

Recent years, there have been witnesses on the rapid growth of digital images due to the increasing power of computing and the fast development of the internet. Because of this tremendous increase of digital images, there is an urgent need of image content description to facilitate automatic retrieval. So a new multimedia application, called content based image retrieval (CBIR) has played an important role to address this urgent issue and attracts more and more attention of researchers. In CBIR, image is described by several low level image features, such as color, texture, shape, special layout or the combination of these features. This thesis investigated a combination of multi-features based image retrieval techniques.

Color, texture and shape features have been the primitive image descriptors in content based image retrieval system. In this thesis, a novel framework for combining and weight all of three i.e. color, texture, and shape features had been proposed. The color feature is extracted by quantifying of YUV color space and the color attributes like a mean value, standard deviation, and image bitmap of YUV color space are represented. The texture features are obtained by the entropy based on the gray level co-occurrence matrix and the edge histogram descriptor of an image. And the Fourier descriptors derived from difference signatures are represented to shape descriptors. Those are also normalized by normalized information distance to adjust distance values into the same level before the linear combination and weight applied to combine all of three features to obtain the similarity as indexing of image. Furthermore, an experimental results indicated a weight variation to achieve higher retrieval efficiency and accuracy.

ACKNOWLEDGMENTS

A Master's thesis writing is a daunting task. While personal perseverance is paramount, the task cannot be fulfilled without a kind and ardent mentor. I would like to take this opportunity to give grateful thanks to Associate Professor Dr. Boonwat Attachoo and Professor Dr. Kasuhiko Hamamoto, who convincingly conveys a spirit of teaching and advising the master dissertation. This dissertation would not have been possible without their guidance and help. In many stages of the writing process, their profound expertise and professional knowledge provides crucial to the technical solutions. The regular discussions with them have helped me to learn not only independent research but also a practice for a professional career.

I gratefully acknowledge the ASEAN University Network/Southeast Asia Engineering Education Development Network (AUN/SEED-Net) for awarding me the scholarship with the financial support for my study at King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand.

I would also like to gratitude and sincere goes also to all Professors, lectures and supporting staffs in the department of Computer Engineering and International College, who always continuously encouraged, helped and gave me a guideline during the whole period of my study at KMITL.

Finally, I would also like to gratitude my parent for their patience, continuous encouragement, guidance, love, and support.

Table of Contents

List of Tables

List of Figures

Figure Page

List of Abbreviations

CBIR : Content-Based Image Retrieval

QBE : Query-By-Example

SCD : Scalable Color Descriptor

HSV : Hue-Saturation-Value

HMMD : Hue-Max-Min-Difference

CSD : Dominant Color Descriptor

CLD : Color Layout Descriptor

TBD : Texture Browsing Descriptor

HTD : Homogeneous Texture Descriptor

EHD : Edge Histogram Descriptor

EMD : Earth Mover's Distance

LLE : Locally-Linear Embedding

MDS : Multidimensional Scaling

MAP : Mean Measures Precision

ANMRR : Average Normalized Modified Retrieval Rank

OPP : Opponent Color Sapce

MTM : Mathematical Transform to Munsell

GLCM : Gray Level Co-occurrence Matrics

GLRLM : Gray Level Run-Length Matrix

NGLDM : Neighboring Gray Level Dependence Matrix

MPEG : Moving Pictures Excerpt Group

IDM : Inverse Difference Moment

ALI : Axis of least inertia

TAR : Triangle-area Representation

AGR : Adaptive Grid Resolution

FDs : Fourier Descriptors

Chapter 1

Introduction

1.1 Introduction

During the last decade there has been a rapid increase in volume of image and video collection. A huge amount of information is available, and daily gigabytes of new visual information are generated, stored, and transmitted. However, it is difficult to access this visual information unless it is organized in a way that allows efficient browsing, searching, and retrieval. Traditional methods of indexing images in databases rely on a number of descriptive keywords, associated with each image. Those manual annotation approaches are subjective and recently, due to the rapidly growing database size is becoming outdated. To overcome these difficulties in the early 1990s, Content-Based Image Retrieval (CBIR) emerged as a promising means for describing and retrieving images. According to its objective, instead of being manually annotated with text-based keywords, images are indexed by their visual content, such as color, texture, shape, and spatial layout.

The importance of content-based retrieval for many applications, ranging from art galleries and museum archives to picture collections, criminal investigation, medical and geographic databases, makes the visual information retrieval one of the fastest growing research fields in information technology. So, in the recent year, many content-based retrieval applications have been created for both research and commercial purposes and attracted more and more attention of researchers. All most of the CBIR research is finding technique to enhance the performance for retrieving more similar image from the image databases of retrieval scheme with developing algorithms exploit the low-level features of the image. Therefore, a content-based image retrieval has been interested and has proposed in this thesis.

1.2 Objective

The purpose of thesis is to investigate new problems and solutions, so as to improve not only the retrieval performance but also the practicality of CBIR system. The specific objective include conducting research on three major issues in the CBIR field, which are 1) feature aggregation for similarity measures, 2) a relevance feedback of the retrieval system, and 3) a retrieval model that incorporates background knowledge.

The first research objective is to investigate the feature aggregation for similarity measures. It is critical to measure image similarity for the purpose of CBIR. Feature aggregation is a promising approach for the similarity measure, which obtain image similarity by combining multiple feature distances measured in individual feature spaces. However, feature aggregation has not been investigated sufficiently in the CBIR field. This thesis addresses three specific problems on feature aggregation: 1) how to choose a suitable feature indexing of visual features before feature aggregation; 2) how to develop a unified framework to unify various aggregation methods; 3) how to effectively exploit the discriminant ability of individual visual features in the aggregation procedure.

The second research objective is to investigate a relevance feedback of the retrieval system. The final objective of content based image retrieval is to find images that are perceptually similar to a given query. In an ideal situation, image retrieval system should have the ability to extract all the relevant semantic features from the image in the same way as a human being does. In this thesis, in addition of a combination multiple low-level visual features, a novel method to assign weights to different features has been proposed for relevance feedback of the retrieval system.

The third research objective is to investigate a retrieval model that incorporates background knowledge. Recent research shows that background knowledge can be incorporated into CBIR schemes to improve retrieval performance. However, it is critical to effectively express the background knowledge of the applications. This thesis explores the expression model of background knowledge for the purpose of CBIR. In particularly, based on information normalized distance to adjust each feature distance before combining feature and weight variation were proposed to improve the retrieval performance more effectively than conventional schemes.

1.3 Image Retrieval System

The prevalence of digital cameras has brought about an exponential growth in the amount of image and videos found on the World Wide Web and in both private and commercial collections. These digital images and videos have found useful application in many fields including business, education, defense and security, to mention a few. The large volume of image and video archives has necessitated the development of efficient methods to enable searching, locating and retrieval. The research community has responded to this challenge and we have witnessed the development of a plethora of techniques and a working system in the last two decades. Image retrieval has become a very active research area since the 1970's, with the thrust from two major research communities: Database Management and Computer Vision.

1.3.1 Image Retrieval Problem

In this computer age, virtually all spheres of human life including commerce, government, academics, hospitals, crime prevention, surveillance, engineering, architecture, journalism, fashion and graphic design, and historical research use images for efficient services. A large collection of image is referred to as an image database. An image database is a system where image data is integrated and stored [1]. Image data include the raw images and information extracted from images by automated or computer assisted image analysis.

The police maintain an image database of criminals, crime scenes, and stolen items. In the medical profession, X-rays and scanned image database are kept for diagnosis, monitoring, and research purposes. Architectural and engineering design, image database exists for design projects, finished projects, and machine parts. In publishing and advertising, journalists create image databases for various events and activities such as sports, buildings, personalities, national and international events, and product advertisements. In historical research, image databases are created for archiving in areas that include arts, sociology, and medicine. For a small collection of images, simple browsing can identify an image. This is not the case for large and varied collection of images, where the user encounters the image retrieval problem, text-based and content-based are the two techniques adopted for search and retrieval in an image database.

1.3.2 Text-Based and Content-Based Image Retrieval

In text-based retrieval, images are indexed using keywords, subject headings,or classification codes, which in turn are used as retrieval keys during search and retrieval [2]. Text-based retrieval is non-standardized because different users employ different keywords for annotation. Text descriptions are sometimes subjective and incomplete because they cannot depict complicated image features very well. Examples are texture image that cannot be described in the text. Textual information about images can be easily searched using existing technology, but requires human to personally describe every image in the database. This is impractical for very large databases, or for images that are generated automatically, e.g. from surveillance cameras. It is also possible to miss images that use different synonyms in their descriptions. Systems based on categorizing images in semantic classes like "horse" as a subclass of "animal" avoid this problem, but still face the same scaling issues [3].

The Content-Based Image Retrieval (CBIR) technique uses image content to search and retrieve digital images. Content-based image retrieval systems were introduced to address the problems associated with text-based image retrieval. Content based image retrieval is a set of techniques for retrieving semantically-relevant images from an image database based on automatically-derived image features [4]. The main goal of CBIR is efficiency during image indexing and retrieval, thereby reducing the need for human intervention in the indexing process. The computer must be able to retrieve images from a database without any human assumption on specific domains (such as texture vs. non-texture, or indoor vs. outdoor).

One of the main tasks for CBIR system is similarity comparison; extracting feature signatures of every image based on its pixel values and defining rules for comparing images. These features become the image representation for measuring similarity with other images in the database. An image is compared to other images by calculating the different between their corresponding features.

1.3.3 Fields of Application

Image retrieval based on content is extremely useful in a plethora of application such as publishing and advertising, historical research, fashion and graphic design, architectural and engineering design, crime prevention, medical diagnosis, geographical information and remote sensing systems, etc. [5]. A typical image retrieval application example is a design engineer who needs to search his organization database for design projects similar to that required by his clients, or the police seeking to confirm the face of a suspected department, before the trademark is finally approved for use, there is need to find out if such or similar ones ever existed. In hospitals, some ailments require the medical practitioner to search and review similar X-rays or scanned images of a patient before proffering a solution.

The most important application, however, is the Web, as a big fraction of it is devoted to images, and searching for a specific image is indeed a daunting task. Numerous commercial and experimental CBIR systems are now available, and many web search engines are now equipped with CBIR facilities, as for example Alta Vista, Yahoo and Google [6].

1.3.4 Principle of CBIR

Content-based retrieval uses the contents of images to represent and access the images. A typical content-based retrieval system is divided into off-line feature extraction and on-line image retrieval. A conceptual framework for content-based image retrieval is illustrated in Figure 1.1 [4]. In off-line stage, the system automatically extracts visual attributes (color, shape, texture, and spatial information) of each image in the database based on its pixel values and stores them in different database within the system called a feature database. The feature data (also known as image signature) for each of the visual attributes of each image is very much smaller in size compared to the image data, thus the feature database contains an abstraction (compact form) of the images in the significant compression of image representation. However, a more important reason for using the signature is to gain an improved correlation between image representation and visual semantics [4].

In on-line image retrieval, the user can submit a query example to the retrieval system in search of desired images The system represents this example with a feature vector. The distance (i.e., similarities) between the feature vectors of the query example, and those of the media in the feature database are then computed and ranked. Retrieval is conducted by applying an indexing scheme to provide an efficient way of searching the image database. Finally, the system ranks the search result and then returns the results that are most similar to the query examples. If the user is not satisfied with search results, he can provide relevance feedback of the retrieval system, which contains a mechanism to learn the user's information needs.

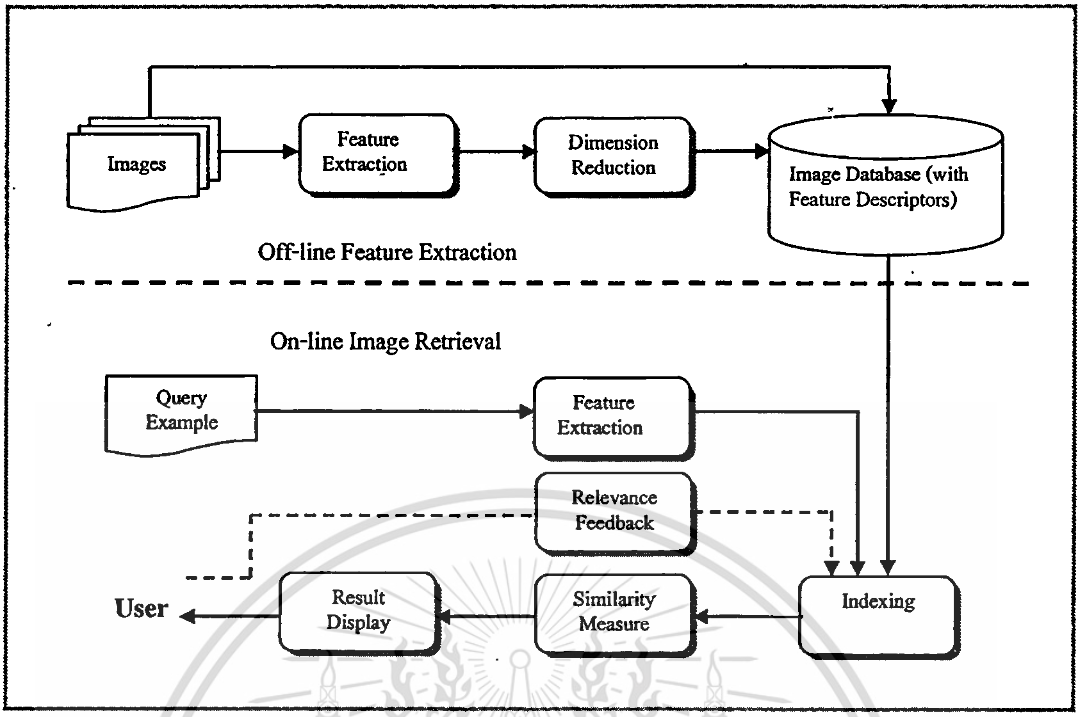

Figure 1.1: A Conceptual Framework for Content-Based Image Retrieval.

1.4 Motivation of This Work

The design and development of effective and efficient CBIR system are still a research problem, because the nature of digital images involves two well-known problems: the semantic gap and the computational load manage large file collection. The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation [7]. It has linguistic and contextual consequences, and mainly depends on the domain knowledge to represent images. On the other hand, the computation load, when large image collections are managed, may make impractical use of CBIR system [8].

The aim of this thesis is to propose a CBIR system; an important task of the system is to reduce the "semantic gap" between low-level image features and the richness of human semantics and to enhance the efficiency and accuracy of the CBIR system. The system used a combination of color, texture, and shape features to extract the features from the images and novel framework to combine those features were normalized by the normalize information distance before combined. The contribution of this work is included:

- Salient low-level features have been a widely using to extract a feature from an image. In many systems, they have been developed algorithms exploit the low-level features of the image such as color, texture, shape, and spatial descriptors of an object to help retrieve images from large databases. Some of them, not only one feature as well represent indexing to measure similarity of the images and to achieve the performance of the retrieval system.

- A set of human perception experiments was designed and conducted to gain insight into human perception of similar images. Analysis of the experimental results by using combination multiple low-level visual features with each feature being weighted differently according to its ability to describe the content of the images.

1.5 Thesis Outline

The rest of this thesis is organized as follows. Chapter 2 reviews the idea behind content-based image retrieval and summarizes some of the related work in the topic of CBIR. The major problems in CBIR paradigm of user query, choice of visual features, design of similarity measures, investigation of image collections, approach of results presentation and measures of performance evaluation. Chapter 3 begins by introducing a color fundamental and described a color attribute to obtain a color feature representing. Chapter 4 presents different texture attributes used in content-based image retrieval and described a feature extraction of gray-level co-occurrence matrix and edge histogram descriptor to obtain the texture feature representing. The different shape attributes and shape feature based on Fourier descriptors are proposed in Chapter 5. The similarity as indexing of the image based on combining and weighting all of three features was described in Chapter 6. The simulation results and evaluation of the each system, as well as the overall systems, are detailed in Chapter 7. Finally, Chapter 8 concludes our work and suggests future work.

Chapter 2

Content Based Image Retrieval: A Literature Review

In content based image retrieval (CBIR) image databases are indexed with descriptors derived from the visual content of the image. Most CBIR systems are concerned with approximate queries where the goal is to find images visually similar to a specified target image. In most cases the goal of CBIR systems is to replicate human perception of image similarity as well as possible.

CBIR presents a challenging problem since it has common elements with both the general image understanding problem (which seems to remain unsolvable for computers at least in the near future) and the field of general information retrieval. Humans excel in image understanding when compared with computers. In contrast, in systematic handling of large databases computers have an edge over us. For this reason CBIR is also potentially very rewarding. This chapter proposed a literature of CBIR and discusses some of the issues related to implementation of CBIR systems. The representation does not even try to be complete or exhaustive. Some important issues are discussed not at all only superficially mentioned.

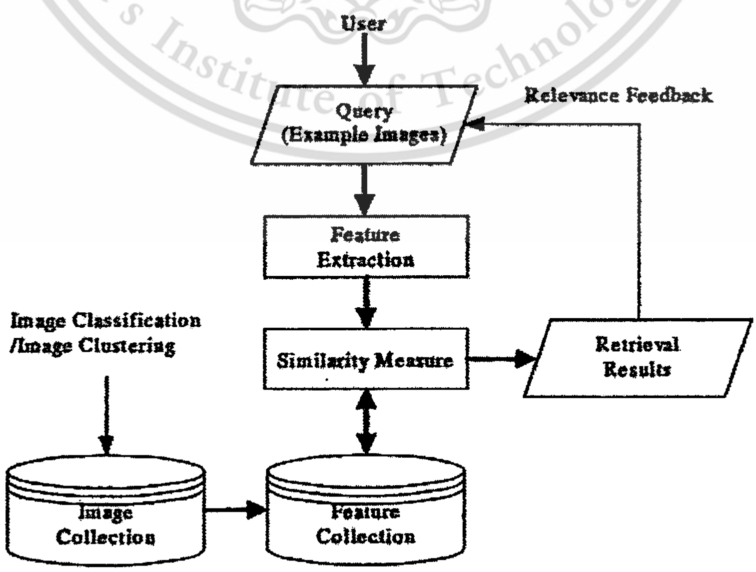

Figure 2.1: A generic system architecture for CBIR.

2.1 A Generic System Architecture of CBIR

Figure 2.1 shows generic system architecture for content-based image retrieval. In a typical retrieval scenario, the user first provides an example image to query the CBIR system. The example image is used to describe the user's information need. To answer the query, the CBIR system searches the image collection for similar images to the example. In the context of CBIR, the similarity is defined on image content which is usually characterized by visual features. Therefore, the CBIR system would extract visual features from the example image. The similarity measurement is performed on the visual features of an example image and that of each image in the collection. The feature extraction for images in the collection is normally conducted beforehand. Finally, the images of the collection are ranked according to their similarities to the query and some top ranked images would be returned as retrieval results. To improve the system performance, some techniques could be introduced into a CBIR system. Relevance feedback is introduced into CBIR systems in order to better capture the user's information need through iterative feedback and query refinement [9]. Image classification and clustering are introduced into a CBIR system to organize unstructured image collections, so as to improve retrieval accuracy and speedup image retrieval [10].

2.2 User Query

How to describe the user's information need and query the retrieval system is an inherent problem of content-based image retrieval. Query-by-example (QBE) is a generally accepted paradigm in the CBIR field [11]. Under the QBE paradigm, the user can use an example to query the CBIR system and expect to obtain similar images to the query example. Given a query, the CBIR system would extract visual features from the example and measure the similarities of images in the collection. Due to the ease of similarity measure, using an example image should be the most representative way [12]. This thesis will adopt the way of query by example image, although other query approaches are reasonable in their applications, such as sketch-based approach [13] and interactive approach [14].

Research has shown that a single example image is unable to sufficiently describe the user's information need and multiple example images are necessary [15]. Multiple example images could be provided by the user at an initial query time [15] or obtained during a number of iterations of relevance feedback [9]. With multiple example images, the retrieval performance can be improved effectively such as by learning a more powerful similarity measure for the specific query. Several research works using multiple example images have been reported and certain success has been achieved in their applications [16].

2.3 Visual Feature

Given an example image as a query, the next question is how to describe the image content. In the context of CBIR, image content is normally described by visual features. The retrieval performance depends on the ability of visual features to describe the image content for a specific application.

Color features, texture features and shape features are three types of visual features well studied and extensively used in the CBIR field [17]. Color features are most widely used in CBIR, which may be ascribed to the superior discriminating potentiality of a three-dimensional domain over the single dimensional domain of gray-level images [11]. Color features can be defined in various color spaces for different applications. Texture features are another set of powerful visual features for CBIR. They could capture the granularity and repetitive patterns of surfaces in an image. In domain specific image retrieval, such as in aerial imagery and medical imaging, texture features are particularly useful due to their close relation to the underlying semantics in these cases [12]. Shape features are also important for CBIR, while they have not been as widely used as color and texture features. A critical problem would be the inaccuracy of segmentation in the extraction of shape features. Shape features can show key attributes of a segmented image region, and their efficient and robust representation could play an important role particularly in object-based image retrieval.

MPEG-7 is a well-known multimedia content description standard. A set of visual descriptors had been tested and included in the MPEG-7 standard [18]. The MPEG-7 color descriptions include a number of histogram descriptors, a dominant color, and a color layout descriptor [18].

The *scalable color descriptor* (SCD) is a color histogram descriptor which is able to capture the color distribution in the image. SCD is defined in the hue-saturation-value (HSV) color space, and uses a Haar transform encoding, in order to address the interoperability issue.

The *color structure descriptor* (CSD) is another color histogram descriptor, which is defined in hue-max-min-difference (HMMD) space to provide information regarding color distribution as well as localized spatial color structure in the image.

The *dominant color descriptor* (DCD) compactly conveys global information regarding the dominant color present in the image. The image is represented as a set of color vecters together with their percentage.

The *color layout descriptor* (CLD) provides information about the spatial color distribution within images. After an image is divided into 64 blocks, CLD is extracted from each of the blocks based on the discrete cosine transform.

The MPEG-7 texture descriptions include texture browsing descriptor, homogeneous texture descriptor and edge histogram texture descriptor [19].

The *texture browsing descriptor* (TBD) characterizes perceptual attributes such as directionality, regularity, and coarseness of texture. A set of scale and orientation selective band-pass filters is applied to filter the image and the filtered outputs are then used to compute the texture browsing descriptor components.

The *homogeneous texture descriptor* (HTD) represents the mean of energy deviation of 30 frequency channels modeled using the Gabor function. The image is filtered by a bank of orientation and scale sensitive filters and the filtered outputs in the frequency domain are used to compute HTD components.

The *edge histogram descriptor* (EHD) captures the edge distribution within an image. EHD extracts edge information from 16 sub-images and categorizes it into five classes for each sub-image: horizontal, vertical, 45 degree, 135 degree and non-directional.

The MPEG-7 shape descriptions include region-based shape descriptor, contour-based shape descriptor and 3D shape descriptor [19].

The *region-based shape descriptor* makes use of all pixels constituting the shape within the image and it can describe the shape with a single connected region and the shape consisting of holes in the objects or several disjoint regions.

The *contour shape descriptor* capture characteristic shape features of an object or region based on its contour. The contour-based shape descriptor uses the curvature scale-space representation which can capture perceptually meaningful features of the shape.

The *3D shape descriptor* provides an intrinsic shape description of 3D mesh models. It exploits some local attributes of the 3D surface.

There are some other visual features proposed for the purpose of CBIR [11], such as salient points and spatial location. Salient points can show the local invariants for image matching. For example, a recent work studied pruned salient point for the purposed of CBIR [20]. Spatial location could be useful particularly in region-based image retrieval [21]. The spatial location and spatial relationship can show the clues for the semantic of an image.

Since each single feature normally describes a specific aspect of image content, multiple visual features are necessary for the general purpose CBIR. Recent research work uses multiple visual features, which can improve the retrieval performance effectively [18]. Although there are a number of good visual features available [11] for CBIR, how to effectively employ multiple visual features in CBIR systems is still an open problem. In particular, a critical problem is to measure image similarity using multiple visual features. This thesis will investigate methods of similarity measure using multiple visual features.

2.4 Similarity Measure

In content-based image retrieval, similarity measure consistent with human perceptual similarity is a fundamental problem because it determines if the system can satisfy the user's information need. A simple way to measure image similarity is to use distance (dissimilarity) measures. In particular, a specific distance measure could be designed for a single visual feature in a certain space to match the perceptual similarity [11, 18]. However, simple distance-based methods are not always effective enough, and more effective and complex method is desirable for CBIR [12].

2.4.1 Distance Measures

In the CBIR field, there are a number of distance measures used for similarity computation, such as the Minkowski metric, Hausdorff distance, K-L divergence, Earth Mover's Distance (EMD), and Integrated Region Matching (IRM). The well-known Euclidean distance

(L_2distance) and Manhattan distance (L_1 distance) are special cases of the Minkowski metric. The weighted Euclidean has been used for the color moments e MARS system [15]. The weighted Euclidean distance between two vectors x and y, can be computed by

$$D_{L2}(x,y) = \sqrt{\sum_j w_j (x_j - y_j)^2} \quad , \tag{2.1}$$

where$w_j$is the weight of the j-th component, and $x_j$and $y_j$are components of x and y, respectively. The Manhattan distance is recommended to compute the distance of two CSD descriptors by MPEG-7 [18]. The Manhattan distance between two vectors x and y, can be computed by

$$D_{L1}(x,y) = \sum_j |x_j - y_j|, \tag{2.2}$$

The Hausdorff distance is applied in region-based image retrieval by Ko and Byun [22]. The Hausdorff distance between two sets of points A and B can be computed by

$$D_H(A,B) = max\left(max_i min_j D(a_i, b_j), max_j min_i D(b_j, a_i)\right), \tag{2.3}$$

where$a_i \in A$, $b_j \in B$, and $D(.,.)$ is any underlying vector distance. The Kullback-Leibler (K-L) divergence is used in the method for measuring similarity of texture features [23]. The KL divergence between two distribution $f()$ and $g()$ can be computed by

$$D_{KL}(f,g) = \int_{-\infty}^{+\infty} f(x) \log\frac{f(x)}{g(x)} dx, D_{KL}(f,g) = \sum_x f(x) \log\frac{f(x)}{g(x)} \quad , \tag{2.4}$$

in the continuous and discrete cases, respectively. The Earth Mover's Distance (EMD) is widely used in region-based image retrieval [24] since it well matches perceptual similarity and can be applied to variable-length representations of distributions. The integrated region matching (IRM) is proposed by Li et al. [25], which can match a region of one image to several regions of another image.

Research has shown that multiple visual features are necessary for the general purpose of CBIR. While simple distance-based methods lead to very efficient computation, they often are not effective enough to be useful [12]. In the case of multiple visual features, it is too ambitious to except a single distance measure to produce meaningful retrieval results for various queries. Different visual features may require different distance measures for similarity matching, such as pairs of visual feature and distance measure recommended by MPEG-7 [18]. Generally speaking, multiple visual features are able to describe the image content more effectively. However, similarity measure based on multiple visual features is a big challenge for CBIR. Recent research tends to develop more effective and complex methods rather than single distance measure.

2.4.2 Feature Fusion

Feature fusion aims to measure image similarity using multiple visual features. There are two feature fusion approaches, early fusion and feature aggregation (also named late fusion [26]). Early fusion uses a unified measure to conduct similarity computation on a unified image representation derived from multiple visual features. It can be based on similarity metric learning, classification formulation and nonlinear manifold learning.

Some methods aim to customize a similarity measure by considering the importance of feature components for a specific query. Based on relevance feedback, early weighting methods use heuristic formulation with empirical parameter adjustment [27]. The basic idea is to assign higher weight to those feature components that well group positive examples and separate the positive and negative. An optimization formulation of the learning process is presented for multiple visual features in the retrieval system [28], which overcomes the difficulties of the well-founded theoretical framework proposed in MindReader. In a region-based scheme [29], each image is represented as a family of fuzzy features derived from color, texture and shape, and image similarity is computed by a unified feature matching measure.

Some methods formulate content-based image retrieval as a classification problem [30] and achieve similarity measure by a classified. For example, based on relevance feedback, content-based image retrieval can be treated as a binary classification problem, where the positive class consists of all relevant images to the query and the negative class includes all irrelevant images. The positive and negative examples obtained during relevance feedback can be used to train a binary classifier such as SVM [31]. Then, the decision functions could be used to measure the image similarity. To address the problem that only positive examples are available, the problem of CBIR was treated as a one-class problem and solved by One-Class SVM [32]. It is reasonable to treat all relevant images in a class, however the negative images are various which cannot form a class. To address this problem, the problem of CBIR was cast as a (1+x)-class classification problem in Zhou and Huang [33]. Moreover, Peng explored the approach which treats the problem of CBIR as a multiclass classification problem [34]. Recent work focused on addressing some practical problems in the way that treats CBIR as a classification problem. Tao et al. proposed CBIR schemes to address the problem of small training sample size and the problem of asymmetric training samples [35]. Wang et al. explored how to optimize kernel parameters in CBIR schemes employing Kernel-based biased discriminant analysis [36].

Some other methods are aimed to search for a nonlinear manifold with which visual perception may be achieved better than the original linear space [12], such as by locally-linear embedding (LLE) and multidimensional scaling (MDS). Then, similarity measurement could be conducted in the new feature space more effectively. The ranking on data manifolds is firstly explored by Zhou et al. [37]. He proposed an incremental semi-supervised subspace learning method for content-based image retrieval by employing locality preserving projection [38]. A multi-resolution manifold distance for invariant image similarity is presented by Vasconcelos and Lippman [39]. He et al. proposed a manifold ranking method for image retrieval [40], which could incorporate the information obtained from relevance feedback. Yu and Tian proposed an optimal non-linear subspace projection to capture the most important properties of the subspaces with respect to classification [41]. Recently, He et al. proposed maximum margin projection [42] to maximize the margin between positive and negative examples at each local neighborhood, which is more suitable for image retrieval.

In contrast to early fusion, feature aggregation adopts a different approach, which combines multiple feature distance to obtain image similarity. Normally each visual feature requires a specific measure to match visual similarity because different features normally describe different aspects of image content. Feature aggregation can take this observation into account. A review of feature aggregation is provided in Section 2.7.

2.5 Image Collection and Result Presentation

The scope of an image collection is related to its applications and can influence the complexity of CBIR system design. The categories of image collections include personal collection, domain-specific collection, enterprise collection, archives and web [12]. Various categories would be different in size, storage and type of data (homogeneous or heterogeneous).

These would be taken into account when developing CBIR techniques for a specific application. In particular, a common consideration in the early research is that image collections are unstructured and retrieval is based on a single similarity measure.

Since a single similarity measure is hard to produce and meaningful ranking of image, more efforts have been made to learning-based techniques, such as supervised image classification and unsupervised image clustering [12]. Image classification [43] and image clustering [44] can be used to organize unstructured image collections. Each image class can use its own similarity measure. Based on the image classes constructed by classification or clustering, only the images that belong to the relevant image class as predicted for the query will be ranked and returned as retrieval results. In a result, image classification/clustering can speed-up retrieval and improve accuracy effectively. However, clustering and classification for the purpose of CBIR received relatively less attention in the early years of research [12].

The presentation of retrieval results could influence user satisfaction of CBIR systems. The most popular way to present retrieval results isbased on the relevance to the query. Images in a collection would be ranked according to their similarities measured on visual features. Other ways include class-based [45] and hierarchical [46]. For example, image clustering could be performed on the highly ranked images by conventional metadata-based method and the retrieval results are presented based on image clusters [45]. Generally speaking, the different ways for presentation of retrieval results should be combined with proper CBIR techniques.

2.6 Evaluation, Systems and Applications

Subjects reviewed in this section include retrieval performance evaluation, real-world CBIR systems and existing/potential applications. In particular, a new research topic in the CBIR field, the security of CBIR systems, will be mentioned.

2.6.1 Performance Evaluation

In the research of CBIR, performance evaluation involves three aspects, evaluation dataset, ground truth and evaluation measures.

There are several image datasets used for CBIR evaluation, such as Corel Stock Photos, Caltech101 [47], Image CLEF, and TRECVID. Corel pictures are most popular for performance evaluation in CBIR. Caltech101 include 101 picture categories, which has been extended to 256 picture categories [48]. Image CLEFis atrack as part of a cross-language evaluation forum, which also provides an image dataset consisting of over 20,000 pictures for performance evaluation of CBIR [50]. The TRECVID benchmark is another popular one in the CBIR community to validate their search and retrieval algorithms.

The construction of ground truth is a practical problem. There are some techniques for creating ground truth [49], such as use of collection with predefined subsets, image grouping, and user judgment. An advantage of using collections with predefined image groups is the ease of obtaining relevance judgments. In contrast, image grouping or use judgment requires the experts or real users to explore the entire image collection, which is normally time-consuming particular for large-scale collections. However, grouping by the collection itself is not always based on visual similarity. Groups may include semantic and visual images. To create ground truth for visual content based retrieval, group should be cleaned to only contain visually similar images. A common technique would be cleaning the predefined subsets by using domain expert knowledge.

Precision and recall are two of the most evaluation measures [49]. Precision is defined as the number of retrieved relevant images divided by the total number of retrieved images. Recall is defined as the number of retrieved relevant images divided by the total number of existing relevant images. Precision and recall have an inverse relationship. Image retrieval results are usually reported by precision-reverse curves. Another popular measure is average precision, which is defined as the mean of all the individual precision scores. Average precision can show the effect of the respective rankings of relevant images in a ranked image list. Furthermore, mean average precision (MAP) can be defined for a set of queries, which is the mean of the average precision for each query. MAP has been shown to have good discrimination and stability. Other measures used in CBIR include such as average normalized modified retrieval rank (ANMRR) [18]. The recent comprehensive overview and discussion on performance measures for CBIR refer to [50]. To show statistical significance, the retrieval performance is computed over a number of queries. Normally, the number of queries is range from 100 to 1000 [45]. The size of the image collection is from 1,000 to 20,000[43]. For large-scale retrieval, the size of the collection can up to 80 million [51].

2.6.2 Systems and Applications

The existing applications of CBIR would be for art culture, medical, personal and the web [52]. In different applications, users may have different interests, which would result in various requirements for CBIR techniques. For example, in the applications for art and culture [53], many art objects have distinct color, texture and shape patterns. In contrast, in the applications for medical [52], a medical professional is interested in a dark round area in a lung X-ray which may mean a particular pathology. The potential applications would be in several areas, such as consumer world, public safety and the professional world. For example, in the professional world, CBIR techniques can provide an easy way for people to retrieve, browse and skim their interested content in their recorded meeting archive [54].

There are a number of CBIR systems developed in academia, such as MARS [15], SIMPLIcity [55] and Picsom [56]. A comprehensive review of early CBIR systems is provided in Veltkamp and Tanase [57]. Recently, two well-known commercial search engines, Google Image Search and Microsoft Bing Image Search, also support to 'find similar images' or 'show similar images' for a specific image. They allow users to query by a single image and return similar images by combining the results of metadata-based image retrieval and content-based image retrieval. CBIR systems are being developed in various domains such as family album management, botany, astronomy, mineralogy and remote sensing [12]. However, not many CBIR systems are deployed for public usage [12], which also becomes motivation for the work on CBIR.

An observation is there are a few efforts addressing security concerns. However, security would be a critical aspect of practical CBIR systems. With the popularity of visual information retrieval in ordinary users, security is becoming more and more important. For example, the copyright protection would be an important security problem in CBIR systems. In medical applications [58], a medical image is used as a query which carries the user's health information and should not be distributed without authorization. This problem is called the user's rights protection.

2.7 Feature Aggregation for Similarity Measure

Feature aggregation is a feature fusion approach, which combines multiple feature distance (dissimilarities) measured in different feature spaces to obtain image similarity. The advantage of feature aggregation is that specific distance metrics can be applied to different visual features. A distance metric can be specifically designed by tanking into account the nature of a visual feature, so as to well visual perceptual similarity. Some work on feature aggregation [15, 55] has been reported in the literature, while there are some unsolved problems.

First, existing feature aggregation methods are superior in different situations [15, 59, 60]. In SIMPLIcity, the similarity between two image regions is obtained by combining multiple feature distances using a heuristic converting function [55]. The aggregation method could be based on simple fixed fusion operations [59], such as Min, Max and Sum. The simple fusion operations are effective in some applications, but they do not take into account the different importance of visual features for various queries. The linear combination was proposed to aggregate multiple feature distance for the purpose of CBIR, which can set the weighting of visual features to describe the importance [15]. In another way, the Boolean logic was applied to combine multiple feature distance in the MARS system and Blobworld [60]. To overcome the limitations of traditional Boolean logic, the combination was treated as a decision problem and solved by a fuzzy logic-based decision method [15]. An unsolved problem is how to select a proper aggregation method to better satisfy a user's information need.

Second, the combination function for feature aggregation is normally learned from multi-example queries. Rui et al. [15] heuristically adjust a linear combination function according to the example images obtained through relevance feedback. An optimization method for the combination function was presented later [28]. A genetic programming framework was proposed to combine MPEG-7 descriptor similarities for the purpose of CBIR. These methods would be applied for all examples in the query. Recently, a localized learning method [61] for the combination function was proposed. It is based on an observation that a feature descriptor may be not equally important to different training example images. This method was proposed for image classification based CBIR scheme. Some CBIR systems may not employ image classification. In the CBIR applications with unstructured image collections, the system should search relevant images according to the example images provided by the user. A new problem is how to design a localized learning method for retrieval by multi-example queries in the CBIR systems without employing image classification.

Third, most existing feature aggregation methods [15, 60, 55] such as linear combination [15], and fuzzy logic-based method [53], adopt the parallel parading. In parallel feature aggregation, the function of individual visual features is limited sine the retrieved images are determined only by the aggregated similarity. The irrelevant images can seriously affect the retrieval performance since all images in the collection are used for ranking. In contrast to the parallel paradigm, the series paradigm can utilize individual visual features in sequence. The series feature aggregation is a novel idea and is investigated in this thesis.

2.8 Retrieval Model by Incorporating Backgroung Knowledge

As mentioned in Section 2.2, a single image is unable to adequately characterize the user's information need so that multiple example images are described [15, 16, 62]. For instance, a user provides an image of 'red car' as a query, it is difficult to determine the user's information need, whether the user is looking for images of 'red car' or images of 'car' regardless of colors?. In contrast to a single example query, the relationships of visual content in a multi-example query could be explored to reveal more on the user's information need. For instance, a user provides images of a 'red car' and other cars with various colors, it could be inferred that the user is looking for images of cars regardless of colors. In the above examples, the words, 'red car' and 'car', is used to describe the visual content of images perceived by people, not the semantic meaning of the car and colors.

Most work on CBIR with multi-example queries aims to improve the retrieval performance by selecting or developing effective combination methods. A number of retrieval experiments showed that a single image is not sufficient to express the user's query and multi-example queries are superior [62]. Jin and French [63] proposed a retrieval scheme to merge multiple image list ranked according to example images instead of finding a query center. Natsev and Smith [64] proposed and investigated some techniques for active selection of query examples and query features. An empirical study showed that the performance of lightweight methods using simple combination rules is comparable to heavyweight SVM methods [64]. Zhu and Zhang [65] presented and evaluated a variety of result combination strategies which includes linear distance combination and non-linear result combination, in order to improve retrieval results using multi-example queries. Abbadni [66] proposed a two-level approach for multi-example queries, in which a linear model was used to fuse multiple representations and a round robin scheme was used to fuse multiple ranking lists obtained from example images. Westerveld and Vries [67] proposed two generation models, query generation and document generation, to theoretically model image retrieval using multiple example images.

Relevance feedback can be viewed as a way of supplying multiple example images by involving users in the retrieval process [15]. Some computing techniques used in relevance feedback, where multiple selected images are handled, could be adapted to process images supplied as examples in multi-example queries, for example, the weight optimization method [28]. In this approach, some difference should be considered in developing retrieval scheme with multi-example queries. In conventional CBIR schemes with relevance feedback [15, 30], the user provides a single image as a query and the system returns the initial retrieval results. Then, the user is required to label some returned images as relevant or irrelevant. The system automatically adjusts internal parameters based on the user's feedback. A session labeling returned images requires several iterations.

In contrast, the retrieval schemes with multi-example queries aim to search relevant images using multiple example images provided by the user as a query in a session with no iterations. From the sampling point of view [68], which only relevant example images are concerned, the approaches of relevance feedback and multi-example querying are different. The approach of obtaining relevant example images using relevance feedback can be viewed as a kind of cluster sampling, because the labeled relevant image would be clustered around the initial query in some way which is determined by the search strategy of CBIR. However, the approach of multi-example querying can be viewed as a kind of simple random sampling, as the example images provided by the user have certain diversity in low-level feature spaces [63].

Although the retrieval schemes for multi-example queries differ, all of them are built upon the same assumption that the relevant images in a multi-example query are provided correctly by the user and exactly express the user's information need. However, in practical CBIR systems, this assumption may not always hold. For instance, an ordinary user nay makes mistakes or supplies some example images that do not precisely characterize the user's query.

2.9 Chapter Summary

This chapter provided a short review of content-based image retrieval, which involves user query, visual features, similarity measures, image collections, evaluation, systems and applications. The focus has been put on two topics: feature aggregation for similarity measures and a retrieval model by incorporating background knowledge. Several research problems were also pointed out and discussed in the review. In summary, there are two importance research topics which will be investigated in this thesis.

- The first research topic is feature for similarity measures. In feature aggregation, the aggregation methods can use heuristic functions, machine learning, or parallel model. There are three specific research problems. First, how to adjust the aggregation function according to different queries. Second, how to learn a local aggregation functioned for each query example. Third, how to sufficiently utilize the discriminant power of individual visual features.
- The second research topic is about retrieval model. Background knowledge can be incorporated into unsupervised multi-example queries and relevance feedback to further improve the retrieval performance in the terms of accuracy and efficiency. A specific research problem is how to supply the user's information need.

Chapter 3

Feature Extraction

Feature extraction means obtaining useful information that can describe the image with its content. We mean by image features with the characteristic properties. For example, the image of a beach can be described by its color and some texture of seas and beaches. Objects in the image can be considered as shapes that can be a feature of the image. To describe an image, we have considered its main features that selecting image features are an important step to represent the content of the image as well. Color, texture, and shape are some features considered for content image description. In this chapter, we will introduce three main features of the feature extraction.

3.1 Color Feature Extraction

3.1.1 Color Fundamental

Color is the brain reaction to a specific visual stimulus. Although we can precisely describe color by measuring its spectral power distribution (the intensity of the visible electro-magnetic radiation at many discrete wavelengths) this leads to a large degree of redundancy. The reason for this redundancy is that the eye's retina samples color using only three broad bands, roughly corresponding to red, green and blue light. The signals from these color sensitive cells (cones), together with those from the rods (sensitive to intensity only), are combined in the brain to give several different "sensations" of the color.

- Brightness: the human sensation by which an area exhibits more or less light.
- Hue: the human sensation according to which an area appears to be similar to one, or to proportions of two, of the perceived colors -- red, yellow, green and blue.
- Colorfulness: the human sensation according to which an area appears to exhibit more or less of its hue.

- Lightness: the sensation of an area's brightness relative to a reference white in the scene.
- Chroma: the colorfulness of an area relative to the brightness of a reference white.
- Saturation: the colorfulness of an area relative to its brightness.

The tri-chromatic theory describes the way of three separate lights, red, green and blue, can match any visible color – based on the eye's use of three color sensitive sensors. This is the basis of on which photography and printing operate, using three different colored dyes to reproduce color in a scene. It is also the way that most computer color spaces operate, using three parameters to define a color.

3.1.2 Color Spaces

A color space is a method by which we can specify, create and visualize color. As humans, we may define a color by its attributes of brightness, hue and colorfulness. A computer may describe a color using the amounts of red, green and blue phosphor emission required to match a color. A printing press may produce a specific color in terms of the reflectance and absorbance of cyan, magenta, yellow and black inks on the printing paper. A color is thus usually specified using three co-ordinates, or parameters. These parameters describe the position of the color within the color space being used. They do not tell us what the color is, that depends on what color space is being used.

One can describe color spaces using the notion perceptual uniformity. Perceptually uniform means that two color spaces which are equally distant in the color space are perceptually equally distant. Perceptual uniformity is a very important notion when a color space is quantized. When a color space is perceptually uniform, there is less chance that the difference in color value due to the quantization will be noticeable on a display or on a hard copy.

- **RGB Color Space**

The first candidate color space is the RGB color space. The RGB color space is a widely used for image display, there are consists of three color components red, green, and blue. Which is called "additive primaries" since a color in RGB space is produced by adding them together. However, the RGB space is not perceptually uniform. As such, the proximity of colors in RGB color space does not indicate color similarity. Since RGB is complete, other complete color spaces are generated by transforming RGB using T_c, which is not necessarily a linear transform.

- **Linear Transforms Color Spaces**

Several linear color transformations are in wide use for representing, transmitting and broadcasting color image and videos [69]. Many of these transforms are computed by simple matrix multiplication.

OPP color space: the opponent color space (OPP) is obtained by linear transformation of the RGB color space [70]. There is evidence that human color vision uses an opponent-color model by which the responses of the R, G, and B cones are combined into two opponent color pathways. The transformation from RGB to OPP is obtained by $w_c = T_c^{OPP} v_c$, where

$$T_c^{OPP} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & 2 \\ 1 & -2 & 1 \end{bmatrix}, \quad (3.1)$$

The advantage with the OPP color space is that it is obtained easily by T_c^{OPP}. The disadvantages are that it is not uniform or natural. The color distance in OPP color space does not provide a robust measure of color dissimilarity. One component of OPP is a luminance channel. The two chrominance channels do no correspond to hue and saturation, but rather to, blue *vs.* yellow and red *vs.* green.

Linear color transforms standards: the YIQ, YUV, and YCrCb linear color transforms have been adopted in recent color picture system. Each of these produces a linear transform for RGB which generates one luminance channel and two chrominance channels. The transformations were designed specifically to the parameters of the expected display devices: YIQ NTSC color television, YUV PAL and SECAM color television, and YCrCb color computer display. None of these color spaces is uniform. As such, the color distances in these transform color spaces do not correspond to color dissimilarities.

YIQ color transform: the YIQ color space was developed for the NTSC composite color TV standard [71]. The transformation from RGB to YIQ is obtained by $w_c = T_c^{OPP} v_c$, where

$$T_c^{YIQ} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.528 & 0.312 \end{bmatrix}$$

$$, \tag{3.2}$$

YCrCb color transform: the YcrCb color space is used in the JPEG digital image standard. The transformation from RGB to YCrCb is obtained by $w_c = T_c^{OPP} v_c$, where

$$T_c^{OPP} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix}, \tag{3.3}$$

- **CIE Color Spaces**

Towards the goal of deriving uniform color spaces, the CIE * in 1976 defined the CIE 1976 ($L^*u^*v^*$) and CIE 1976 ($L^*a^*b^*$) color spaces [73]. These are generated by transforming the RGB color space first by linear transformation to the XYZ color space followed by one of two nonlinear transformations: to CIE 1976 ($L^*u^*v^*$) color space or to CIE 1976 ($L^*a^*b^*$) color space. The first linear transformation between RGB and XYZ is given by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.290 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \tag{3.4}$$

The nonlinear transformations from XYZ to CIE 1976 ($L^*a^*b^*$) and CIE 1976 ($L^*u^*v^*$) are given in [73]. "A" which is determined by relation to a nominally white object color stimulus gives the tri-stimulus values X_n, Y_n, Z_n. In this case, the lightness is given by

$$L^* = 116\left(\frac{Y}{Y_n}\right)^{1/3} - 16 \quad , \tag{3.5}$$

The values for u, v and a, b are given as follows:

CIE 1976 ($L^*u^*v^*$):

$$\begin{aligned} u^* &= 13L^*(u' - u_n') \\ v^* &= 13L^*(v' - v_n') \end{aligned} \quad , \tag{3.6}$$

where

$$\begin{aligned} u' &= \frac{4X}{X + 15Y + 3Z} u_n' = \frac{4X_n}{X_n + 15Y_n + 3Z_n} \\ v' &= \frac{9Y}{X + 15Y + 3Z} u_n' = \frac{9X_n}{X_n + 15Y_n + 3Z_n} \end{aligned} \quad , \tag{3.7}$$

The color distance between two color stimuli is calculated from

$$\Delta E^*_{uv} = [(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2]^{1/2}, \tag{3.8}$$

CIE 1976 ($L^*a^*b^*$):

$$a^* = 500\left[\left(\frac{X}{X_n}\right)^{1/3} - \left(\frac{Y}{Y_n}\right)^{1/3}\right]$$

$$b^* = 200\left[\left(\frac{Y}{Y_n}\right)^{1/3} - \left(\frac{Z}{Z_n}\right)^{1/3}\right] \quad , \tag{3.9}$$

The color distance between two color stimuli is calculated from:

$$\Delta E^*_{ab} = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2}\,, \tag{3.10}$$

The CIE color spaces represent with equal emphasis, three color variants that best perceptually characterize color: hue, lightness and saturation. However, the CIE color spaces are inconvenient due to the necessary non-linearity in forward and reverse transformations with RGB space.

- **HSV Color Space**

Another form of the hue, lightness and saturation transform from RGB to HSV is given in [69]. The transform to HSV is non-linear, but it is easily invertible. The HSV color space is natural and approximately perceptually uniform. Therefore, we can define a quantization Q_c of HSV to produce a collection of colors that is also compact and complete. The HSV transformation T_c from RGB to HSV is accomplished though the following equations [69]: let the color triple $v_c = (r, g, b)$ be the to a color point in RGB space and let's $w_c = (h, s, v)$ be the to the transformed color point in HSV color space, such that $w_c = T_c(v_c)$, then the convergence formulas are as follows:

$$H = \cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right\}, \tag{3.11}$$

$$S = 1 - \frac{3}{R+G+B}[min(R,G,B)]\,, \tag{3.12}$$

$$V = \frac{1}{3}(R+G+B) \quad , \tag{3.13}$$

- **YUV Color Space**

Different colors spaces are better for different application for example some equipment has limited factors that decide the size and type of color space that can be used. Some color spaces are perceptually linear, i.e. a one-unit change in stimulus will produce the same change in perception wherever it is applied. Many color spaces, particularly in computer graphics, are not linear in this way. Among these color spaces, the YUV color space is bit unusual. The vast majority of DVDs already store information in the YUV color space. The Y component decides the brightness of the color (referred to as luminance or luma), while the U and V range from -0.5 to 0.5 (or -128 to 127 in signed digital form representation, or 0 to 255 in unsigned form). Some standards moreover limit the ranges, so the out-of-bounds values indicate special information like synchronization. Since the human eye is much responsive to brightness than it is to color, many loss image compression formats throw away half or more of the sample in the Chroma channel to reduce the amount of data to deal with, without severely destroying the image quality. Therefore, only the Y component has used for our preliminary study. There are a lot of slightly different formulas to convert between YUV and RGB. The only major difference is a few decimal places. The following equations are used to convert from RGB color space to YUV spaces:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \tag{3.14}$$

3.1.3 Color Image Features Representing

Color is an important feature for image representation which is widely used in image retrieval. Many researchers have used color feature to depict image contents for region matching, semantic categorization, and similarity searches. However, the performance is poor for a CBIR system, which has only used a color feature to search for similar images from a huge database.In this proposed, we describe the global and local features of the color descriptor to obtain the color features: the mean value and the standard deviation have been used to extract the global feature from an image and the image bitmap as the local to describe the local characteristics of the image.

- **Global Feature Extraction**

Each image in an image database may be different from all others, but at the same time all images may share certain common characteristics. Hence, we need the statistical description of images to capture these common characteristics and use them to represent an image with fewer bits. The color distribution of pixels in an image contains sufficient information. The mean value of pixel colors states the principal colors of the image, and the standard deviation of pixel colors can depict the variation of pixel colors [74]. The variation degree of pixel colors in an image is called the color complexity of the image. We can use these two features to represent the global properties of an image.

In the proposed, each pixel of a color image is represented by a vector

$$P_i = \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} , \qquad (3.15)$$

Where P_i is the *i*th pixel of the image, $1 \leq i \geq M$, the size of the image is M and the components of P_i are the color space components of the color image.

The mean value (μ) and the standard deviation (σ) of the color image are determined as follows:

$$\mu = \frac{1}{M}\sum_{i=1}^{M} P_i , \qquad (3.16)$$

$$\sigma = \left[\frac{1}{M-1}\sum_{i=1}^{M}(P_i - \mu)^2\right]^{1/2}, \qquad (3.17)$$

where $\mu = [\mu X, \mu Y, \mu Z]^T$ and $\sigma = [\sigma X, \sigma Y, \sigma Z]^T$, each component of μ and σ indicates the color space information, respectively.

- **Local Feature Extraction**

In addition of the global property of an image, the local color properties in an image play also an important role to improve the retrieval performance. Hence, a feature called image binary bitmap can be used to capture the local color information of an image. The basic concept of image binary bitmap comes from the block truncation coding [75], which is a relatively simple image coding technique and has been successfully employed in many image processing applications. There are three steps to generate the image binary bitmap. This method first divides an image into several non-overlapping blocks. Let $B_j = \{b_1, b_2, \dots, b_k\}$ be the jth block of the image, where 1 ≤j≥ m; k represents the total number of pixels in the block, an m is the total number of blocks in the image. The second step is to compute the mean value for each block. Let μ_{B_j} be the mean value of the block B_j, which is defined as follows:

$$\mu_{B_j} = \frac{1}{k}\sum_{i=1}^{k} b_i \ , \tag{3.18}$$

Where $\mu_{B_j} = \left[\mu X_{B_j}, \mu Y_{B_j}, \mu Z_{B_j}\right]^T$ and the components of μ_{B_j} are the color space components. The final step is the scheme to compare each pixel of the block with the mean value of the block μ_{B_j} and summarized the normalized value of each block to determine the characterization of the block B_j for generating the image binary bitmap. Let $I = [IX, IY, IZ]^T$ be the image binary bitmap. Each component in I is expressed as $IX = [IX_1, IX_2, \dots, IX_m]$, $IY = [IY_1, IY_2, \dots, IY_m]$, and $IZ = [IZ_1, IZ_2, \dots, IZ_m]$, respectively. The entries are represented by

$$IX_j = \begin{cases} 1, \text{if } X_i \geq \mu X_{B_j} \\ 0, \text{ otherwise} \end{cases} , \tag{3.19}$$

$$IY_j = \begin{cases} 1, \text{if } Y_i \geq \mu Y_{B_j} \\ 0, \text{ otherwise} \end{cases} , \tag{3.20}$$

$$IZ_j = \begin{cases} 1, \text{if } Z_i \geq \mu Z_{B_j} \\ 0, \text{ otherwise} \end{cases} , \tag{3.21}$$

3.2 Texture Feature Extraction

3.2.1 Introduction to Texture

Although no formal definition of texture exists, there are several of intuitive properties of texture, which are generally assumed to be true [76]. Texture is a property in the area therefore its definition must involve gray values in a spatial neighborhood. The size of the neighborhood depends on the texture type, or the size of the primitives defining the texture. Texture involves also the spatial distribution of gray levels, therefore two-dimensional histograms or co-occurrence matrices are reasonable texture analysis tools. There are several properties such as coarseness, contrast, directionality, and smoothness, which play an important role in describing texture. Coarseness measures texture scale (average size of regions that has the same intensity), contrast measure vividness of the texture (depends on the variance of the gray-level histogram), and directionality gives the eventual main direction of the image texture. Formulas to calculate these attributes can be found in [77]. Texture analysis is important, since the texture is useful in various applications such as automated inspection, medical image processing, remote sensing, defect detection, and similarity evaluation. They include: recognition of image regions using texture properties, known as texture classification, recognition of texture boundaries using texture properties, known as texture segmentation, texture synthesis, and generation of texture from known texture models.

3.2.2 Texture Attributes

Many applications in image processing and computer vision depend on texture images. They use them for recognition of image regions and extracting image shapes using texture properties, recognition of texture boundaries using texture properties, and generation of texture images from texture models. Therefore, since there is no accepted mathematical definition for texture, many different methods are proposed for computing texture. Among these methods, no single method works best with all types of texture. Some common methods are used for texture feature extraction such as statistical, model-based, and transform-based methods.

Model-based methods try to generate a model for the process that generates the texture. To create a model, part of image model parameters are assumed according to some hypotheses, and other model parameters are estimated so that the model works in the best way. Currently, there are popular models as Markov models, fractals, and the multi-resolution autoregressive features.

Transform-based model use the frequency content of the image to determine texture features. These methods aim to represent images in mathematical models. For example, Fourier transform is used to describe the global frequency content of the image. It converts the signal into sine waves of various frequencies. Multi-resolution transforms such as wavelet and Gabor use a window function to transform a signal to a shifter and scaled vision of the original signal. Texture Descriptors in MPEG-7 is an ISO/IEC standard developed by Moving Picture Excerpts Group (MPEG), which aims to create a standard for describing the content of multimedia data [78]. The MPEG-7 has introduced three texture descriptors: homogeneous texture, texture browsing, and edge histogram [79]. The *homogeneous texture descriptor* (HTD) provides a precise quantitative description of a texture that can be used for search and retrieval [78, 79]. In order to obtain an image signature, first the image is filtered with a bank of Gabor filters with five scales and six orientations. The first and second moments of energy in the frequency bands are, then, used as the components of the descriptor. The *texture browsing* descriptor categorizes texture in terms of regularity (highly regular, regular, slightly regular, irregular), coarseness (fine, medium, coarse very coarse), and directionality $(0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ)$ similar to a human characterization. It can be used for browsing type of application, but it may not be appropriate for similarity ranking. One possible scheme for similarity retrieval is the use of the texture browsing descriptor to find a set of candidates and then use homogeneous texture descriptor to get a similar ranking among the candidate images [78, 79]. The *edge histogram descriptor* (EHD) is represented the spatial distribution of five type of edge, namely, four type of directional edge (*i.e.*, horizontal, 45°diagonal, and 135°diagonal) and a non-directional edge (*i.e.*, isotropic) are considered [80]. It is useful for image-to-image matching, where underlying texture is not homogeneous. More detailed description the method will be provided in the next section.

3.2.3 Texture Features Representing

Texture is an important attribute that refers to innate surface properties of an object and their relationship to the surrounding environment. If we could choose appropriate texture descriptors, the performance of the CBIR system should be improved. We use a gray level co-occurrence matrix (GLCM) and the edge histogram descriptor (EHD), which are simple and effective method for representing texture feature.

- **Gray-Level Co-occurrence Matrix**

Texture features based on the spatial co-occurrence of pixel values are probably the most widely used in the analysis of remote sensed imagery. First proposed by Haralick in 1973, they characterize texture using a variety of quantities derived from second-order image statistics. Co-occurrence texture features are extracted from an image in two steps. First, the pair-wise spatial co-occurrences of pixels separated by a particular angle and/or distance are tabulated using a gray co-occurrence matrix (GLCM). Second, the GLCM is used to compute a set of scalar quantities that characterize different aspects of the underlying texture. While these quantities can be interpreted using intuitive notions of texture, their main benefit is providing a quantitative description that can be used for image analysis, such as forming a high-dimensional feature vector to support content retrieval.

In the following, we will use $\{I(x,y), 0 \le x \le N_x - 1, 0 \le y \le N_y - 1\}$ to denote an image with G gray levels. The G x G gray level co-occurrence matrix P_d^θ for a displacement vector d = (dx, dy) and direction θ is defined as follows. The element (i, j) of P_d^θ is the number of occurrences of the pair of gray levels i and j which the distance between i and j following direction θ is d.

$$P_d^\theta(i,j) = \#\{((r,s),(t,v)): I(r,s) = i, I(t,v)| = j\}, \qquad (4.19)$$

Where $(r,s),(t,v) \in N_x \times N_y$; $(t,v) = (r + dx, s + dy)$and # denotes the number of occurrences inside the window, with m, y and j the intensity levels of the first pixel and the second pixel at positions (x_1, y_1) and (x_2, y_2), respectively.

0	1	0	2
0	2	1	1
3	1	0	0
0	0	2	3

4 x 4 image

2	1	3	0
2	1	0	0
0	1	0	1
0	1	0	0

GLCM

0.16	0.08	0.25	0
0.16	0.08	0	0
0	0.08	0	0.08
0	0.08	0	0

Nomalized GLCM

Figure 3.1: GLCM of a 4 x 4 image for distance d = 1 and direction $\theta = 0$

Figure 3.1 shows the co-occurrence matrix P_d^θ with distance d = 1 and the direction is horizontal($\theta = 0$). This relationship ($d = 1, \theta = 0$) is nearest horizontal neighbor. There will be $(N_x = 1)$ neighboring resolution cell pairs for each row and there are $N_y$rows providing $R = (N_x - 1) \times N_y$ nearest horizontal pairs. The co-occurrence matrix can be normalized by dividing each of its entries by R.

In addition, there are also co-occurrence matrix for the vertical direction $(\theta = 90)$ and both diagonal directions$(\theta = 45, 135)$. If the direction from bottom to top and from left to right is considered, there will be eight directions (0, 45, 90, 135, 180, 225, 270, 315) (Figure 3.2). In order to simplify and reduce the computational effort, we computed the GLCM according to one direction (i.e., $\theta = 0$) with a given distance d (= 1) .

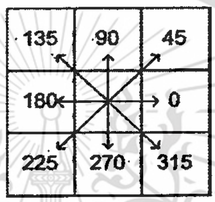

Figure 3.2: Eight directions of adjacency

While GLCMs provide a quantitative description of a spatial pattern, they are too unwieldy for practical image analysis. Haralick has thus proposed a set of scalar quantities for summarizing the information contained in a GLCM. He originally proposed a total of 14 quantities, or features; however, typically only a subset of these is used. The following five GLCM derived features are described: angular second moment (ASM), contrast, inverse different moment (IDM), entropy, and correlation.

The *angular second moment (ASM)* is known as uniformity or energy. It measures the uniformly of an image. When pixels are very similar, the ASM value will be large.

$$ASM = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P_{d,\theta}(i,j)^2 \quad , \tag{4.20}$$

It will have a larger value when the co-occurrence frequencies $I(t, v)$ are concentrated in a few places in the GLCM. This often occurs along the diagonal of a constant image or off diagonal for structured images. A near random or noisy image will have an evenly distributed GLCM with a low ASM.

The *contrast feature* is a measure of intensity or gray-level variability between the reference pixel and its neighbor. In the visual perception of the real world, contrast is determined by the difference in the color and brightness of the object and other objects within the same field of view.

$$contrast = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P_{d,\theta}(i,j) \right\} \text{, where } n = |i-j| \tag{4.21}$$

Where i and j are equal, the cell is on the diagonal and $i - j = 0$. These values represent pixel entirely similar to their neighbor, so they are given a weight of 0. If i and j differ by 1, there is a small contrast, and the weight is 1. If i and j differ by 2, the contrast is increasing and the weight is 4. The weights continue to increase exponentially as $(i - j)$ increases.

The *inverse difference moment (IDM)* is usually called homogeneity that measures the local homogeneity of an image. IDM feature obtains the measures of the closeness of the distribution of the GLCM elements to the GLCM diagonal.

$$IDM = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{1}{1+(i-j)^2} P_{d,\theta}(i,j) \quad , \tag{4.22}$$

IDM weight value is the inverse of the contrast weight, with weights decreasing exponentially away from the diagonal.

The *entropy feature* is a difficult term to define. The concept comes from thermodynamics; it refers to the quantity of energy that is permanently lost to heat every time a reaction or a physical transformation occurs. Entropy cannot be recovered to do useful work. Because of this, the term can be understood as the amount of irremediable chaos or disorder. The equation of entropy is:

$$ENT = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P_{d,\theta}(i,j) \quad \log\left(P_{d,\theta}(i,j)\right) \quad , \tag{4.23}$$

The entropy is larger for an image with an evenly distributed GLCM. Thus, a near random or noisy image will have larger entropy.

The *correlation feature* shows the linear dependency of gray level values in the co-occurrence matrix. It presents how a reference pixel is related to its neighbor, 0 is uncorrelated, 1 is perfectly correlated.

$$COR = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P_{d,\theta}(i,j) \ \frac{(i-\mu_x)(j-\mu_y)}{\sigma_x \sigma_y} \quad , \tag{4.24}$$

Where μ_x, μ_y and σ_x, σ_y are the means and standard deviations of p_x and p_y.

$$\mu_x = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} i.P_{d,\theta}(i,j) \qquad \mu_y = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} j.P_{d,\theta}(i,j) \quad , \tag{4.25}$$

$$\sigma_x = \sqrt{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i-\mu)^2 P_{d,\theta}(i,j)} \qquad \sigma_y = \sqrt{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (j-\mu)^2 P_{d,\theta}(i,j)} \quad , \tag{4.26}$$

A texture feature vector for an image can be formed from the scalar features derived GLCMs computed for different offsets.

- **Edge Histogram Descriptors**

Edge in an image is an important low-level feature, if can describe both shape and texture feature, which are also essential elements for content-based image retrieval. One way to represent such an important edge feature is the use of a histogram. The edge histogram in the image space represents the frequency and the directionality of the brightness change in the image. The edge histogram descriptor (EHD) [80] represents the distribution of the 5 type of edges in each local area called a sub-image. As shown in Figure. 3.3, the sub-image is defined by dividing the image space into 4 ×4 non-overlapping blocks. Thus, the image partition always yields 16 equal-sized sub-images regardless of the size of the original image. To characterize the sub-image, we then generate a histogram of edge distribution for each sub-image. Edges in the sub-images are categorized into 5 types: vertical, horizontal, 45-degree diagonal, 135-degree diagonal and non-directional edges (figure. 3.4). Thus, the histogram for each sub-image represents the relative frequency of occurrence of the 5 edge types in the corresponding sub-image. As the results, it is shown in figure. 3.4, each local histogram contains 5 bins. Each bin corresponds to one of 5 edge types. Since there are 16 sub-images in the image, a total of 5 ×16 = 80 histogram bins is required, (figure. 3.5). Note that each of the 80-histogram bins has its own semantics in terms of location and edge type. For example, the bin for the horizontal type edge in the sub-image located at (0, 0) in Figure 3.6 carries the information of the relative population of the horizontal edges in the top-left local region of the image.

The semantics of the 1-D histogram bins form the normative part of the MPEG-7 standard descriptor. Specifically, starting from the sub-image at (0, 0) and ending at (3, 3), 16 sub-images are visited in the raster scan order and corresponding local histogram bins are arranged accordingly. Within each sub-image, the edge types are arranged in the following order: vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional. Table 3.1 summarizes the complete semantics for the EHD with 80 histogram bins. Of course, each histogram bin value should be normalized and quantized. For normalization, the number of edge occurrences for each bin is divided by the total number of image-blocks in the sub-image.

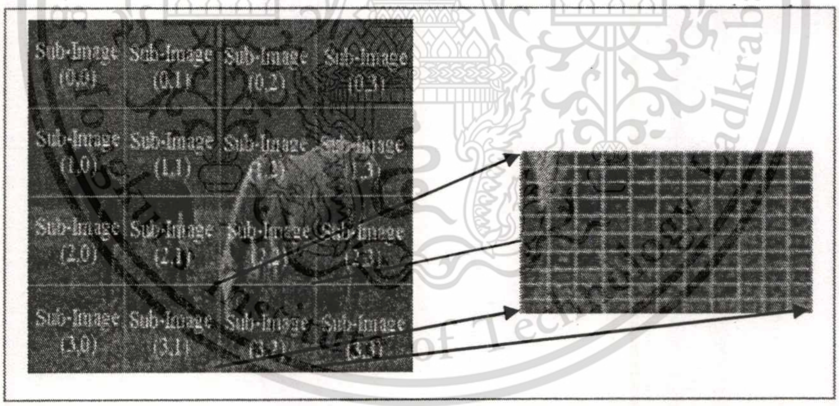

Figure 3.3: Definition of sub-image and image-block.

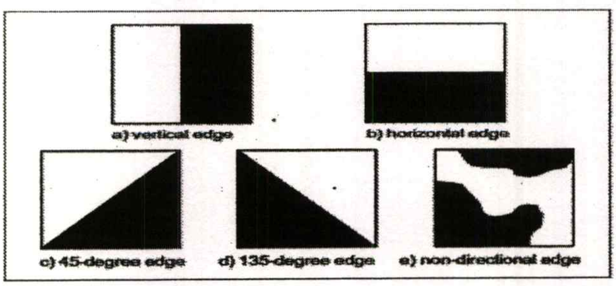

Figure 3.4: Five types of edges.

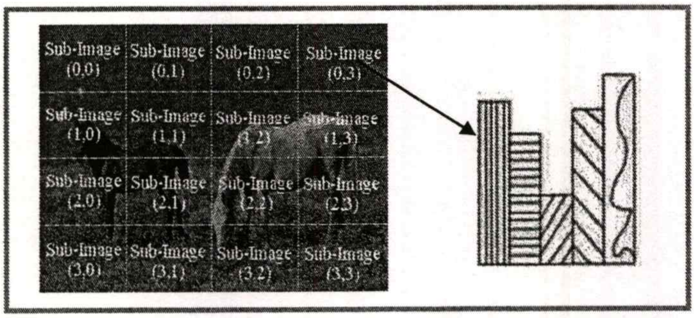

Figure 3.5: Five types of edge bins for each sub-image.

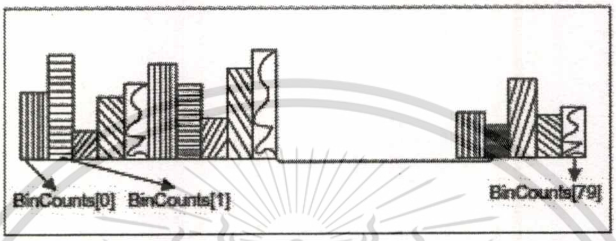

Figure 3.6: D-array of 80 bins of EHD.

The image-block is a basic unit for extracting the edge information. That is, for each image-block, we determine whether there is at least an edge and which edge is predominant. When an edge exists, the predominant edge type among the 5 edge categories is also determined. Then, the histogram value of the corresponding edge bin increases by one. Otherwise, for the monotone region in the image, the image-block contains no edge. In this case, that particular image-block does not contribute to any of the 5 edge bins. Consequently, each image-block is classified into one of the 5 types of edge blocks or a non-edge block. Although the non-edge blocks do contribute to any histogram bins, each histogram bin value is normalized by the total number of image-blocks including the non-edge blocks. This implies that the summation of all histogram bin values for each sub-image is less than or equal to 1. This, in turn, implies that the information regarding non-edge distribution in the sub-image (smoothness) is also indirectly considered in the EDH.

Table 3.1: Semantics of local edge bins.

Histogram bins	Semantics
BinCounts[0]	Vertical edge of sub-image at (0, 0)
BinCounts[1]	Horizontal edge of sub-image at (0, 0)
BinCounts[2]	45-degree edge of sub-image at (0, 0)
BinCounts[3]	135-degree edge of sub-image at (0, 0)
BinCounts[4]	Non-directional edge of sub-image at (0, 0)
BinCounts[5]	Vertical edge of sub-image at (0, 1)
:	:
BinCounts[74]	Non-directional edge of sub-image at (3, 2)
BinCounts[75]	Vertical edge of sub-image at (3, 3)
BinCounts[76]	Horizontal edge of sub-image at (3, 3)
BinCounts[77]	45-degree edge of sub-image at (3, 3)
BinCounts[78]	135-degree edge of sub-image at (3, 3)
BinCounts[79]	Non-directional edge of sub-image at (3, 3)

3.3 Shape Feature Extraction

3.3.1 Introduction to Shape

The shape of an object is a binary image representing the extent of objects. Shape representation techniques used in similarity retrieval are generally characterized as being region-based and boundary-based. The former considers the shape being composed of a set of two-dimensional regions while the latter presents the shape by its outline. Region-based feature vectors often result in shorter feature vectors and simpler matching algorithms. However, generally they fail to produce efficient similarity retrieval. On the other hand, feature vectors extracted from boundary-based representations provide a richer description of the shape. Unlike the traditional classification, the approaches of shape-based feature extraction and representation are classified according to their processing approaches. The figure 5.1 shows the hierarchy of the classification of shape feature extraction approaches.

Selecting a set of features from the shape representation to characterize an object for a certain application is not easy, since one must take into consideration the variability of the shapes and the specific characteristics of the application domain. Feature comparison can be understood as a way of quantifying the similarity/dissimilarity between corresponding objects. This is a very difficult problem since it tries to mimic the human perception of similarity between objects [81]. In addition, several shape features have been proposed in the literature for shape characterization [82, 83]. Many of these techniques, however, cannot be used for content-based image retrieval due to their complexity or because they lack a counterpart in the human visual system. Therefore, techniques based on simple and visually meaningful shape features have been used in several content-based retrieval systems [81].

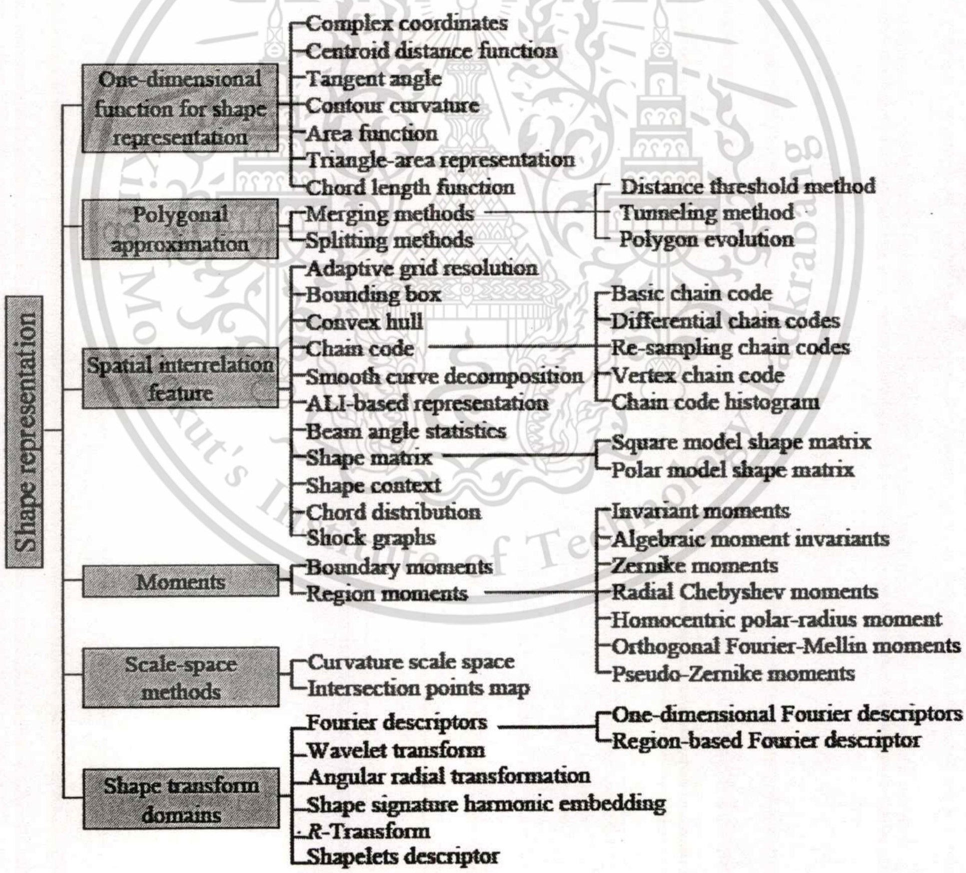

Figure 3.7: An overview of shape description techniques

3.3.2 Shape Attributes

Shape of the object is the characteristic surface configuration as represented by the outline or contour. Shape recognition is one of the modes through which human perception of the environment is executed. It is important in CBIR because it corresponds to region of interests in images. Shape feature representations are categorized according to the techniques used. There are boundary based and region based [84]. In region based techniques, all the pixels within a shape are taken into account to obtain the shape representation. Common region based methods use moment descriptors to describe shapes. Region moment representation interprets a normalized gray level image function as a probability density of a 2-D random variable. The first seven invariant moment is derived from the second and third order normalized central moments, because moments combine information across an entire object rather than providing information just at a single boundary point, they capture some of the global properties missing from many pure contour based representations. The contour based is very popular using for shape representation. Contour based shape representation only exploits shape boundary information. Simple contour based shape descriptors include area, perimeter, compactness, eccentricity, elongation, and orientation. Complex boundary based descriptors include Fourier descriptors, and chain codes [85]. In our proposed system, we have considered a contour based shape representation that three shape-invariant is obtained the features by Fourier descriptors derived from the spectral transform on different shape signatures.

3.3.3 Shape Feature Representing

In general, shape descriptor is some set of numbers that are produced to describe a given shape feature. A descriptor attempts to quantify shape in ways that agree with human intuition (or task-specific requirements). Good retrieval accuracy requires a shape descriptor to be able to effectively find perceptually similar shapes from a database. Usually, the descriptors are in the form of vectors. Shape description should meet the following requirements: 1) the descriptor should be as complete as possible to represent the content of the information items. 2) The descriptor should be represented and stored compactly. The size of descriptor vector must not be too large. 3) The computation of the distance between descriptors should be simple; otherwise the execution time would be too long.

- **Shape Signatures**

The shape signature is any one-dimensional function representing two-dimensional areas. Three shape signatures are considered in our case, these are centroid distance, complex coordinated, and curvature signature which is derived from shape boundary coordinates. In the following, we assume the shape boundary coordinates $(x(t), y(t)), t = 0, 1, \ldots, L-1$, have been extracted in the preprocessing stage [86, 87].

A *complex coordinate function* is simply the complex number generated from the boundary coordinates:

$$z(t) = x(t) + iy(t) \tag{5.61}$$

In order to eliminate the effect of bias, we use the shifted coordinate function:

$$z(t) = [x(t) - x_c] + i[y(t) - y_c] \tag{5.62}$$

Where (x_c, y_c) is the centroid of the shape, which is the average of the boundary coordinates

$$x_c = \frac{1}{L}\sum_{t=0}^{L-1} x(t), \quad y_c = \frac{1}{L}\sum_{t=0}^{L-1} y(t) \ , \tag{5.63}$$

This shift makes the shape representation invariant to translation.

The *centroid distance function* is expressed by the distance of the boundary point from the centroid of the shape

$$r(t) = \sqrt{(x(t) - x_c)^2 + (y(t) - y_c)^2} \ , \tag{5.64}$$

Due to the translation of centroid, which represents the position of the shape, from boundary coordinates, the centroid distance representation is also invariant to translation.

A *curvature function* is a function of angular changes of a boundary tangent. The curvature at a boundary can be defined as the differentiation of successive boundary angles calculated in window *w*:

$$K(t) = \theta(t) - \theta(t-1) \ , \tag{5.65}$$

Where

$$\theta(t) = \tan^{-1}\frac{y(t) - y(t-w)}{x(t) - x(t-w)}\,, \tag{5.66}$$

- **Fourier Descriptor-Based Feature**

Fourier transformation on shape signatures is widely used for shape analysis. There are also some recent attempts to exploit it for shape retrieval [88, 89]. The Fourier transformed coefficients from the Fourier descriptors of the shape. These descriptions represent the shape of the object in the frequency domain. The lower frequency descriptors contain information about the general features of the shape, and the higher frequency descriptors contain information about the finer details of the shape. Although the number of coefficients generates from the transform is usually large, a subset of coefficients is enough to capture the overall features of the shape. The very high frequency information describes the small details of the shape. It is not so helpful in shape discrimination. Therefore, they can be ignored. As the result, the dimensions of the Fourier descriptors used for indexing shapes are significantly reduced. So, the shape signatures are described by the discrete Fourier transform (DFT), let's $s(t), t = 0, 1, \ldots, L,$ is a sampling stage. The discrete Fourier transform of $s(t)$is given by

$$u_n = \frac{1}{N}\sum_{t=0}^{N-1} s(t)exp\left(\frac{-j2\pi nt}{N}\right), \quad n = 0, 1, \ldots, N-1\,, \tag{5.67}$$

The coefficients u_n, $n = 0, 1, \ldots, N-1$, are usually called Fourier descriptors (FD) of the shape, denote as FD_n, $n = 0, 1, \ldots, N-1$.

Before applying Fourier transform on the shape signature, shape is first sampled by equal arc length sampling to fix the number of points [90]. The equal arc length sampling method selects candidate point spaced at equal are length along the shape boundary. The space between two consecutive candidate points is given by P/K, where P is the perimeter of the shape boundary and K is the total number of candidate points. For each shape, we selected $N = 2^8$candidate points with equal arc length space between them.

In shape retrieval, user is only interested in the outline features of similar shapes, the position, size and rotation of the shapes is not important. In order to make the model shape and data shapes comparable, the shape representations must be invariant to translation, rotation and scaling. Shape invariant is difficult to achieve under spatial domain, most invariance techniques in spatial domain, especially rotation invariance techniques, involve a large amount of computation. However, shape invariance is easy to achieve for the FDs. All the three shape signatures described in above mentioned are invariant under translation, therefore, the corresponding FDs are also translation invariant. Rotation invariant of the FDs is also achieved by ignoring the phase information and by taking only the magnitude values of the FDs.

For complex coordinated signature, all the N descriptors except the first one (DC component) are needed to index the shape. The DC component depended only on the position of the shape, it is not useful in describing shape this is discarded. Scale normalization is achieved by dividing the magnitude value of all the other descriptors by the magnitude value of the second descriptors. The invariant feature vector used to index the shape is given by

$$fd = \left[\frac{|FD_2|}{|FD_1|}, \frac{|FD_3|}{|FD_1|}, \ldots, \frac{|FD_{N-1}|}{|FD_1|}\right], \tag{5.68}$$

For centroid distance signature and curvature signature, the function are real valued, there are only $N/2$ different frequencies in the Fourier transform, therefore, only half of the FDs is needed to index the shape. Scale invariance is obtained by dividing the magnitude values of the first half of FDs by the DC component

$$fd = \left[\frac{|FD_1|}{|FD_0|}, \frac{|FD_2|}{|FD_0|}, \ldots, \frac{|FD_{N/2}|}{|FD_0|}\right], \tag{5.69}$$

3.4 Chapter Summary

In this chapter we presented the feature extraction of image to represent an image feature. There are involved three different techniques to extract the features including color, texture, and shape features. The color feature extraction, we used the global and local features of the color image to obtain the suitable color space with retrieving images from the database. The mean value and the standard deviation have been used to extract the global feature from an image and the image bitmap as the local to describe the local characteristics of the image. The texture features are extracted by the gray-level co-occurrence matrix and edge histogram descriptors. The gray level co-occurrence matrix (GLCM) have been extracted the texture from the image and obtained the texture features by statistical features. The edge histogram descriptors (EHD) are represented the local edge distribution in the image following the MPEG-7 standard. For shape feature extraction, a shape feature in accordance with human perception is not an easy task. Due to the fact that human vision and perception are an extraordinary complicated system, it is to hope that machine vision has a super excellent performance with small complexity. According to the shape requirements for shape retrieval, the Fourier descriptors have been proposed to represent the shape feature.

Chapter 4

Feature Similarity and Indexing

4.1 Introduction

The similarity between two images (represented by their feature values) is defined by a similarity measure. Selection of similarity metrics has a direct impact on the performance of content-based image retrieval. The kind of the feature vectors selected determines the kind of measurement that will be used to compare their similarity [89]. If the features extracted from the images are presented as multi-dimensional points, the distance between corresponding multi-dimensional points can be calculated. Euclidean distance is the most common metric used to measure the distance between two points in multi-dimensional space [90].

For other kinds of features such as color histogram, Euclidean distance may not be an ideal similarity metric or may not be compatible with the human perceived similarity. Histogram intersection was proposed by Swain and Bllard [91] to find known objects within images using color histograms. A number of other metric, such as Mahalanobis distance, Minkowski distance, Earth Mover's distance, and Proportional Transportation distance, have been proposed for specific purposes.

In this thesis, we organized four experiments of the similarity indexing techniques: The first of these is the color similarity indexing, which were used to obtain the best retrieval color model. The second of matching technique is a texture similarity; this technique is used to find the best retrieving images from database. The third of the technique, shape similarity between images were used to retrieve relevance shape images, which according to the suitable shape requirements, it proposed in the chapter 5. The fourth, the combination all of three feature similarity and weighting to achieve the best retrieving scheme.

4.2 Color

The color similarity, we used two different measurements to evaluate the similarity, which are the global and local features. For the global features S_{c2}, μ and σ, we used the Euclidean distance to calculate the similarity. On the other hand, the local feature S_{c1} is used the Hamming distance to evaluate the distance between the two bitmaps. After that, before the linear combining of these two similarity values, the normalized information distance (NID) [92] are applied tonormalize the value of each distance adjusted to the same level: the distance has a value between 0 and 1.

For two images, the Hamming distance H used to evaluate the image bitmap similarity is defined by

$$H(BM^d, BM^q) = \frac{1}{3\times m}\left\{\sum_{j=1}^{m}(IX_j^d - IX_j^d) + \sum_{j=1}^{m}(IY_j^d - IY_j^d) + \sum_{j=1}^{m}(IZ_j^d - IZ_j^d)\right\}, \quad (6.1)$$

$$S_{c1}(d,q) = \frac{H(BM^d, BM^q) - min\{H(BM^d, BM^q)\}}{max\{H(BM^d, BM^q)\}}, \quad (6.2)$$

Where IX_j^d is the jth component of the bitmap IX of the image d from the database and the IX_j^q is the jth component of the bitmap IX of the query image q . BM means the image bitmap feature of both images. So, the distance similarity measurement of global color featureis formed as follows:

$$E(d,q) = \sqrt{\sum_{t\in(X,Y,Z)}(\mu_t^d - \mu_t^q)^2 + \sum_{t\in(X,Y,Z)}(\sigma_t^d - \sigma_t^q)^2}, \quad (6.3)$$

$$S_{c1}(d,q) = \frac{E(d,q) - min\{E(d,q)\}}{max\{E(d,q)\}}, \quad (6.4)$$

Where μ_t and σ_t represent the mean value and standard deviation of the image from the database d and the query image q in color space respectively. After that, we used the linear combination between the local and global features to represent the distance of the color features(d_{color}).

4.3 Texture

The texture similarity, we used a linear combination two simple and effective features, there is the entropy derived from the gray co-occurrence matrix and the edge histogram descriptor. For similarity measure between the query image and the image in the database of the entropy E_t, we used the Euclidean distance and the Hamming distance H_t were used to evaluate the distance between two edge histogram descriptors. Before the combination, there are always normalized by the normalized information distance to adjust the value distance to the same level. The Euclidean distance between two images with the entropy similarity can be defined by follow:

$$E_t(d,q) = \sqrt{(ENT^d - ENT^q)^2} \ , \tag{6.5}$$

For the normalized distance of the entropy is defined by

$$S_{t1} = \frac{E_t(d,q) - min\{E_t(d,q)\}}{max\{E_t(d,q)\}} \ , \tag{6.6}$$

Where ENT^d and ENT^q represents the entropy of the image from the database d and the query image q respectively.

The Hamming distance between two images with the edge histogram descriptors can be defined by follow:

$$H_t(d,q) = \sum_{j=0}^{79} \left| EHD_j^d - EHD_j^q \right| \ , \tag{6.7}$$

For the normalized distance of the edge histogram descriptor can be defined as follows:

$$S_{t2} = \frac{H_t(d,q) - min\{H_t(d,q)\}}{max\{H_t(d,q)\}} \ , \tag{6.8}$$

Where EHD^d and EHD^q represents the edge histogram description of the image from the database d and the query image q respectively. In the end, the linear combination was used to combine all of two features to represent similarity of the texture feature($d_{texture}$).

4.4 Shape

The shape similarity, we described all of three shape invariant including the translation, rotation, and scaling invariants: the location, rotation and scaling changing of the shape must not affect the extracted features. The shape representation was used the Fourier descriptors derived from the spectral transform on the shape signature. Other advantages of the FD method include easy normalization and information preserving. So, the DC component was used to normalize the Fourier descriptors to obtain the shape invariant likely the equation 5.68 and 5.69. After the normalization, the shape indexing included by FD feature $fd = [FD_1, FD_2, \dots, FD_N]$. The similarity between a query shape q and a database shape d is determined by the Euclidean distance $fd(d, q)$ between their FDs:

$$fd(d,q) = \sqrt{\sum_{i=1}^{N} \left(FD_i^q - FD_i^d\right)^2} , \tag{6.7}$$

In addition, to adjust the distance the component of shape descriptors, the normalized information distance was used after the linear the combination of them. That can be defined as follows:

$$d_{shape} = \frac{fd(d,q) - min\{fd(d,q)\}}{max\{fd(d,q)\}} , \tag{6.8}$$

4.5 Indexing Representing using the Combination Method

Indexing a database consists of building and utilizing secondary data structures relating to the contents of the actual records in the database. These indices facilitate efficient retrieval of the database records according to given queries. An index can be seen as an ordering of the records in the database according to a group of attributes. In a database of people a simple example of an index would be a list where the people are ordered according to their height. Another index could order the people according to their weight. In practical indices the different attributes of records are usually combined. Some of the indexing methods partition the space to which the database records belong while other methods operate directly on the records themselves.

To achieve the efficiency of the retrieval system, we used a combination and weight variation all of three feature including the color, texture and shape features to represent similarity indexing, there are can be by follows:

$$SIM(d,q) = w_c \times d_{color} + w_t \times d_{texture} + w_s \times d_{shape}, \tag{6.9}$$

Where w_c , w_t and $w_s$represents the weight of the color, texture, and shape respectively.

4.6 Proposed CBIR Methodology

The content-based image retrieval methodology in this system goes throughout the steps shown in the block diagram Figure 6.1. The user enters a query image for which the system extracts three features including color, texture, and shape features as explained in the previous sections, the feature vectors of database images are previously extracted and stored. Using the similarity metrics defined for color, texture, and shape, the similarity distances between the query image and every image in the database are calculated according to Equation 6.9 and then are sorted in ascending order. The first N similar target images (with a smallest distance value of the query) are retrieved and shown to the user, where N is the number of the retrieved images required by the user.

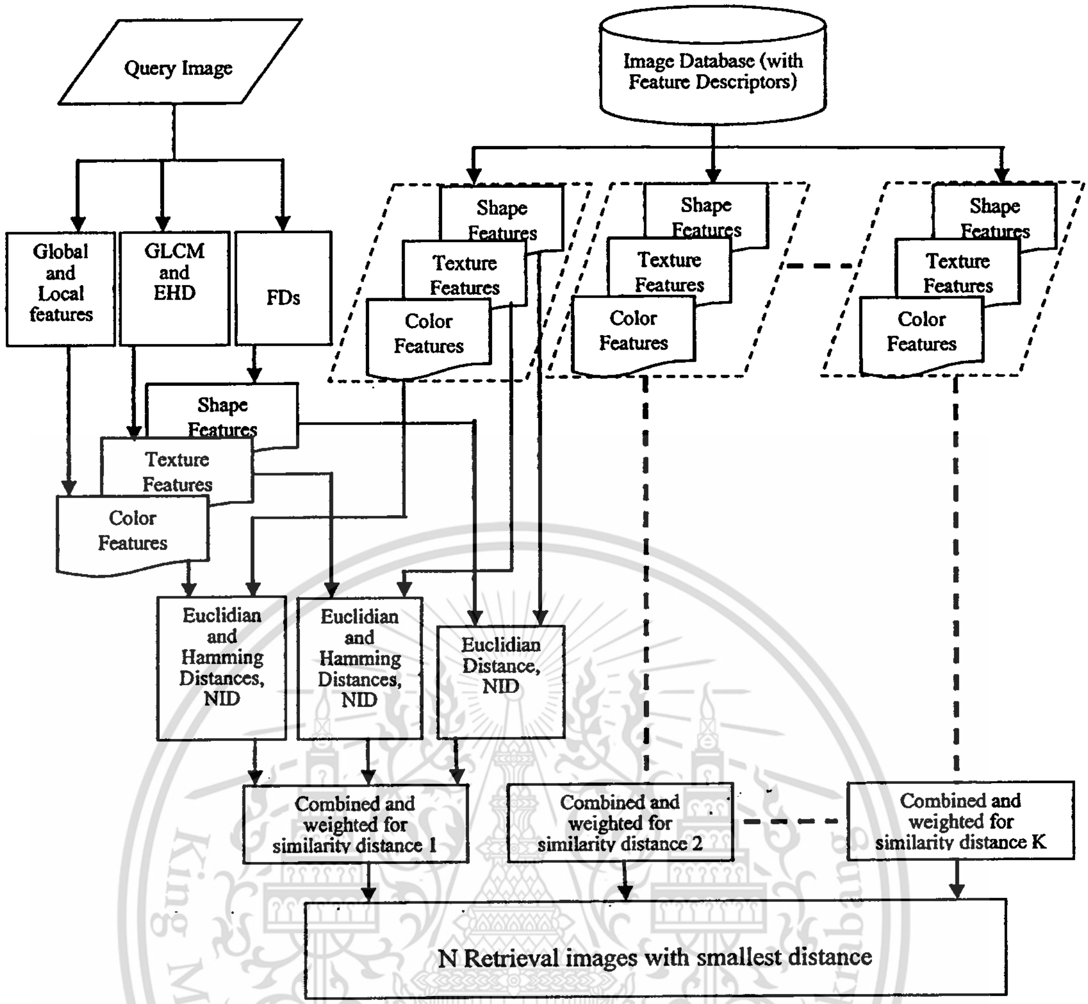

Figure 4.1: Block diagram of proposed CBIR system

4.7 Chapter Summary

Descriptive view is obtained the feature similarity and indexing by a comparison a suitable that those are proposed in the chapter 5. This Chapter, we organized three similarities of the feature including color, texture, and shape; the color derived from a linear combination of the global and local feature, the texture also derived from the linear combination of the entropy measure randomness of the GLCM and the edge histogram descriptors, and the shape derived from the normalized information distance of the Fourier descriptors. In addition, the goal of this chapter is obtaining the indexing representation, we used the linear combination all of them similarity and given weight variation to achieve higher retrieval experimental results in the next chapter.

Chapter 5

Result and System Evaluation

In this chapter, we will present the evaluation of our proposed system that was introduced in the previous chapter. We introduce the database we select to test our system, and we have compared our system result with other already existing CBIR systems that most of them use the same image database.

5.1 Image databases

The database we used in our evaluation is a SIMPLIcity project or WANG database [55]. The WANG database is an image database that the images are manually selected from the Corel database. In WANG database, the images are divided into 10 classes, each class contains 100 images. It is widely used for testing CBIR systems. Classification of the image in the database into 10 classes makes the evaluation of the system easy. Figure 5.1 shows one example image for each class.

Class 1: Africans people

Class 2: Beaches

Class 3: Buildings

Class 4: Buses

Class 5: Dinosaurs

Class 6: Elephants

Class 7: Flowers

Class 8: Horses

Class 9: Mountains

Class 10: Foods

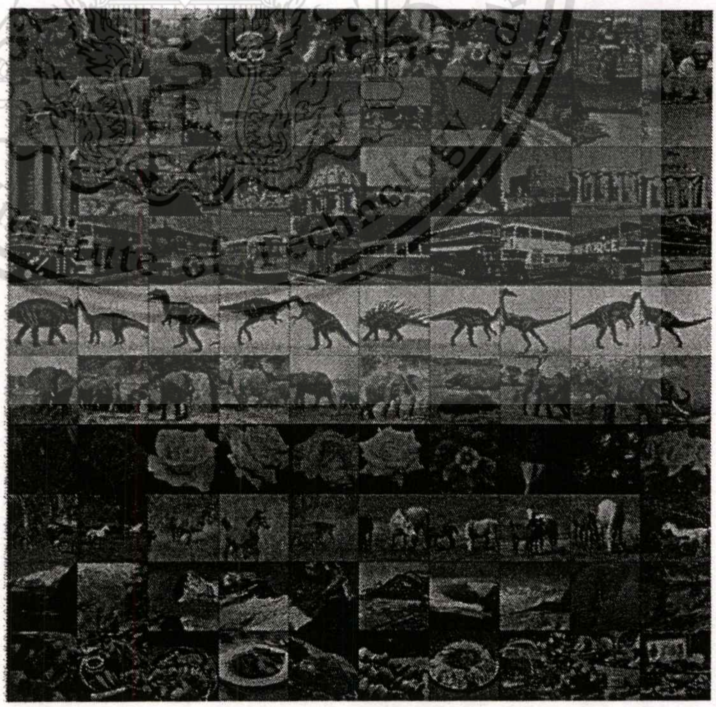

Figure 5.1: One example images from each of the 10 classes of WANG database.

This database was created by the group of Professor James Wang from the Pennsylvania State University. This database is free and available to download. It was used extensively to test many CBIR systems. This popularity is because the size of the database and availability of class information allows for performance evaluation. Sine this database is a subset of the Corel database, the images are of size 384×256 or 256×384 pixels as well.

As we said before, we choose this database for many reasons. Many CBIR systems use this database, so we can easily compare these results with our systems. This database contains large images in different classes (100 images per class). Diversity in the images helps us to get more results and enhance our system. Also, we can explain the results we get according to the database that the number of images we get from testing and compare them with the images in the database. The most interesting point is that the images are familiar to human, and they are very friendly to us.

5.2 Implementation Environment

Our proposed system is implemented using a Matlab image processing program of version R2010a. We use a platform of Intel Core i7 Processing power of 2.10 GHz CPU with 8 GB RAM during the implementation. 1000 images went through the implemented system to extract the features and stored them. The extracted features are used to find the similarity indexing which stored in the image database with the feature descriptors in the offline zone that the database now is already for testing and evaluating our proposed system.

5.3 Performance Evaluation of CBIR

When we want to evaluate a CBIR system, we may face some problems. One major problem we face for CBIR system performance evaluation is that neither a standard database nor a unique performance measure is available. There are many image databases that are used to represent results for CBIR system. So, no standard image databases are available for CBIR system. Therefore, it is impossible to compare the performance of different systems using different image databases. Furthermore, the early CBIR systems were restricted their results.

They presented one or more example queries, which are used to give a positive impression about the efficiency of the system. This is neither a quantitative nor an objective measure that can be used for system performance evaluation.

In CBIR, the most commonly used performance measures are Precision and Recall. **Precision** gives information about the relative effectiveness of the retrieval scheme, which is defined as the ratio of the number of retrieved relevant images to the total number of retrieved images. **Recall** giving information about the absolute accuracy of the retrieval scheme, it defined as the ratio of the number of retrieved relevant images to the total number of relevant images in the database [93]. We denote to the precision by p and the recall by r , there are can be defined as follows:

$$p_m = \frac{n_r}{M} \text{ and} \tag{7.1}$$

$$r_m = \frac{n_r}{N_r}, \tag{7.2}$$

Where M is the total number of retrieved images, n_r is the number of relevant correct images retrieved, N_r represents the total number of relevant images available in the database. It is assumed that the user inspects the m first images of the ranked list. So, for every $m = 1, 2, \ldots, M$, precision p_m and the recall r_m are calculated.

In CBIR, if the precision score is 1.0, this means that every image retrieved by a search is relevant, but we do not know if the search retrieves all images relevant to the query. If the recall score is 1.0, this means that all relevant images are retrieved by the search, be we do not know the number of irrelevant images were also retrieved.

In classification task, a precision of score 1.0 for a class X means that every object labeled as belonging to class X does indeed belong to class X, but we do not know the number of objects from class X that were not labeled correctly. Also, a recall of score 1.0 for a class X means that every object from class X was labeled as belonging to X, but we do not know the number of objects were also labeled as belonging to class X incorrectly. There is a trade-off between precision and recall. In search engine, for example, the user prefers to increase the number of results relevant to the query by retrieving more documents. This will increase the recall. On the other hand, the results will have a quite large number of irrelevant documents retrieved. This will decrease the precision [95].

Usually, precision and recall scores are not discussed in isolation. Instead, either values for one measure are compared to a fixed level at the other measure (e.g. Precision at a recall level of 0.75) or both are combined into a single measure, such as a precision / recall graph. Precision-recall pair is a good standard of performance evaluation. It provides meaningful result when the database type is known and has been effectively used in some earlier research. For other data sets, especially those that have been created by collecting user generated images, the result may vary due to different human concepts of image classification. Because the ground truth is known for the whole database, every image in the database can be used as the query. For each query, the precision of the retrieval at each level of the recall is obtained.

5.4 Image Retrieval Experiments

This section is the most important section of the thesis. Here, we test our proposed system and show the results. We explain the results, comment on them, and compare our system with other existing systems.

In this experiment, we organized as four experiments. The first one is the color feature experimental, it is to compare efficient performance of the color space such as RGB, HSV, YUV, YIQ, YCbCr, and L*a*b*, for retrieving images from the large database. The second is the texture features, which is showing the effect of the standard statistical of texture feature with retrieving images. The third is the shape feature that the shape descriptors derived from the Fourier descriptors, there are also to obtain knowledge the effect of the shape feature on the retrieving images from the general image database. The last one, it is a combination all of three features and given weight variation to achieve a higher retrieval performance based on our proposed system.In addition, to check retrieval effectiveness of the proposed system, we have to test it by selecting some images randomly and retrieve some images. Also, we have to calculate the two performance evaluation metrics that are Precision and Recall. Finally, we compare our proposed system with other existing systems and show the efficiency of the proposed system.

In each experimental of our proposed system, the evaluation of the retrieval precision is performed ten category images were randomly selected from each specific category in the database as query images and count the precision and recall separately after getting the return result.

5.4.1 Experiment I

This section assesses the performance of content-based image retrieval in the six color spaces (RGB, HSV, YUV, YIQ, YCbCr and L*a*b*) which is based on the global and local features to measure their performance. In each color space experimental, the evaluation of the retrieval precision is performed ten images were randomly selected from each specific category in the database as query images. It can be also shown that it has the same good retrieval result in most of the other images in the database if they are chosen as queries. The precision values of the retrieval results for top 20retrieved corresponding images are given in Table 5.1.

Table 5.1:The precision values of the top 20 retrieved images different query images.

Category	RGB	HSV	YUV	YIQ	YCrCb	L*a*b*
Africa people and village	0.63	0.58	0.73	0.35	0.68	0.65
Beaches	0.73	0.45	0.53	0.33	0.55	0.61
Buildings	0.45	0.30	0.46	0.36	0.45	0.48
Buses	0.73	0.83	0.98	0.46	0.66	0.95
Dinosaurs	1.00	1.00	1.00	1.00	1.00	1.00
Elephants	0.75	0.60	0.75	0.38	0.67	0.70
Flowers	0.90	0.35	0.61	0.20	0.65	0.58
Horses	0.78	0.41	0.81	0.51	0.78	0.76
Mountain and glaciers	0.50	0.43	0.55	0.33	0.53	0.52
Foods	0.63	0.58	0.73	0.21	0.85	0.78
Average	0.71	0.55	0.72	0.41	0.68	0.70

As can be seen in Table 5.1, each color space can be affected for retrieving images from database. It can be also shown that it has some good retrieval results for each image category with any color space. But it cleared to know that the average precision of YUV color spacehas the performance higher more than other color spaces and the YIQ losses than other color spaces to retrieve the same image categories.

To further evaluate a comparison the color spaces with retrieving images, 20 images are randomly selected as queries from the 10 semantic classes in the database for top 5, 10, 20,and 50 retrieved images from the database, for the precision of retrieval at each level of the recall is obtained by gradually increasing the number of retrieved images. The performance comparison of the color space retrieval results are averaged to give the final precision/recall chart of Figure 5.2.

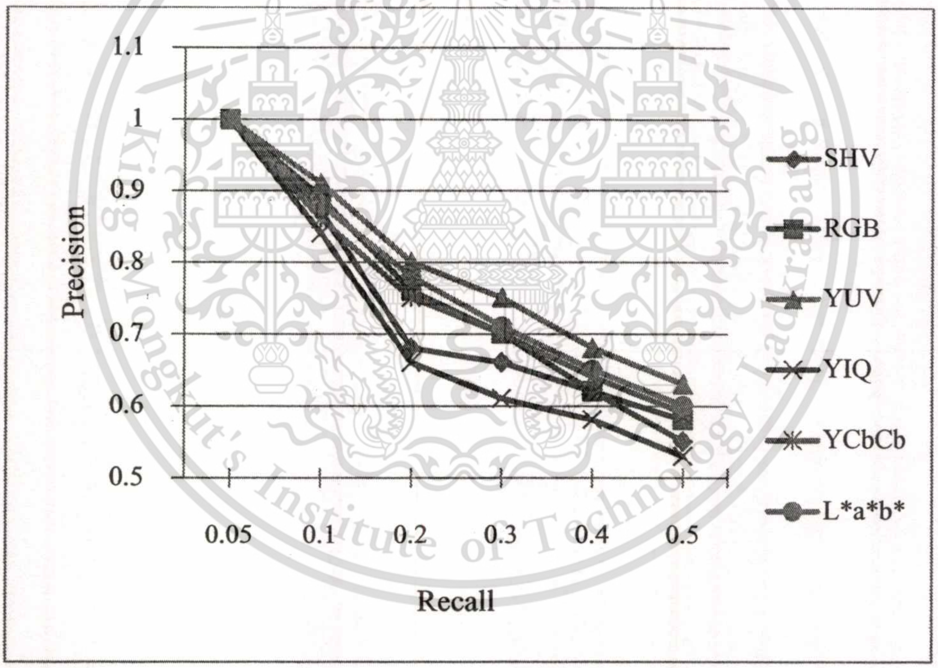

Figure 5.2: The average precision/recall chart of six color space system.

From figure 7.2, we notice that the YUV color space system has good precision result then another color space over the different values of recall. The maximum average precision of 1 at recall value of 0.05, and the value decreases to 0.63 at 0.5 of recall. For the losses color space is the YIQ, the maximum average precision of 1 at recall value of 0.05 and the value decreases to 0.53 at 0.5 of recall.

5.4.2 Experiment II

This section presents an application of the gray level co-occurrence matrix (GLCM) and edge histogram descriptors to texture-based similarity evaluation. Retrieval results of the texture feature are also used with the same database that to evaluate the effect of texture feature for retrieving images from general image database. For the texture feature representing, we used the combination of entropy derived from the GLCM and edge histogram descriptors. In order to calculate the performance, we used the approach as that of the experiment I. For each category in the 1000 database images, we randomly selected 20 images as queries. Since we have 10 categories in the database, we have 200 query images. For each query, we examined the precision of the retrieval based on the relevance of the semantic meaning between the query and the retrieved images. The 200 retrieval results are averaged to give the final precision/recall chart of figure 5.3

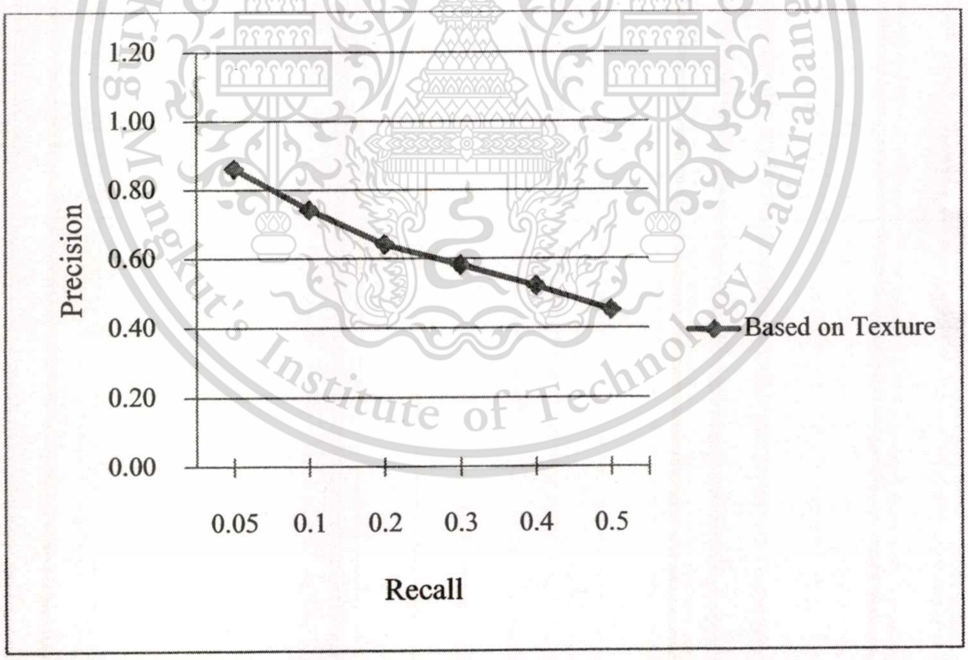

Figure 5.3: The average precision/recall chart of a CBIR system based on texture feature.

From figure 5.3 it can be noted that the CBIR system based on texture feature is can effected for retrieving images from general database, it also more effect over different recall levels, but there is losses performance than the color feature. It has a maximum average precision of 0.86 at recall level of 0.05, this value decreases to 0.44 precision value at 0.5 of recall levels.

5.4.3 Experiment III

This section provides an important application for the proposed technique on content-based image retrieval based on shape features, and assuming the effect of shape characteristics retrieving with the general image database. The shape representation was used the Fourier descriptors derived from the spectral transform on the shape signature. The experiments for shape similarity evaluation were performed on the same database image with the randomly selected 20 images as queries from 10 image categories in the database. We also use the precision/recall curve to evaluate the performance of CBIR based on shape feature using the same steps we used in testing in the experiment I and experiment II. The average precision/recall curve for the CBIR based on shape feature is shown in Figure 5.4.

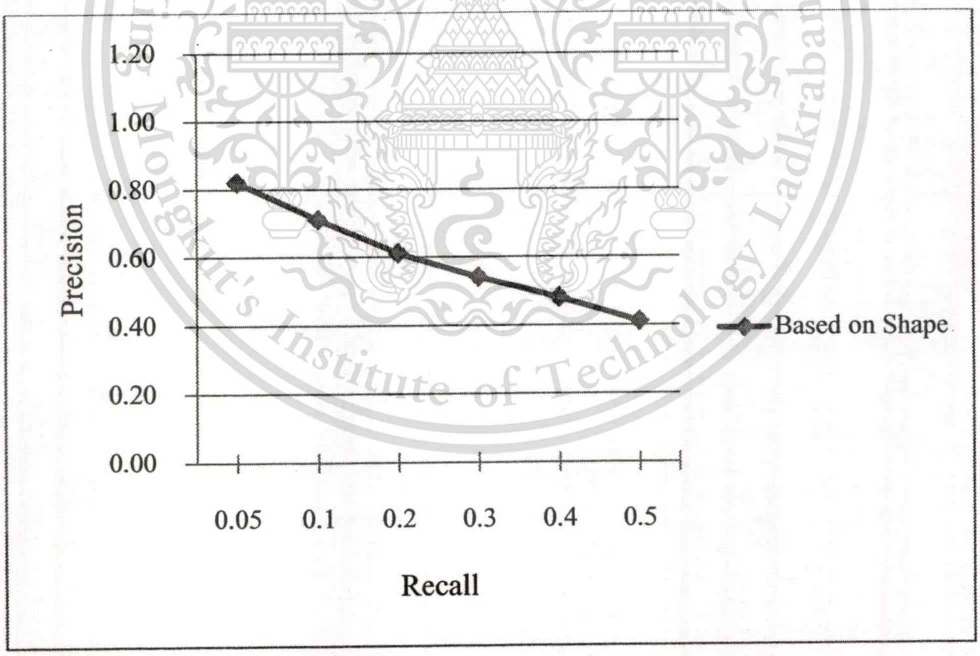

Figure 5.4: The average precision/recall chart of a CBIR system based on shape feature.

From figure 5.4, we notice that the CBIR system based on shape feature is also affected for retrieving images from general database, but the performance is losses than the color and texture features based on our proposed approach. It has a maximum average precision of 0.81 at recall level of 0.05, this value decreases to 0.41 precision value at 0.5 of recall level.

5.4.4 Experiment IV

Based on three experiments was mentioned above, some retrieval results of the image category have losses performance for retrieving images from general image database. In this section is very important of our approach because we have improved the efficiency of our proposed CBIR using a combination all of three features and achieve a higher performance by using weight variation. To show that based on our approach can be increase the efficiency and accuracy of the retrieval results, we randomly selected images from 10 different semantic classes in the database as queries. The responding of image, we used the top 20 retrieved images to compute the precision and recall. For the efficiency and accuracy of our retrieval system is shown in the Figure 5.5, which is weight variation depends on the best weight of the color feature. In this case, we try given weight for experimentation to get the best weight. It is clear that the efficiency and accuracy of retrieval scheme depended on the weight of color feature; the best weight of the color feature is 70%. After that, the weight variations of texture and shape features were varied seven times to achieve higher performance of the CBIR system.

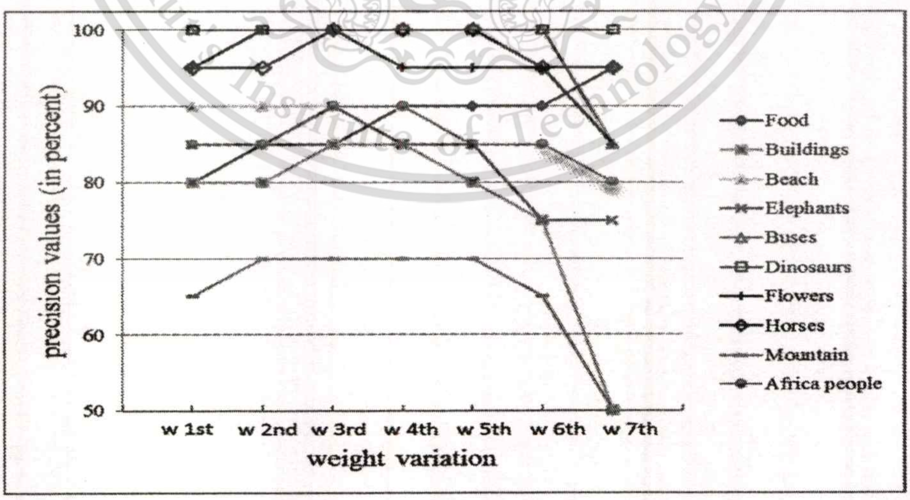

Figure 5.5: Precision values of retrieving images in each image category

Figure 5.5 gives a crossover point of precision value of CBIR based on our method; it depends on the weight variation. The weight of color feature distance was constant at 70% and varied the weight of texture and shape feature distances from 30% to 0% and 0% to 30% respectively. Based on that way, the higher performance was achieved when we set w_c= 70%, w_t= 20%, w_s = 10%, there are the weight of the color, texture and shape feature distances. The weight factor of color feature distance is higher than the weight factor of texture feature and shape feature distance because the database was used to experiment with our system consists of most natural images.

One traditional graph that describes the performance of the system is the Precision/Recall graph. It provides a meaningful result when the database is known and has been used be some earlier systems. We used the Precision/Recall graph to show the performance of our proposed system. It also randomly selected 20 images from 10 different semantic classes in the database as queries to calculate the precision and recall. For each image, the precision of retrieval result is obtained increasing the number of retrieved images. Figure 5.6 shows the Precision/Recall graph.

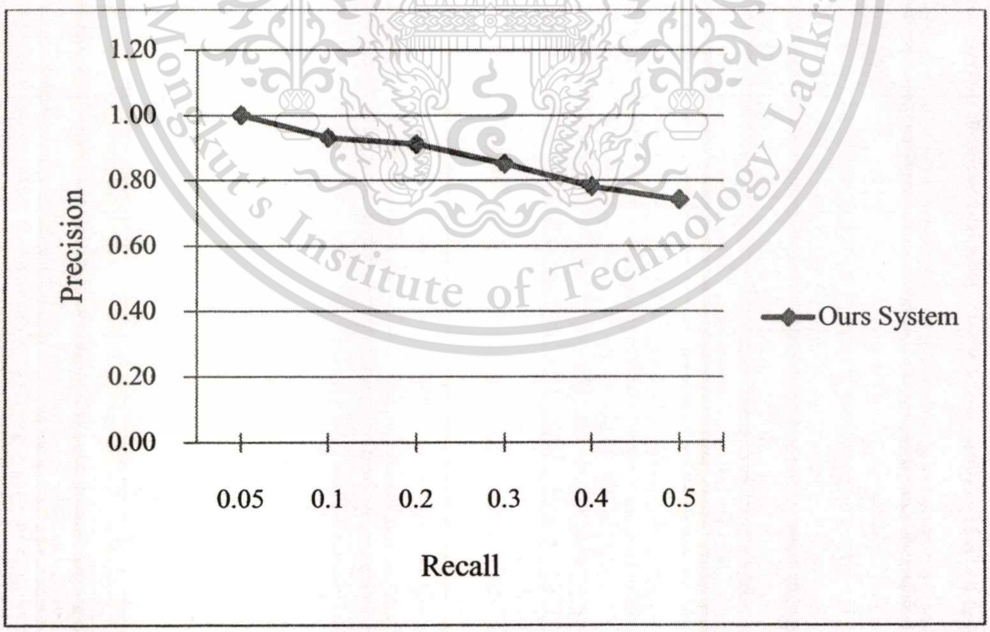

Figure 5.6: The average precision/recall chart of our proposed system

From figure 5.6, we notice that the system has good precision results over the different values of recall. The maximum average precision of 1 at recall value of 0.05, and the precision value decreases a little to 0.74 at 0.5 of recall. For example, if a user submits a query image and he want just 20 relevant images from 100 images retrieved by the system, the graph shows us that the user will get 91 relevant images to the query image instead of 20 images. In other words, for an average recall value of 20%, we have an average precision 91%. This means that we intend to get 20% of the relevant images in the database and we will get 91% of the retrieved images that are relevant to the query image.

To further evaluate our proposed CBIR system, some images are randomly selected as queries from different classes, namely Flowers, Dinosaurs, Buses, Horses, Elephants, and Foods. Each query returns the top 50 images from database. The first test is selecting a random image from the Buses class. We submit the image to the system and retrieve the top 50 images that are similar to the query image; it can be seen from Figure 5.7 that the most of retrieved images belonged to the same class at top 5 to 40. When the corresponding images increase to 50, the retrieved images are a bit incorrect.

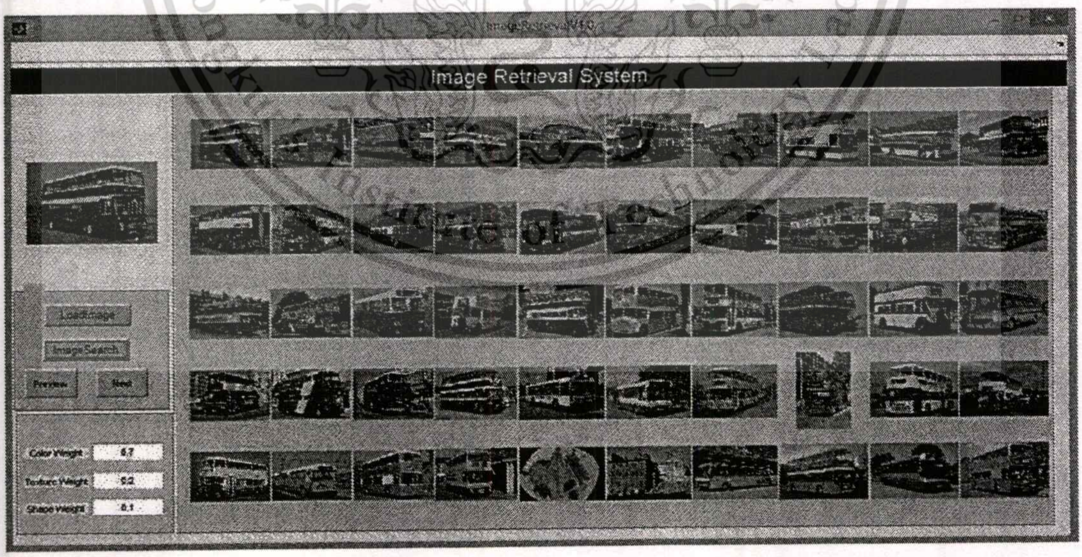

Figure 5.7: Buses Query, the top 50 retrieved images

Figure 5.8 shows another query image and its result. This is a second test for the system. We select an image randomly from the Dinosaurs class and all images retrieved by the system are relevant to the query image. Notice that the retrieved images are belonging to the same class of the query image.

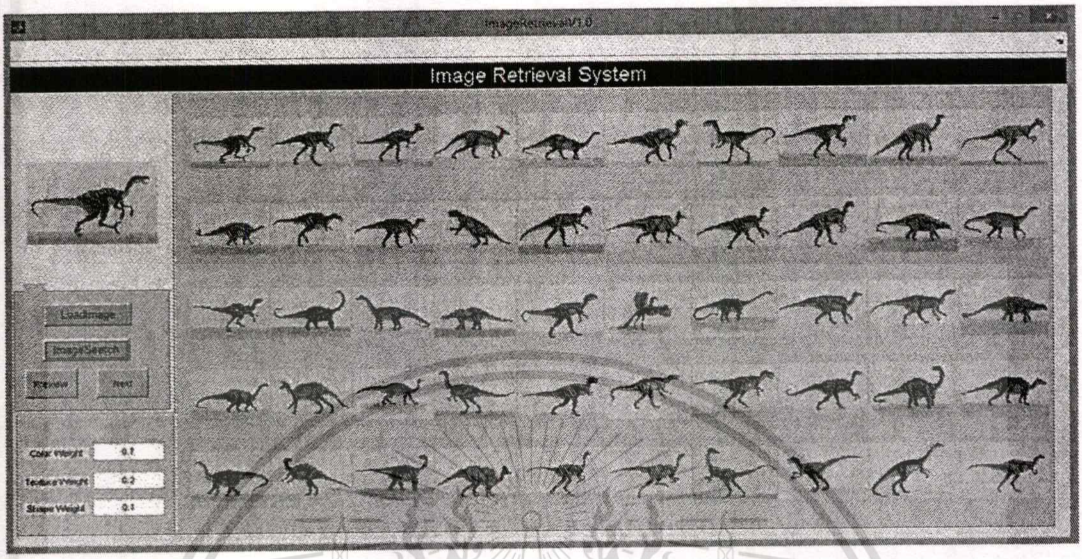

Figure 5.8: Dinosaurs Query, the top 50 retrieved images

We also selected randomness image from Flowers class at third test. We submit the image and retrieve 50 images similar to the query image. Figure 5.9 shows the result of retrieving images similar to the query image. As we can see from Figure 5.9, all the images are similar to the query image and belong to the same class.

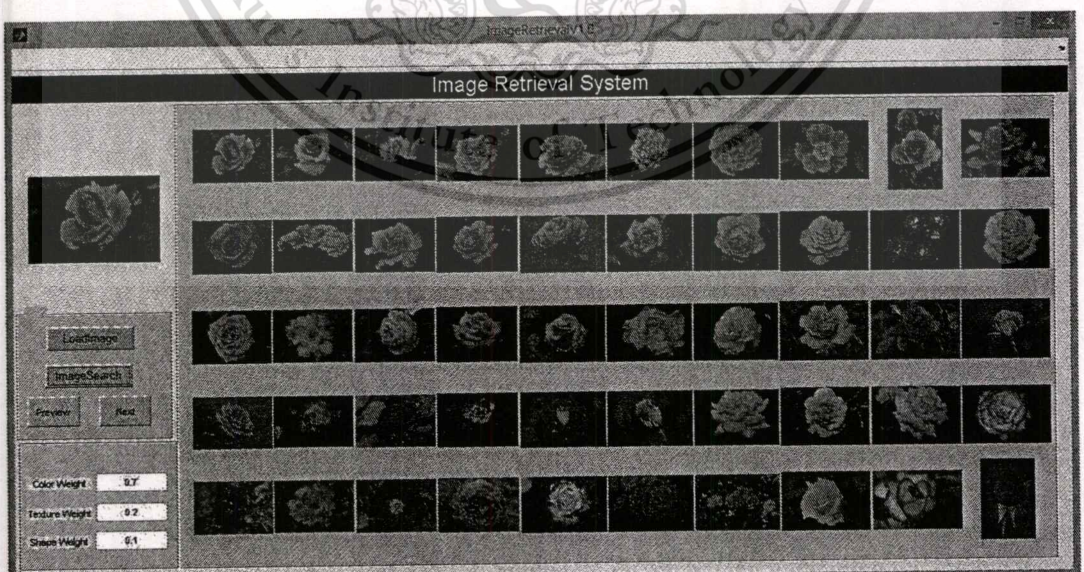

Figure 5.9: Flowers Query, the top 50 retrieved images

For the next test and final test, it shows in the Figure 5.10 and Figure 5.11. We can be noticed that when the retrieved images increases some images do indeed belong to the class, but the other images are not. The system retrieves the incorrect images because they are similar to the query image in their color.

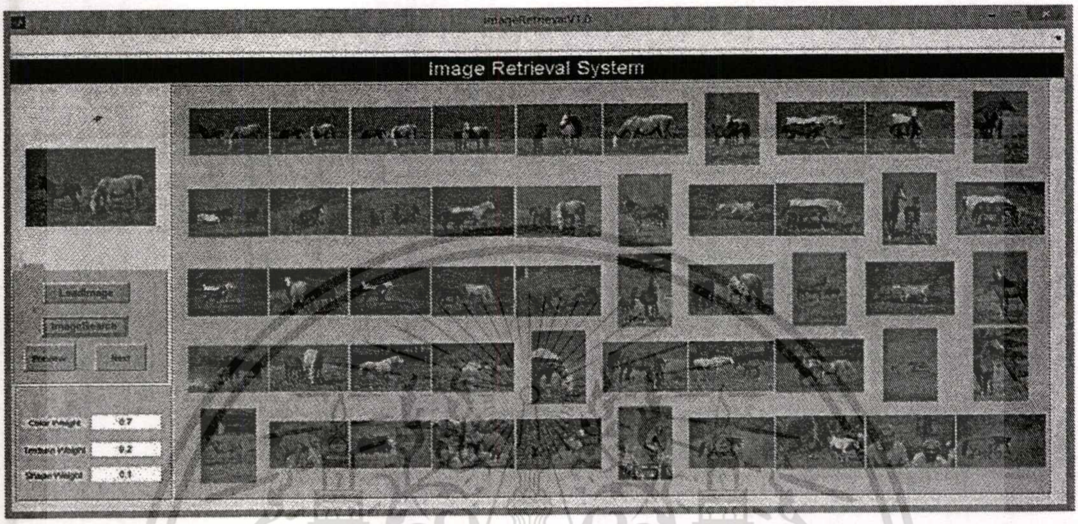

Figure 5.10: Horses Query, the top 50 retrieved images

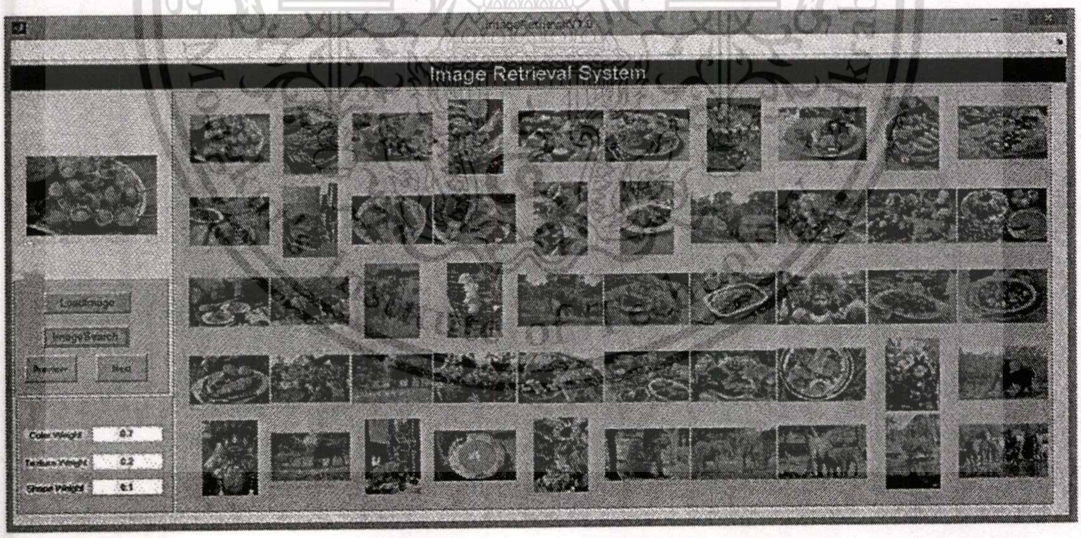

Figure 5.11: Foods Query, the top 50 retrieved images

5.5 Comparison with Other Systems

In this section, we present some earlier CBIR systems' results and compare them with our proposed system. The existing systems we chose for comparison use visual feature to represent images, and they also use the WANG database to evaluate their proposed systems. To evaluate our proposed system, we also randomly selected images from 10 different semantic classes in the database to be a query image and submit it to the system with the same case of retrieved images. We calculate the Precisions for each query in all classes. Then for each class we take the average of all calculated precisions at top 20 retrieved images which are shown in the Table 5.2.

The result of this study is compared against the performance of the existing system. The first system is a user-oriented image retrieval system based on an interactive genetic algorithm proposed by Chin Lai and Ying-Chuan Chen [74], which the color attributes, the entropy based on the gray level co-occurrence matrix and the edge histogram of an image are combined to represent the similarity measure and applied the interactive genetic algorithm to achieve better retrieval efficiency of the retrieval system. The second one is content based image retrieval using color, texture and shape features proposed by P.S. Hiremath and J. Pujari [94], which the local color, the texture descriptors in a coarse segmentation framework of grids, and the shape descriptors in terms invariant moments computed on the edge image were used to achieve the performance of the system.

Table 5.2: Comparison of Precision of the Proposed System with Previously Existing Systems.

Category	Chin Lai and Y. Chuan Chen	P.S. HiremathandJ. Pujari	Proposed System
Africa people and village	0.61	0.48	0.90
Beaches	0.93	0.34	0.90
Buildings	0.85	0.36	0.85
Buses	0.71	0.61	1.00
Dinosaurs	1.00	0.95	1.00
Elephants	0.80	0.48	0.90
Flowers	0.82	0.65	1.00
Horses	0.79	0.70	1.00
Mountain and glaciers	0.56	0.40	0.70
Foods	0.99	0.50	0.85
Average	0.80	0.55	0.91

In the Table 5.2 shows that most of the image classes and the average precision of our proposed system perform significantly better than other systems. Exceptionally, there are some image classes does not increase and losses likely the case of "Buses" and "Foods" categories, the performance of our approach is slightly inferior to that obtained in the Chin Lai and Ying-Chuan Chen [74]. This is because their feature sets derived from the interactive genetic algorithm to achieve better performance, which lets the user evaluate the retrieved images as more or less relevant to the query one.

Figure 5.12 shows the comparison of our proposed system with other system. We represent the precision value of each system for each class using the value from Table 5.2 by a vertical bar. We can see that our proposed system outperforms other existing systems, over most classes, that use the low-level visual features to represent the images and also they use the same database for evaluation.

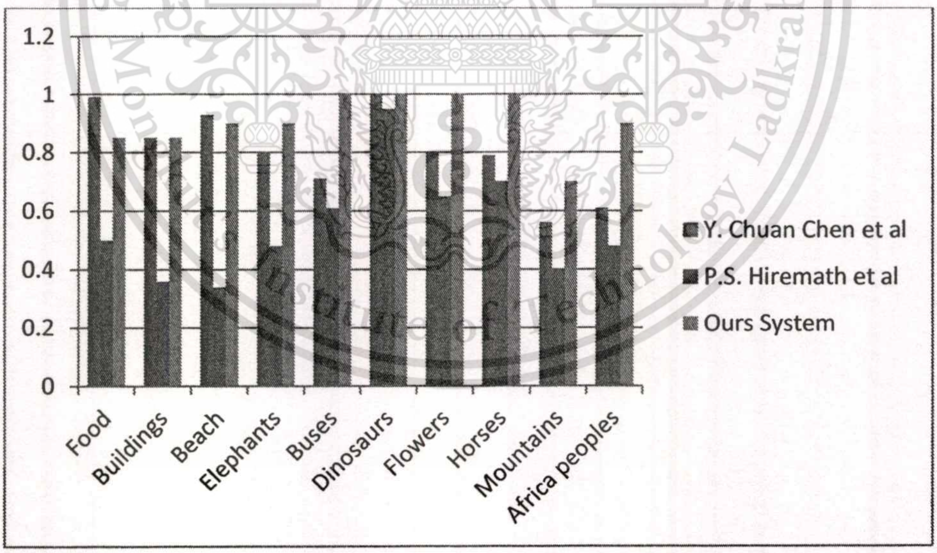

Figure 5.12: Comparison of Precision of the proposed system with previously existed systems.

In addition, we have a comparison our proposed system with the Precision/Recall graph of four experiments to ensure that based on our proposed approach; the combination all of three features including color, texture, and shape features to represent the similarity of the image and the weight variations to achieve higher performance. It can be concluded that our system works well, which shows in the Figure 5.13.

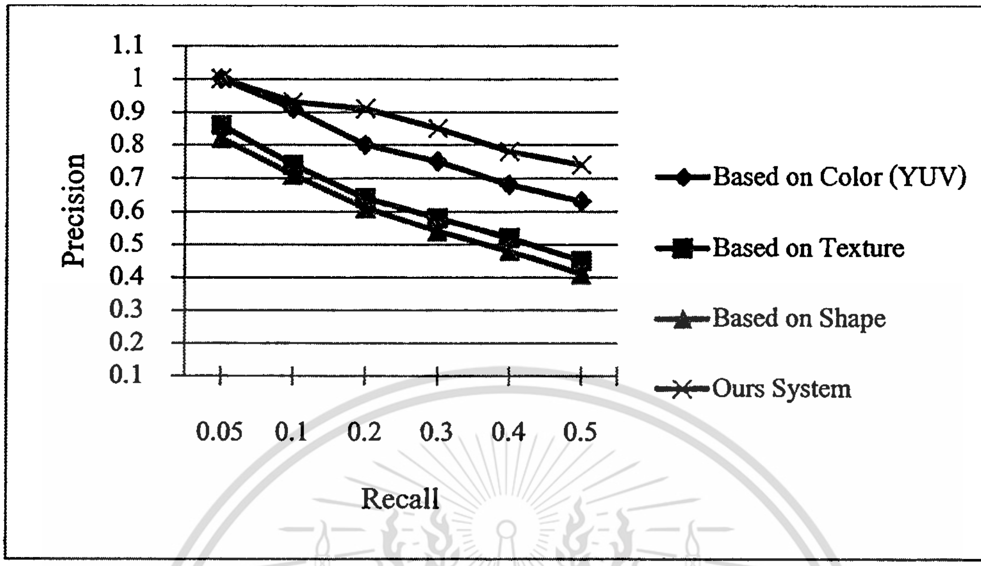

Figure 5.13: Comparison of Precision/Recall of the proposed system with any experiments.

7.6 Chapter Summary

This chapter presented an evaluation of our proposed system that we introduced the database to experiment with our approach. The weight variation technique was proposed to obtain the higher performance for retrieving images based on reduce semantic gap between user and machine. The performance is measured in the term of precision and recall and we organizing four experimental to ensure that our proposed system performed significantly better than other systems. In addition, we compared our system results with other already existing CBIR systems that most of them use the same image database. As a conclusion, the results show that our proposed system works well compared with other existing systems.

Chapter 6

Conclusion and Future Prospects

In this Chapter, we will present a summary of our work, talking about the contribution we have achieved and a conclusion for the work. We also present some recommendation and future works.

6.1 Conclusion

Nowadays, the content-based image retrieval is an active research area in the image processing field. Many researches have been done to develop some algorithms that solve some problems and achieve the efficiency, accuracy when retrieving images and distinguishing between them. Many proposed algorithms use images to extract features and use their features for similarity matching. However, the CBIR is a challenging method of capturing relevant images from a large storage space. Although this area has been explored for decades, no technique has achieved the accuracy of human visual perception in distinguishing images. Whatever, the size and content of the image database enables human being to easily recognize images of the same category.

The goal of this thesis has been to give an overview of the content-based image retrieval process while mainly focusing on reducing the "semantic gap" between low-level image features and the richness of human semantics and enhancing the efficiency and accuracy of the CBIR system. In the thesis, two contributions have been presented: the first one is related to low-level features including color, texture and shape features analyzed and the second one is an analysis of the experimental results by using combination multiple low-level visual features with each feature being weighted differently according to its ability to describe the content of the images.

The first contribution of the thesis concerns on low-level visual feature analysis and feature retrieval including color, texture, and shape features. The color features are extracted from the global characteristic of an image such as the mean value and standard deviation, and the image bitmap is represented to the local feature, those are obtained to compare the efficiency of the color spaces for retrieving images from general database. The texture features are extracted from the entropy based on the Gray Level Co-occurrence Matrix (GLCM) and the edge histogram descriptorsto obtain the efficiency of the texture feature in the experimental results. The shape feature are extracted by the Fourier descriptors derived from the spectral transform on the shape signature to represent three shape invariant; translation, rotation, and scaling invariants. While applying this technique to extract the features from images, it was noticed that the results obtained correspond to visual perception.

The second contribution is obtained from a set of human perception experiments which has been designed and conducted to gain insight into human perception of similar images by the linear combination of all three features and weight variations. The combining descriptors are necessary to get a better retrieval performance and the weight variation is obtained more efficient for the retrieval scheme.

In this work used WANG database that is widely used in CBIR. This database contains 1000 images divided into 10 classes where there are 100 images for each class. Images in the database are friendly and used for evaluating many CBIR systems. We use it to evaluate the performance of our system by calculating the Precision and Recall metrics. In experimental results of all phases, after combination three features, the weight variation is varied to achieve higher retrieval efficiency. It can be concluded that the retrieval performance is depended on the color feature weight 70%, 20% of texture feature weight and 10% of shape feature weight. In addition, we also compared our proposed system with other existing CBIR systems that use the same database, our approach had better than the performance of existing systems such as Chin Lai and Ying-Chuan Chen [74], and P.S. Hiremath and J. Pujari [94].

In this thesis a combination of color, texture and shape features have been used for content-based image retrieval with the case of the general image database. Therefore, the retrieval performance depends on the weight of the color feature. If this algorithm is applied to other databases that are not general image databases, the retrieval performance may be not depended on the weight of the color feature.

6.2 Future Prospects

The approach taken in this thesis and results obtained have generated ideas for new directions and improvements to this research. Possible improvements and further studies on the proposed methods are addressed below:

- To obtain the human perception, the corresponding time for retrieving images would be described and optimized corresponding time by incorporating such as image classification or image clustering.

- To further improve the retrieval results, the texture and shape analysis would be found the possible features for comparing them to get the best algorithm before a combination features.

Reference

[1] H. Tamura, and N. Yokoya, "Image Database Systems: A Survey,"*Pattern Recognition*, vol. 17, no 1, pp.29–49, Sep. 1984.

[2] S. Gerard, C. Buckely, "Term-Weighting Approaches in Automatic Text Retrieval," *Information Processing and Management*, vol. 24, no.5, pp. 513-523, Jan. 1988.

[3] Y. Chen, J. Wang, "Image Categorization by Learning and Reasoning with Regions," *Journal of Machine Learning Research*, vol. 5, pp. 913–939, May 2004.

[4] F. Long, H. Zhang, H. Dagan, and D. Feng, "Fundamentals of content based image retrieval," *Multimedia Signal Processing Book*, Chapter 1, Springer-Verlag, Berlin Heidelberg New York, 2003, pp.1-26.

[5] V. Gudivada and V. Raghavan, "Content-based image retrieval systems," *IEEE Computer*, vol. 28, no 9, pp18-22, Sep. 1995.

[6] M. Kherfi, D. Ziou, and A. Bernardi, " Image Retrieval From the World Wide Web: Issues, Techniques, and Systems," *ACM Computing Surveys*, vol. 36, no. 1, pp. 35–67, March 2004.

[7] A. Smeulders, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349-1380, May. 2000.

[8] J. Caicedo, F. Gonzalez, E. Romero, E. triana, "Design of a Medical Image Database with Content-Based Retrieval Capabilities," *In Proceedings of the 2nd Pacific Rim conference on Advances in image and video technology*, Santiago, Chile, December 17-19, 2007.

[9] Y. Rui, T.S. Huang, et al. "Relevance feedback: A power tool for interactive content-based image retrieval." *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 644-655, September 1998.

[10] A. Vailaya, A. K. Jain et al. "Image classification for content-based indexing." *IEEE Transactions on Image Processing*, pp. 117-130, January 2001.

[11] A. W. M. Smeulders, M. Worring, S, Santini, A. Gupta, and R. Jain. "Content-based image retrieval at the end of the early years." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1349-1380, December 2000.

[12] R. Data, D. Joshi, J. Li, and J. Z. Wang. "Image retrieval: Ideas, influences and trends of the new age."*ACM Computing Surveys*, pp. 1-60, 2008.

[13] A. Chalechale, G. Naghdy, and A. Mertins. "Sketch-based image matching using angular partitioning." *IEEE Transactions on Systems, Man, AndCybernet*, pp. 28-41, January 2005.

[14] Y. Fang, D. Geman, and N. Boujemaa. "An interactive system for mental face retrieval." *ACM SIGMM International workshop on Multimedia Information Retrieval*, pp. 193-200, Singapore, November 2005.

[15] S. M. M. Tahaghoghi, J. A. Thom, and H. E. Williams. "Are two pictures better than one?"*Australasian Database Conference*, pp. 138-144, Queensland, Australia, January 2001.

[16] X.S. Zhou and T. S. Huang. "Relevance feedback in image retrieval: A comprehensive review." *Multimedia Systems*, pp. 536-544, April 2003.

[17] Y. Rui, T. S. Huang, and S. Chang. "Image retrieval: Current techniques, promising direction, and open issues." *Journal of Visual Communication and Image Representation*, pp. 39-62, March 1999.

[18] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada. "Color and texture descriptors." *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 703-715, June 2001.

[19] MPEG-7 overview (version 10). International Standard JTC1/SC29/WG11/ISO/IEC, October 2004.

[20] H. Zhang, R. Rahmani, S. R. Cholleti, and S. A. Goldman. "Local image representations using pruned salient points with application to CBIR." *ACM International Conference on Multimedia*, pp. 287-296, Santa Barbara, USA, October 2006.

[21] Y. Lui, D. S. Zhang, G. J. Lu, and W. Y. Ma. "A survey of content-based image retrieval with high-level semantics." *Pattern Recognition*, pp. 262-282, January 2007.

[22] B. Ko and H. Byun. "Integrated region-based image retrieval using region's spatial relationships." *IEEE International Conference on Pattern Recognition*, pp. 196-199, Quebec, Canada, August 2002.

[23] M. N. Do and M. Vetterli. "Wavelet-based texture retrieval using generalized Gaussian density and kullback-leibler distance." *IEEE Transactions on Image Processing*, pp. 146-158, February 2002.

[24] Y. Bubner, C. Tomasi, and L. J. Guibas. "The earth mover's distance as a metric for image retrieval." *International Journal of Computer Vision*, pp. 99-121, 2000.

[25] J. Li, J. Z. Wang, and G. Wiederhold. "Integrated region matching for image retrieval." *ACM International Conference on Multimedia*, pp. 147-156, Los Angeles, AUS, October 2000.

[26] C. G. M. Snork, M. Worring, and A. W. M. Smeulders. "Early versus late fusion in semantic video analysis." *ACM International Conference on Multimedia*, pp. 399-402, Singapore, November 2005.

[27] J. Peng, B. Bhanu, and S. Qing. "Probabilistic feature relevance learning for content-based image retrieval." *Computer Vision and Image Understanding*, pp. 150-164, July 1999.

[28] Y. Rui and T. S. Huang. "Optimizing learning in image retrieval." *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 39-62, March 2000.

[29] Y. Chen and J. Z. Wang. "A region-based fuzzy feature matching approach to content-based image retrieval." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1252-1267, September 2002.

[30] L. Zhang, F. Lin, and B. Zhang. "Support vector machine learning for image retrieval." *IEEE International Conference on image Processing*, pp. 721-724, Thessaloniki, Greece, October 2001.

[31] S. Tong and E. Chang. "Support vector machine active learning for image retrieval." *ACM International Conference on Multimedia*, pp. 107-118, September 2001.

[32] Y. Chen, X. S. Zhou, and T. S. Huang. "One-class SVM for learning in image retrieval." *IEEE International Conference on Image Processing*, pp. 34-37, October 2001.

[33] X. S. Zhou and T. S. Huang. "Small sample learning during multimedia retrieval using bias map." *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 11-17, Kauai, USA, December 2001.

[34] J. Peng. "Multiclass relevance feedback content-based image retrieval." *Computer Vision and Image Understanding*,pp. 42-67, April 2003.

[35] D. Tao, X. Tang, X. Li, and X. Wu. "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1088-1099, July 2006.

[36] L. Wang, K. L. Chan, and P. Xue. "A criterion for optimizing kernel parameters in KBDA for image retrieval." *IEEE Transactions on Systems, Man, and Cybernetics*, pp. 556-562, June 2005.

[37] D. Zhou, J. Weston, A. Gretton, Q. Bousquet, and B. Schokopf. "Ranking on data manifolds."*Conference on Neural Information Processing Systems*, December 2003.

[38] X. He. "Incremental semi-supervised subspace learning for image retrieval." *ACM International Conference on Multimedia*, pp. 2-8, New York, UAS, October 2004.

[39] N. Vasconcelos and A. Lippman. "A multi-resolution manifold distance for invariant image similarity." *IEEE Transactions on Multimedia*, pp. 127-142, February 2005.

[40] J. He, M. Li, H. J. Zhang, H. Tong, and C. Zhang. "Manifold-ranking based image retrieval." *ACM International Conference on Multimedia*, pp. 9-16, New York, USA, October 2004.

[41] J. Yu and Q. Tian. "Learning image manifolds by semantic subspace projection." *ACM International Conference on Multimedia*, pp. 297-306, October 2006.

[42] X. He, D. Cai, and J. Han. "Learning a maximum margin subspace for image retrieval." *IEEE Transactions on Knowledge and Data Engineering*, pp. 189-201, February 2008.

[43] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H. J. Zhang. "Image classification for content-based indexing." *IEEE Transactions on Image Processing*, pp. 117-130, January 2001.

[44] B. Saux and N. Boujemaa. "Unsupervised robust clustering for image database categorization." *IEEE International Conference on Pattern Recognition*, pp. 259-262, August 2002.

[45] Y. Chen, J. Z. Wang, and R. Krovetz. "Cluster-based retrieval of images by unsupervised learning." *IEEE Transactions on Image Processing*, pp. 1187-1201, August 2005.

[46] D. Cai, X. He, Z. Li, W. Ma, and J. Wen. "Hierarchical clustering of www image search results using visual, textual and link information." *ACM International Conference on Multimedia*, pp. 952-959, New York, UAS, October 2004.

[47] L. Fei-Fei, R. Fergus, and P. Perona. "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories."*Computer Vision and Image Understanding*, pp. 59-70, April 2007.

[48] M. Grubinger. "Analysis and evaluation of visual information systems performance." *PhD Thesis, school of Computer Science and Mathematics Faculty of Health, Engineering and Science Victoria University, Melbourne, Australia*, 2007.

[49] H. Muller, W. Muller, D. M. Squir, S. Marchand-Maillet, and T. Pun. "Performance evaluation in content-based image retrieval." *Pattern Recognition Letters*, pp. 593-601, April 2001.

[50] D. P. Huijsmans and N. Sebe. "How to complete performance graphs in content-based image retrieval"*IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 245-251, February 2005.

[51] A. Torralba, R. Fergus, and W. T. Freeman. "80 million tiny images: A large data set for nonparametric object and scene recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1958-1970, 2008.

[52] M. S. Kamkanhalli and Y. Rui. "Application potential of multimedia information retrieval." *Proceedings of the IEEE*, pp. 712-720, April 2008.

[53] A. Kushki, P. Androutsos, K. N. Plataniotis, and A. N. Venetsanopoulos. "Retrieval of images from artistic repositories using a decision fusion framework." *IEEE Transactions on Image Processing*, pp. 277-292, March 2004.

[54] R. Cutle, Y. Rui, A. Gupta, J. J. Cadiz, I. Tashev. "Distributed meetings: A meeting capture and broadcasting system." *ACM International Conference on Multimedia*, pp. 503-512, December 2002.

[55] J. Z. Wang, J. Li, and G. Wiederhlod. "Simplicity: Semantics-sensitive integrated matching for picture libraries." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 947-963, September 2001.

[56] J. Laaksonen, M. Koskela, and E. Oja. "Picsom-self-organizing image retrieval with MPEG-7 content descriptors." *IEEE Transactions on Neural Networks*, pp. 841-853, July 2002.

[57] R. C. Veltkamp and M. Tanase. "Content-based image retrieval systems: A survey." *Technical report uu-cs-2000-31, Dept. of Computing Science, Utrecht University*, 2000.

[58] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler. "A review of content-based image retrieval systems in medical applications-clinical benefits and future directions." *International Journal of Medical Informatics*, pp. 1-23, February 2004.

[59] K. McDonald and A. F. Smeaton. "A comparison of score, rank and probability-based fusion methods for video shot retrieval." *International Conference Image and Video Retrieval*, pp. 61-70, Singapore, July 2005.

[60] M. Ortega, Y. Rui, K. Chakrabarti, and T. Huang. "Supporting ranked Boolean similarity queries in mars." *IEEE Transactions on Knowledge and Data Engineering*, pp. 909-925, November 1998.

[61] A. Frome, Y. Singer, and J. Malik. "Image retrieval and classification using local distance functions." *Conference on Advances in Neural Information Processing Systems*, pp. 417-424, Vancouver, Canada, December 2006.

[62] T. E. Bjoerge and E. Y.Chang. "Why one example is not enough for an image query." *IEEE International Conference on Multimedia and Expo*, pp. 253-256, June 2008.

[63] X. Y. Jin and J. C. French. "Improving image retrieval effectiveness via multiple queries." *Multimedia Tools and Application*, pp. 221-245, June 2005.

[64] W. W. Y. Ng, A. Dorado, D. S. Yeung, W. Pedrycz, and E. Izquierdo. "Image classification with the use of radial basis function neural networks and the minimization of the locatlized generalization error." *Pattern Recognition*, pp. 19-32, January 2007.

[65] L. Zhu and A. Zhang. "Supporting multi-example image queries in image databases." *IEEE International Conference on Multimedia and Expo*, pp. 697-700, Towers New York City, USA, July 2000.

[66] N. Abbadeni. "An approach based on multiple representation and multiple queries for invariant image retrieval." *International Conference on Advances in Visual Information Systems*, pp. 570-579, Shanghai, China, June 2007.

[67] T. Westerveld and A. P. de Vries. "Multimedia retrieval using multiple examples." *ACM International Conference on Image and Video Retrieval*, pp. 344-352, July 2004.

[68] L. Kish. "Survey Sampling." Wiley, 1995.

[69] R. W. G. Hunt. "Measuring Color." *Ellis Harwood Series in Applied Science and Industrial Technology*, Halsted Press, New York, 1989.

[70] D. H. Ballard and C. M. Brown. "Computer Vision" *Prentice-Hall, Inc*, English wood Cliffs, NJ, 1982.

[71] J. C. Russ. "The Image Processing Handbook." *CRC Press*, Boca Raton, 1995.

[72] A. H. Munsell. "Munsell Book of Color." *Munsell Color Company*, 1942.

[73] A. R. Robertson. "Historical development of CIE recommended color difference equations." *COLOR Research and Application*, June 1990.

[74] C. Chin Lai, Y. Chuan Chen. "A user-oriented image retrieval system based on interactive generic algorithm." *IEEE Transactions. Measurement*, vol. 60, No. 10, October 2011.

[75] E. J. Delp and O. R. Mitcell. "Image coding using block truncation coding." *IEEE Transactions on Communication*, vol. COM-27, No. 9, pp. 1335-1342, September 1979.

[76] M. Tuceryan, and A. K. Jain. "The handbook of pattern recognition and computer vision." *World Scientific Publishing Co.*, pp. 207-248, 1998.

[77] A. Del Bimbo. "Visual Information Retrieval." Morgan Kaufmann Publishers, San Francisco, California, 1999.

[78] MPEG-7 Requirements Group. "MPEG-7: Context, objectives and technical roadimap," *MPEG Vancouver Meeting. ISO/IEC SC29/WG11*, July 1999.

[79] B. S. Manjuneth, J-R. Ohm, V. V. Vasudevan, and A. Yamada. "Color and texture descriptors." *IEEETransactions. Circuits and Systems for Video Technology*, pp. 703-715, 2001.

[80]T. Sikora. "The MPEG-7 visual standard for content description anoverview" *IEEE Transactions. Systems for Video Technology,*vol. 11, no. 6, pp. 696-702, June 2001.

[81] M. Trimeche, "Shape representations for image indexing and retrieval." *Master of Science Thesis, Tampere University of Technology*, May 2000.

[82] R. Datta, J. Li, and J. Wang. "Content-based image retrieval – approaches and trends of the new age." *ACM Computing Surveys*, vol. 40, no. 2, pp. 1-60, April 2008.

[83] C. Teh and T. Roland. "On image analysis y the methods of moments." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 4, pp. 496-513, 1988.

[84] S. Namdagopalan, B. Adiga, and N. Deepak. "A Univeral Model for Content-Based Image Retrieval" *Proceeding. World Academy of Science, Engineering and Technology,* Vol. 36, pp. 659-662, Dec 2008.

[85] M. Lew, N. Sebe, C. Djeraba and R. Jain. "Content-based multimedia information retrieval: State of the art and Challenges." *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 2, no. 1, pp. 1-19, February 2006.

[86] T.-C. Lu and C.-C. Chang, "Color image retrieval technique based on color features and image bitmap," *Inf. Process. Manage.*, vol. 43, no. 2, pp. 461–472, Mar. 2007.

[87] A. Sajjanhar. "A technique for similarity retrieval of shapes."*Master Thesis*, Monash University, Australia, 1997.

[88] C.-L. Huang and D. H. Huang. "A content-based image retrieval system." *Image and Vision Computing*, pp. 149-163,1998.

[89] J. Peter, V. Octterloo. "A contour-oriented approach to shape analysis." *Prentice Hall International (UK)*, Ltd. C 1991.

[90] F. Long, H. Zhang, H. Dagan, and D. Feng, "Fundamentals of content based image retrieval," *Multimedia Signal Processing Book*, Chapter 1, Springer-Verlag, Berlin Heidelberg New York, pp.1-26, 2003.

[91] G. Qian, S. Sural, Y. Gu, and S. Pramanik, "Similarity between Euclidean and cosine angle distance for nearest neighbor queries," *Proceedings of ACM Symposium on Applied Computing*, vol. 12, no. 22, pp. 1232-1237, 2004.

[92] M. Swain, D. Ballard, "Color indexing," *International Journal of Computer Vision*vol. 7, no 1, pp 11-32, Nov 1991.

[93] K. Kaabneh, A. Abdullah, Z. Al-Halalemah. "Video classification using normalized information distance" *IEEE Computer*, pp. 34-40, Jul 2006.

[94] Y. Liu, D..Zhang, G. Lu, and W. Ma, "A survey of content based image retrieval with high-level semantics," *Journal of Pattern Recognition*, vol. 40, pp. 262–282, Nov 2007.

[95] Wikipedia, "Precision and Recall," last modified on 2 September 2013, Available at: http://en.wikipedia.org/wiki/Precision_and_recall

Appendix A

Biography

Personal Information

Name & Surname:	Mr. MekchoneBounthanh
Nationality:	Lao
Birth of date:	June 08, 1985
Place of birth:	Champasack province, Lao P.D.R

Education

High diploma

Field:	Computer Engineering
Duration:	2005 - 2009
Institute:	Department of Computer Engineering and Information Technology, Faculty of Engineering, National University of Laos

Bachelor degree (IT Bridging course)

Field:	Computer Engineering
Duration:	2009 - 2011
Institute:	Department of Computer Engineering and Information Technology, Faculty of Engineering, National University of Laos

Master degree

Field:	Computer Engineering
Duration:	2011 - 2013
Institute:	International College, King Mongkut's Institute of Technology, Ladkrabang, Thailand

Research Interests

Image processing, Image retrieval, Pattern recognition, Image enhancement, and Image classification.

List of International Conference and Proceeding Papers

1. MekchoneBounthanh,Boonwatattachoo, KazuhikoHamamoto, ThaBounthanh "Content-based image retrieval system based on color and image bitmap features", *International Conference on Engineering, Applied Sciences, and Technology* , November 21-24, 2012, Bangkok, Thailand
2. MekchoneBounthanh,Boonwatattachoo, Kazuhiko Hamamoto, ThaBounthanh, "Content-based image retrieval system based on combination multi-features." , *The 2013 International Conference on Embedded System and Intelligent Technology (ICESIT2013)*,January 13-15, 2013, NongKhai, Thailand
3. MekchoneBounthanh,Boonwatattachoo, Kazuhiko Hamamoto, ThaBounthanh, "Content-based image retrieval system based on combined and weighted multi-features" , *The 13th International Symposium on Communications and Information Technology (ISCIT)*, 4-6 September 2013, Samui Island, Thailand

Content-Based Image Retrieval System Based on Color and Image Bitmap Features

Mekchone Bounthanh
International College
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, THAILAND
mekchone_b@fe-nuol.edu.la

Kazuhiko Hamamoto
Department of Information Media Technology,
School of Information and Telecommunication Engineering,
Tokai University, Tokyo 108-8619, JAPAN
hama@keyaki.ce.u-tokai.ac.jp

Boonwat Attachoo
Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, THAILAND
boonwat@kmitl.ac.th

Tha Bounthanh
Faculty of Engineering
National University of Lao
Vientiane 3166, LAOS
tha@fe-nuol.edu.la

Abstract—**The content-based image retrieval (CBIR) field has been actively studied in the several decades. Most research has to find a technique to measure the performance for retrieving more similar image from the digital image databases such as color descriptor, texture descriptor, edge histogram, sharp, etc. In this paper, we proposed the color attributes like the mean value, the standard deviation, and image bitmap of color space to represent the characteristics of the image retrieval. In addition we have to use the characteristics of them for comparison the performance of the color model such as RGB, HSV, YUV, YIQ, YCbCr, and L*a*b* color spaces to get the best color model for relevant retrieved images. Furthermore, An experimental results of our scheme indicated performance of each color space can be retrieved what the type of image.**

Keywords-component; Content-based image retrieval (CBIR), Image bitmap, Color distribution, RGB, HSV, YUV, YIQ, YCbCr, L*a*b*

I. INTRODUCTION

The image retrieval system is a computer system for searching, browsing, and retrieving images from a large database of digital images. Since the primary goal of the content-based image retrieval (CBIR) system is to construct meaningful descriptions of physical attributes from images to facilitate efficient and effective retrieval. Furthermore, a technique which uses visual contents to search images has been an active and a fast advancing research area since the 1990s.

The CBIR system have progressed in four major directions such as global image properties based, region-level feature based, relevance feedback, and semantic based [1],[2]. Some of them have to develop algorithms exploit the low-level features of the image such as color, texture, edge, and shape of an object to help retrieve images. In this paper, we compare the performance of color space for Content-Based Image Retrieval. The color space of color image has to consider six colormodels such as RGB, HSV, YUV, YIQ, YC_bC_r, and L*a*b*, color models are also compared their performance with color attribute and image bitmap features to retrieve images from database. In the next section, we provide a review of related works about CBIR. Then, section III consider a color space or color model and the component of each colormodels. The proposed method to measure the similarity of images is presented in section IV. Section V shows the experimental results and provides comparative performances. Finally, section V is conclusion.

II. RELATED WORKS

The color space is an important feature for perceptual object recognition and classification of images. It has been used in CBIR in conjunction with texture, shape, and structure for retrieving images. Color information is represented as points in three-dimensional color spaces (such as RGB, HSV, YIQ, L*u*v*, L*a*b* [3],[4]). Some of them are hardware-oriented the devices used to reproduce colors. Others are used-inspired (e.g., *L*u*v**, *L*a*b**) as they were defined to quantify color differences as perceived by humans. The signature-based color-spatial image retrieval system has been proposed by Yoo[5]. Color and the spatial distribution within the image are used for the features. In [6], a CBIR scheme based on the global and local color distribution and in an image is presented. The scheme has to use their method with RGB color space to query image and saved a total memory space. "Vadivel*et al.*[7]" have introduced an integrated approach for capturing spatial variation of both color and intensity levels and shows its usefulness in image retrieval application. In addition, a color space features has been used conjunction with many methods for relevant image query from the large database image. "Chun *et al.* [8]" proposed a CBIR method based on an efficient combination of multiresolution color and texture features.As its colorfeatures, color correlograms of the hue and saturation component images in HSV color space are used."Ji-quan ma *et al* [9]" proposed a CBIR system by using HSV color space and wavelet transform approach for feature

AUN/SEED-Net, JICA

extraction.In [10], a L*a*b* color space-used to generate a new method for image analysis, matching, and retrieval. In [11], the Block Truncation Coding (BTC) represent a technique to retrieved iamgs from databases. The result showed that BTC could not be used for color image compression only, BTC also could be used for Content Base Image Retrieval .

III. Image Features

The images in database might be difference from all the others, conversely all images might share certainly common characteristics. Hence, we need the statistical description of images to capture common characteristics and use represent an image with fewer bits. This section is to present a briefly review about the color space to retrieve images .

A. The RGB color space

RGB color space is a widely used for image display. RGB consists of three color components *red, green,* and *blue*. which is called "*additive primaries*" since a color in RGB space is produced by adding them together.However, a major drawback of the RGB space is senseless[4].

B. The HSV color space

HSV color space is widely used in computer graphics and is a more intuitive way of describing color. The three color components are *hue*, *saturation* (lightness) and *value* (*brightness*), HSV is commonly used in image retrieval systems [12]. Hue component is used to distinguish colors, whereas saturation gives a measure of the percentage of white light added to a pure color.

C. The YUV color space

YUV color space is used by the PAL (Phase Alternation Line), NTSC (National Television System Committee), and SECAM (Sequential Couleur Avec Memoire or Sequential Color with Memory) composite color video standards [13]. The black-and-white system used only luminance (Y) information; color information (U and V) was added in such a way that a normal black-and-white picture. Color receivers decoded the additional color information to display a color picture.

D. The YIQ color space

YIQ color space is derived from the YUV color space and is optionally used by the NTSC composite color video standard [14]. The I stands for “in-phase” and the Q for “quadrature” which is the modulation method used to transmit the color information.

E. The YCbCr color space

The YCbCr color space was developed as part of ITU-R BT.601 during the development of a world-wide digital component video standard [15]. The Y in YCbCr denotes the luminance component, and Cb and Cr represent the chrominance factors. The difference between YCbCr and RGB is that the first represents color as brightness and two color difference signal, while the second represent color as red, green, and blue.

F. The L*a*b* color space

The *L*a*b** color space is the most one commonly used color spaces and is modeled based on the human vision system [16]. The *L* component in the *L*a*b** color space corresponds to brightness ranging from 0 (black) to 100 (white), the *a* component corresponds to the measurement of redness(positive values) or greenness (negative values), and the *b*component corresponds to the measurement of yellowness (positive values) or blueness (negative values).

IV. The Proposed Method

This paper uses two features, the color distribution and the image bitmap of an image to represent the similarity between image.

A. The color distribution feature

The color distribution of pixels in an image contains sufficient information. The mean value of pixel colors states the principal color of the image, and the standard deviation of pixel colors can depict the variation of pixel colors [17]. The variation degree of pixel colors in an image is called the color complexity of the image. In the propose scheme, we supposed the P_i represent the color space component of the color image. The mean (μ) and the standard deviation (σ) of a color image are defined as follows:

$$\mu = \frac{1}{M}\sum_{i=1}^{M} P_i, \quad \text{and} \tag{1}$$

$$\sigma = \left[\frac{1}{M-1}\sum_{i=1}^{M}(P_i-\mu)^2\right]^{1/2} \tag{2}$$

Where $\mu = [\mu X, \mu Y, \mu Z]^T$ and $\sigma = [\sigma X, \sigma Y, \sigma Z]^T$ the components of μ and σ are the color space components. The mean value and the standard deviation are the global feature of the image that depicts the global characteristics of the images. For the purpose of enhancing the retrieval accuracy, this paper adopts image bitmap as the local to describe the local characteristics of the image.

B. The image bitmap feature

In the first step to generate the image bitmap, the scheme divides the image into several non overlapping blocks [18]. Let $B_j = \{b_1, b_2, \ldots, b_k\}$ be the *j*th block of the image, where $1 \le j \le m$. The symbol *k* is the total number of pixels in the block, and m is the total number of blocks in the image. In the second step, the scheme computes the mean value for each block. Let μ_{B_j} be the mean value of the block B_j that is computed using the expression

$$\mu_{B_j} = \frac{1}{k}\sum_{i=1}^{k} b_i. \tag{3}$$

Where $\mu_{B_j} = \left[\mu X_{B_j}, \mu Y_{B_j}, \mu Z_{B_j}\right]^T$ and the components of μ_{B_j} are the color space components. In the third step, the scheme compares the mean values of the block μ_{B_j} with the mean value of the image μ to determine the characterization of the block B_j for generating the image bitmap. Let $I = [IX, IY, IZ]^T$ be the image bitmap. Each component in I is expressed as $IX = [IX_1, IX_2 \dots IX_m]$, $IY = [IY_1, IY_2 \dots IY_m]$,and $IZ = [IZ_1, IZ_2 \dots IZ_m]$, respectively. The entries are represented by

$$IX_j = \begin{cases} 1, \text{ if } \mu X_{B_j} \geq \mu X \\ 0, \text{ otherwise} \end{cases} \tag{4}$$

$$IY_j = \begin{cases} 1, \text{ if } \mu Y_{B_j} \geq \mu Y \\ 0, \text{ otherwise} \end{cases} \tag{5}$$

$$IZ_j = \begin{cases} 1, \text{ if } \mu Z_{B_j} \geq \mu Z \\ 0, \text{ otherwise} \end{cases} \tag{6}$$

C. The similarity measure of the features

In the experiment system, we use two different measurements to evaluate the similarity which are the global features and the local features. For the global features, μ and σ, the scheme uses Euclidean distance to calculate the similarity. On the other hand, the local feature is used the Hamming distance to evaluate the distance between the two bitmaps. After that, the overall similarity is obtained by linearly combining of these two similarity values.

For two images, the Hamming distance H used to evaluate the image bitmap similarity is defined by

$$H(BM^d, BM^q) = \sum_{j=1}^{m}(IX_j^d - IX_j^q) + \sum_{j=1}^{m}(IY_j^d - IY_j^q) + \sum_{j=1}^{m}(IZ_j^d - IZ_j^q) \tag{7}$$

Where IX_j^d is the *j*th component of the bitmap *IX* of image *d* from the database and the IX_j^q is the *j*th component of the bitmap *IX* of the query image *q*. *BM* means the image bitmap feature of both images. Finally, the similarity measure is formed as follows:

$$SIM(d,q) = \sqrt{\sum_{t\in(X,Y,Z)}(\mu_t'^d - \mu_t'^q) + \sum_{t\in(X,Y,Z)}(\sigma_t'^d - \sigma_t'^q)} + \frac{H(BM^d, BM^q)}{3\times m} \tag{8}$$

Where μ_t' and σ_t' represent the normalized mean value and standard deviation of the image from the database d and the query image q in color space respectively.

V. EXPERIMENTAL RESULT

This section assesses the performance of content-based image retrieval in the six color space (RGB, HSV, YUV, YIQ, YCbCr and L*a*b*) which is based on the global and local features to measure their performance. In this experiment, we used the database of the SIMPLIcity project covering a wide range of semantic categories of natural scenes with artificial objects for experiments [19]. The database is partitioned into ten categories, including African people and village, beach, building, buses, dinosaurs,elephants, flowers,horses, mountains and glaciers, food, etc., and each category contains 100 images (Fig. 1).

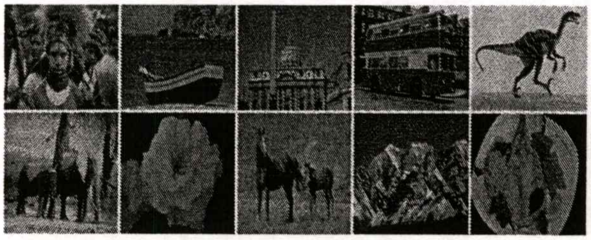

Fig. 1.Sample images of each category of the image database.

To evaluate The performance of proposed CBIR methods, we examined by using the evaluated point value of the precision rate p and recall rate r for considering the number of relevant images to query the retrieval precision and recall are defined as

$$p_M = \frac{n_r}{M}, \text{ and} \tag{9}$$

$$r_M = \frac{n_r}{N_r} \tag{10}$$

Where M is the total number of retrieved images, $n_r$is the number of relevant correct images retrieved,N_r represents the total number of relevant images available in the database. It is assumed that the user inspects the m first images of the ranked list. So for every $m = 1, 2, \dots, M$, precision $p_m$and the recall r_m are calculated. In each experiment, evaluation of the retrieval precision is performed ten images were randomly selected from each specific category in the database as query images and we used the top 20 retrieved images to compute the precision and recall.

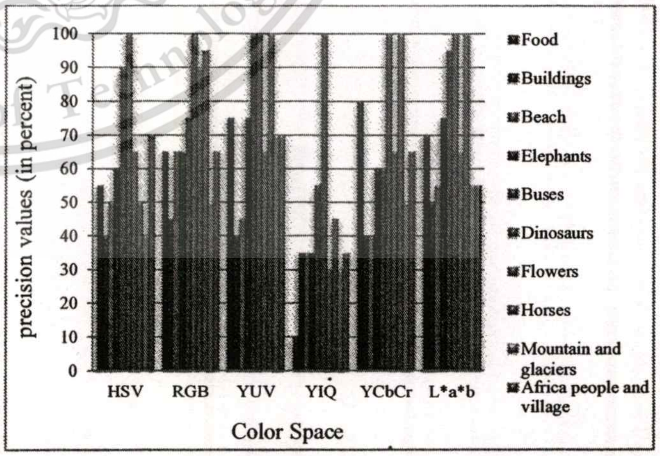

Fig. 2.Performance precision of retrieving images in the each color space.

Fig.2 gives crossover points of precision value of the color distribution and image bitmap feature based on CBIR techniques for all considered difference color space testing on the Wang image database. The result showed that dinosaurs category is better than another color space in higher crossover point values.

The precision crossover value of YIQ color space is lower than other color space to retrieved the same image categories. In addition,TABLE I and TABLE II is average of precision value and average of recall value respectively, the resultsare average of value in three times for one categary by using image query random. These are compared the performance of all color spaces for relevant image retrieved from database.

TABLE I. AVERAGE OF PRECISION VALUE (IN PERCENTAGE)

Category	RGB	HSV	YUV	YIQ	YCbCr	Lab
Africa people and village	63.3	58.2	73.3	35	68.3	65
Beach	73.3	45	53.3	33.3	55	61.7
Buildings	45	30	46.7	36.7	45	48.3
Buses	73.3	83.3	98.3	46.7	66.7	95
Dinosaurs	100	100	100	100	100	100
Elephants	75	60	75	38.3	66.7	70
Flowers	90	35	61.7	20	65	58.3
Horses	78.3	41.7	81.7	51.7	78.3	76.7
Mountain and glaciers	50	43.3	55	33.3	53.3	51.7
Food	63.3	58.3	73.3	21.7	85	78.3
Average	71.7	55.2	72.7	41.7	68.3	70.5

TABLE II. AVERAGE OF RECALL VALUE

Category	RGB	HSV	YUV	YIQ	YCbCr	Lab
Africa people and village	11.7	12.7	14.7	7	13.7	13
Beach	9	14.7	10.7	6.7	10.7	12.3
Buildings	6	9	9.3	7.3	9	9.7
Buses	16.7	14.7	19.7	8.7	13.3	19
Dinosaurs	20	20	20	20	20	20
Elephants	12	15	15	7.7	13.3	14
Flowers	7	18	12.3	4	13	11.7
Horses	8.3	15.7	16.3	10	15.7	15.3
Mountain and glaciers	8.7	10	11	6.7	10.7	10.3
Food	11	13.7	16.3	5.3	17	15.7
Average	11	14.3	14.5	8.4	13.6	14.1

VI. CONCLUSION

This paper proposed a content-based image retrieval system which is based on the color distribution and bitmap in different color spaces to query image with a Wang image database. The color distribution features are the global characteristic of an image such as the mean value and the standard deviation. The image bitmap represented the local feature to depict the local characteristics of the image. We used both features to compare the performance of six color space such as RGB, HSV, YUV, YIQ, YCbCr, and L*a*b* color space for retrieving image from the image database. The experimental results of the proposed approach have shown the percentage in retrieval performance. The YUV color space is better than the performance of other color spaces and the YIQ color space is smaller than other space.

ACKNOWLEDGMENT

The authors would like to thank the AUN/SEED-Net for the financial support to Mr.MekchoneBounthanh to study at king Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, THAILAND.

REFERENCES

[1] J. Han, K. N. Ngan, M. Li, and H. J. Zhang, "A memory learning framework for effective image retrieval, " *IEEE Trans. Image Process.*, vol. 14,no. 4, pp. 511-524, Apr, 2005.

[2] M. Antonelli, S. G. Dellepiane, and M. Goccia, " Design and implementation of Web-based systems for image segmentation and CBIR," *IEEE Trans. Instrum. Meas*, vol. 55, no. 6, pp. 1869–1877, Dec. 2006.

[3] A. del Bimbo. *Visual Information Retrieval*. Morgan Kaufmann Publishers, San Francisco, CA, USA, 1999.

[4] K. SHIH, C. HUNG, AND CHUAN-HO KAO,"An Intelligent Content-based Image Retrieval System Based on Color, Shape and Spatial Relations," *Proc. Natl. Sci. Counc. ROC(A)*, Vol. 25, No.4, pp. 232-243, Nov, 2000.

[5] H.-W. Yoo, H.-S. Park, and D.-S. Jang, "Expert system for color image retrieval," *Expert Syst. Appl.*, vol. 28, no. 2, pp. 347–357, Feb. 2005.

[6] T.-C. Lu and C.-C. Chang, "Color image retrieval technique based on color features and image bitmap," *Inf. Process. Manage.*, vol. 43, no. 2, pp. 461–472, Mar. 2007.

[7] A. Vadivel, S. Sural, and A. K. Majumdar, "An integrated color and intensity co-occurrence matrix," *Pattern Recognit. Lett.*, vol. 28, no. 8, pp. 974–983, Jun. 2007.

[8] Y. D. Chun, N. C. Kim, and I. H. Jang, "Content-based image retrieval using multiresolution color and texture features," *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1073–1084, Oct. 2008.

[9] Ji-quan ma. "Content-based image retrieval with HSV color space and texture feature," *International Conference*..

[10] A. Mojsilovic, J. Hu, and E. Soljanin, "Extraction of perceptually important colors and similarity measurement for image matching, retrieval, and analysis",*IEEE Trans.Image process.*, vol. 12, No. 1, Nov, 2002.

[11] Guoping Qiu, "Color image indexing using BTC"*IEEE Trans. Image process.*, vol. 12, No. 1, Jan, 2003.

[12] G. Deshpande, M. Borse, "Image Retrieval with the use of different color spaces and the texture feature," *Inter. Conf. IACSIT Press, Singapore*. vol. 9, pp 273-278, 2011.

[13] R. Sudhir, "An efficient CBIR technique with YUV color space and texture features," *Computer Engineering, ISSN 2222-1719*, vol 2, No 6, 2011.

[14] F. Liu, "A novel similarity measure method for CBIR system," *IEEE Trans. Image Process.*, vol 5, pp. 335-335, Aug, 2009.

[15] D. Chai, and A. Bouzerdoum, "A bayesian approach to skin color classification in YCbCr color space," *IEEE Trans. Image Process.*, vol 2, pp. 421-424, 2000.

[16] J. Pujari, S. N. Pushpalatha, "Content-based image retrieval using color and shape descriptors," *IEEE Trans. Image Process.*, pp. 239-242, Dec, 2010.

[17] C. Chin Lai, Y. Chuan Chen, " A user-oriented image retrieval system based on interactive genetic algorithm," *IEEE Trans. Measurement.*, vol. 60, No. 10, Oct, 2011.

[18] E. J. Delp and O. R. Mitchell, "Image coding using block truncation coding," *IEEE Trans. Commun.*, vol COM-27, No. 9, pp. 1335-1342, Sep, 1979.

[19] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantic sensitive integrated matching for picture libraries," *J. Inf. Tchnol. Appl.*, vol. 1, No. 3, pp. 205-214, Dec, 2006.

Content-Based Image Retrieval System Based on Combination Multi-Features

Mekchone Bounthanh
International College
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, THAILAND
mekchone_b@fe-nuol.edu.la

Kazuhiko Hamamoto
Department of Information Media Technology,
School of Information and Telecommunication Engineering,
Tokai University, Tokyo 108-8619, JAPAN
hama@keyaki.ce.u-tokai.ac.jp

Boonwat Attachoo
Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, THAILAND
boonwat@kmitl.ac.th

Tha Bounthanh
Faculty of Engineering
National University of Laos
Vientiane 3166, LAOS
tha@fe-nuol.edu.la

***Abstract*— Color, edge and texture features have been the primitive image descriptors in Content Based Image Retrieval (CBIR) systems. In this paper, we proposed a novel framework for combining all of three i.e. color, edge and texture features to achieve higher retrieval efficiency. The color feature is extracted by quantifying the YUV color space and the color attributes like the mean value, the standard deviation, and the image bitmap of YUV color space is represented. The texture features are obtained by the entropy based on the gray level co-occurrence matrix and the edge histogram descriptor of an image. When computing the similarity between the query image and target image in the database, normalization information distance is also used for adjusting distance values into the same level. And then the linear combination has used to combine the normalized distance of the color, edge and texture features to obtain the similarity as the indexing of image. Furthermore, An experimental results indicated, the proposed technique indeed outperforms other schemes in terms of the accuracy and efficiency.**

Keywords-component; Content-based image retrieval (CBIR), Image bitmap, Color distribution, YUV, Gray level co-occurrence matrix

I. INTRODUCTION

The image retrieval system is a computer system for searching, browsing, and retrieving images from a large database of digital images. Due to the tremendous growth of digital image, results to the traditional text-based retrievals based on keywords are not sufficient enough to resolve image retrieval. The content-based image retrieval (CBIR) can greatly enhance the accuracy and efficiency of retrieving and managing the huge number of image data. So the CBIR system attracts more and more attention of researchers. In the recent year, content-based image retrieval (CBIR) has been an active research area in image processing. Content based image retrieval has many application areas such as architectural design, education, commerce, military, medical diagnosis, biomedicine and web image classification.

All most of The CBIR research is finding technique to measure the performance for retrieving more similar image from the image databases of retrieval scheme [1, 2, 3]. Some of them have been developed algorithms exploit the low-level features of the image such as color, texture,edge and shape of an object to help retrieve images. In this paper, we have been combination three features including color, edge and texture features to achieve higher retrieval efficiency of retrieval scheme. In addition, we have been comparing the accuracy and efficiency of the scheme with other the retrieval scheme. In the next section, we provide a review of related works about CBIR. Then, the proposed method to measure the similarity of images is presented in section III. Section IV shows the experimental results and provides comparative performances. Finally, section V is a conclusion.

II. RELATED WORKS

This section introduces some important literatures review which content based image retrieval (CBIR) was first introduced by Kato in 1992 [4]. The term of CBIR is widely used for retrieving desired images from a large collection, which is based on extracting features such as color, texture, edge and shapes from images themselves. For the color feature extraction, the YUV color space is widely used for CBIR scheme. In [5, 6] the YUV color space were extracted a characteristic of images by discrete wavelet transform (DWT) and compares the most significant wavelet coefficients of the YUV components of the query image and those of the images in the database to integrate similarity scores result from the matching processes. Some of experiment result of them were compares the performance for retrieving image of the YUV color space higher more than the RGB and HSV color space when has used the same work. In addition of color space, texture and edge are also important features for retrieving image of the retrieval scheme. In [7], a texture was used to retrieve images from database, which is considered six

AUN/SEED-Net, JICA

different texture features produced from the co-occurrence matrix and used normalized correlation to adopt as a similarity function. In [8, 9], a texture also was used to retrieve image from database, but in these cases has found a novel technique of texture applied to the CBIR field. Some of them had a combined texture feature with edge histogram to get better the performance in the retrieving image of the retrieval scheme. " Wang Beizhen *et al* [10]" have improved dominant color histogram and edge histogram and evaluates the retrieving result with the criteria provided by MPEG-7. Although, the visual feature is important in the retrieving scheme, but the feature extraction and the method to extract the characteristic is also helping to improve the performance of the retrieval scheme. So, in [11], a CBIR scheme based on the global and local color distribution and in an image is presented. The scheme had used their method with RGB color space to query image and saved a total memory space. "Vadivel *et al.*[12]" have introduced an integrated approach for capturing spatial variation of both color and intensity levels and shows its usefulness in image retrieval application. In addition, a color space features has been used conjunction with many methods for relevant image query from the large database image. In [13], the Block Truncation Coding (BTC) represented a technique to retrieve images from databases. The result showed that BTC could not be used for color image compression only, BTC also could be used for Content Base Image Retrieval .

III. The Proposed Method

One of the important issues in the image retrieval system is the choice of appropriate the feature extraction method and corresponding similarity measure. In This section, we propose the feature extraction approach of the low-level visual features.

A. The color distribution feature

In the proposed approach, the RGB images are first transformed to the YUV color space for two reasons: 1) efficiency and 2) ease of extracting the feature based on the color tones. It is well-know that the Y component of an image is much more significant than the U and V component can be much less than those kept for the Y component without sacrificing the accuracy rate. Moreover, compared with the features derived from the RGB color space, the features derived from the YUV color space are more suitable for the purpose of screening the image based on a certain color tone by weighting the corresponding features. So, we had choices the YUV color space apply to this research and it can be defined in [5, 6].

The color distribution of pixels in an image contains sufficient information. The mean value of pixel colors states the principal color of the image, and the standard deviation of pixel colors can depict the variation of pixel colors [14]. The variation degree of pixel colors in an image is called the color complexity of the image. In the propose scheme, we supposed the P_i represent the color space component of the color image. The mean (μ) and the standard deviation (σ) of a color image are defined as follows:

$$\mu = \frac{1}{M}\sum_{i=1}^{M} P_i \quad , \text{ and} \tag{1}$$

$$\sigma = \left[\frac{1}{M-1}\sum_{i=1}^{M}(P_i-\mu)^2\right]^{1/2} \tag{2}$$

Where $\mu = [\mu Y, \mu U, \mu V]^T$ and $\sigma = [\sigma Y, \sigma U, \sigma V]^T$ the components of μ and σ are the color space components. The mean value and the standard deviation are the global feature of the image that depicts the global characteristics of the images. For the purpose of enhancing the retrieval accuracy, this paper adopts image bitmap as the local to describe the local characteristics of the image.

In the first step to generate the image bitmap, the scheme divides the image into several non overlapping blocks [15]. Let $B_j = \{b_1, b_2, \ldots, b_k\}$ be the jth block of the image, where $1 \le j \le m$. The symbol k is the total number of pixels in the block, and m is the total number of blocks in the image. In the second step, the scheme computes the mean value for each block. Let μ_{B_j} be the mean value of the block B_j that is computed using the expression

$$\mu_{B_j} = \frac{1}{k}\sum_{i=1}^{k} b_i. \tag{3}$$

Where $\mu_{B_j} = \left[\mu Y_{B_j}, \mu U_{B_j}, \mu V_{B_j}\right]^T$ and the components of μ_{B_j} are the color space components. In the third step, the scheme comparing the each pixel of the block with the mean value of the block μ_{B_j} and summarized the normalized value of each block to determine the characterization of the block B_j for generating the image bitmap. Let $I = [IY, IU, IV]^T$ be the image bitmap. Each component in I is expressed as $IY = [IY_1, IY_2 \ldots IY_m]$, $IY = [IU_1, IU_2 \ldots IU_m]$,and $IV = [IV_1, IV_2 \ldots IV_m]$, respectively. The entries are represented by

$$IY_j = \begin{cases} 1, \text{if } Y_i \ge \mu Y_{B_j} \\ 0, \text{ otherwise} \end{cases} \tag{4}$$

$$IU_j = \begin{cases} 1, \text{if } U_i \ge \mu U_{B_j} \\ 0, \text{ otherwise} \end{cases} \tag{5}$$

$$IV_j = \begin{cases} 1, \text{if } V_i \ge \mu V_{B_j} \\ 0, \text{ otherwise} \end{cases} \tag{6}$$

B. Texture descriptor feature

Texture is also a visual feature that refers to innate surface properties of an object and their relationship to the surrounding environment. In this case, we are using gray-level co-occurrence matrix (GLCM) for extracting texture features of images [16], which is a simple and effective method for representing texture. Co-occurrence matrices display the gray level spatial dependency along different angular relationships, horizontal, vertical, and two diagonal directions, in a block or an image. The GLCM is specified by the relative frequencies

$P(i, j, d, \theta)$ with which two pixels, separated by a distance *d*, occur in a texture along the direction of angle , one with gray level *I* and the other with gray level *j*, The GLCM is mathematically defined as follows:

$$P(i, j, d, \theta) = \#\{(x_1, y_1)(x_2, y_2) | g(x_1, y_1) = i, g(x_2, y_2) = j, \\ |(x_1, y_1) - (x_2, y_2)| = d, \angle((x_1, y_1)(x_2, y_2)) = \theta\} \quad (7)$$

Where # denotes the number of occurrences inside of the window, with *i* and *j* being the intensity levels of the first pixel and the second pixel at positions (x_1, y_1) and (x_2, y_2), respectively.

In order to simplify and reduce the computational effort, we computed the GLCM according to one direction (i.e., = 0°) with a given distance *d* (= 1) and only select one of the features as retrieval feature: Entropy. The entropy (*E*) measures the disorder of an image and it achieves its largest value when all elements in *P* matrix and are defined as follows:

$$E = -\sum_{i,j} P_{i,j} \log P_{i,j} \quad (8)$$

C. Edge descriptor feature

Edge in an image is an important low-level feature, it can describe both shape and texture features, which are also essential elements for content-based image retrieval. One way to represent such an important edge feature that is used a histogram. The edge histogram in the image space represents the frequency and the directionality of the brightness change in the image. The edge histogram descriptor (EHD) [17] represents the local edge distribution in the image, it's useful in image matching when the texture is not homogeneous. The feature extraction process of EHD consists of the following stages.

1) *An image is divided into 4 x 4 sub-images.*

2) *Each sub-image is further partitioned into non-overlapping image blocks with a small size.*

3) *The edge in each image block is categorized into five types: vertical, horizontal, 45° diagonal, 135° diagonal, and non-directional edge.*

4) *Thus, the histogram for each sub-image represents the relative frequency of occurrence of the five types of edges in the corresponding sub-image.*

5) *After examining all image blocks in the sub-image, the five-bin values are normalized by the total number of blocks in the sub-image. Finally, the normalized bin values are quantized for the binary representation.*

D. The similarity measure of the features

In the experimental system, we use linear combinations of three feature distance measurements to evaluate the similarity which is the distance of color features, the distance of texture feature, and the distance of the edge histogram descriptor feature. When computing the similarity of each feature between the query image and the target image in the database. The normalized information distance (NID) [18] was used to normalize the value of each distance adjusted to the same level: the distance has a value between 0 and 1. Firstly, the distance of the color features have described in the term of global features and the local features. The local feature (S_1) is used the Hamming distance to evaluate the distance between the two binary image bitmaps. On the other hand, the global feature (S_2), μ and σ, the scheme use Euclidean distance to calculate the similarity.

For two images, the Hamming distance *H* used to evaluate the image bitmap similarity is defined by

$$H(BM^d, BM^q) = \frac{1}{3 \times m}\left\{\sum_{j=1}^{m}(IY_j^d - IY_j^q) + \sum_{j=1}^{m}(IU_j^d - IU_j^q) + \sum_{j=1}^{m}(IV_j^d - IV_j^q)\right\} \quad (9)$$

$$S_1(d, q) = \frac{H(BM^d, BM^q) - min\{H(BM^d, BM^q)\}}{\max\{H(BM^d, BM^q)\}} \quad (10)$$

Where IY_j^d is the *j*th component of the bitmap *IY* of image *d* from the database and the IY_j^q is the *j*th component of the bitmap *IY* of the query image *q*. *BM* means the image bitmap feature of both images. So, the global color feature distance similarity measure is formed as follows:

$$E(d, q) = \sqrt{\sum_{t \in (Y,U,V)} (\mu_t^d - \mu_t^q)^2 + \sum_{t \in (Y,U,V)} (\sigma_t^d - \sigma_t^q)^2} \quad (11)$$

$$S_2(d, q) = \frac{E(d, q) - min\{E(d, q)\}}{max\{E(d, q)\}} \quad (12)$$

Where μ_t and σ_t represent the mean value and standard deviation of the image from the database *d* and the query image *q* in color space respectively. After that, we used the linear combination between the local and global features to represent the distance of the color features (d_{color}).

Second one, the normalized distance of the texture can be defined as follows:

$$d_{texture}(d, q) = \frac{|E^d - E^q| - min\{|E^d - E^q|\}}{max\{|E^d - E^q|\}} \quad (13)$$

Where E^d and E^q represents the entropy of the image from the database *d* and the query image *q* respectively.

The third one, the normalized distance of the edge histogram descriptor can be defined as follows:

$$d_{EHD}(d, q) = \frac{|EHD^d - EHD^q| - min\{|EHD^d - EHD^q|\}}{max\{|EHD^d - EHD^q|\}} \quad (14)$$

Where EHD^d and EHD^q represents the edge histogram descriptor of the image from the database d and the query image q respectively. Finally, the similarity measurement is obtained by linearly combining these of all similarity values.

$$SIM(d,q) = d_{color} + d_{texture} + d_{EHD} \quad (15)$$

IV. EXPERIMENTAL RESULT

In this experiment, we used the database of the SIMPLIcity project covering a wide range of semantic categories of natural scenes with artificial objects for experiments [19]. The database is partitioned into ten categories, including African people and village, beach, building, buses, dinosaurs,elephants, flowers,horses, mountains and glaciers, food, etc., and each category contain 100 images (Fig. 1). All the images are in RGB color space. They are in the JPEG format with size in 256×384 and 384×256 pixels.

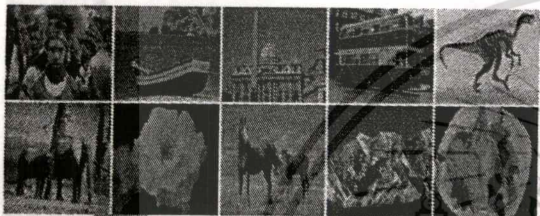

Fig. 1 Sample images of each category of the image database.

A. The retrieval accuracy and efficiency

To evaluate The performance of proposed CBIR methods, we examined by using the evaluated point value of the precision rate p and recall rate r for considering the number of relevant images to query the retrieval. The precision gives information about the relative effectiveness of the retrieval scheme and the recall giving information about the absolute accuracy of the retrieval scheme, which are defined as

$$p_m = \frac{n_r}{M}, \text{ and} \quad (9)$$

$$r_m = \frac{n_r}{N_r} \quad (10)$$

Where M is the total number of retrieved images, n_r is the number of relevant correct images retrieved, N_r represents the total number of relevant images available in the database. It is assumed that the user inspects the m first images of the ranked list. So, for every $m = 1, 2, ..., M$, precision p_m and the recall r_m are calculated. In each experiment, evaluation of the retrieval precision is performed ten category images were randomly selected from each specific category in the database as query images and count the precision and recall separately after getting the return result. The responding of image, we always used the top 20 retrieved images to compute the precision and recall. For the efficiency and accuracy of our retrieval scheme is shown on the Fig. 2.

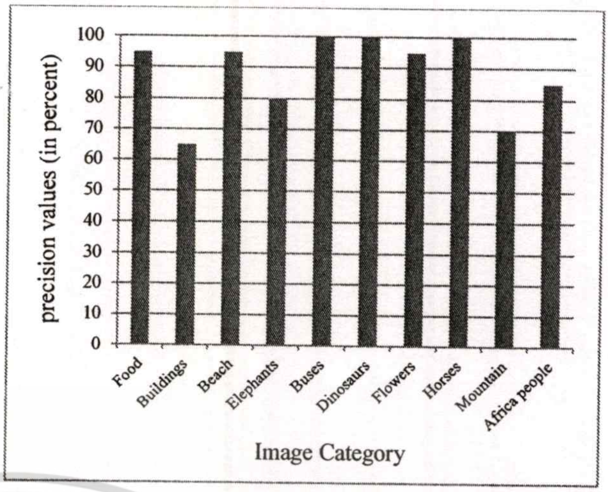

Fig. 2 Performance precision of retrieving images in the each image category.

Fig.2 gives crossover points of precision value of CBIR based on our method, it is clear that the performance of our proposed approach can be given the higher performance. For the precision values of each image categories are more than sixty percentage. Furthermore, the result showed that dinosaurs, buses, and horses categories are better than other image category in higher crossover point values.

B. Comparison with other method

To evaluate the superiority of our approach, we had compared our experimental result with other retrieval systems. The first system is a user-oriented image retrieval system based on an interactive genetic algorithm [14], which the color attributes, the entropy based on the gray level co-occurrence matrix and the edge histogram of an image are combined to represent the similarity measure and applied the interactive genetic algorithm to achieve better retrieval efficiency of the retrieval scheme. Second one is image retrieval based on color and image bitmap features [20]. They have also used the YUV color space to generate the feature extraction and used the global and local feature of the color space to represent the similarity measurement of their retrieval scheme. The comparison is considered the performance during the term of the precision and recall, there are shown in TABLE I and TABLE II respectively. Actually, the propose approach of our scheme has achieved better than the other systems. Exceptionally, in the case of the "buildings" and "food" categories, the performance of our approach is slightly inferior to that obtained in [14]. This is because their feature sets derived from the interactive genetic algorithm to achieve better performance, which lets the user evaluate the retrieved images as more or less relevant to the query one. For our approach also used the same low-level visual features, but some of the features, we used different technique to extract a feature and the YUV color space is suitable for image retrieval more than the other color space in the same work (certify in [5, 6, 20]). So, our proposed approach scheme had a higher performance more than the other systems illustrated.

TABLE I. AVERAGE OF PRECISION VALUE (IN PERCENTAGE)

Category	[20]	[14]	Ours
Africa people and village	73	61	85
Beach	53	93	95
Buildings	47	85	65
Buses	98	71	100
Dinosaurs	100	100	100
Elephants	75	80	80
Flowers	62	82	95
Horses	82	79	100
Mountain and glaciers	55	56	70
Food	73	99	95
Average	71.8	80.6	88.5

TABLE II. AVERAGE OF RECALL VALUE

Category	[20]	[14]	Ours
Africa people and village	15	12	17
Beach	11	18	19
Buildings	9	17	13
Buses	19	14	20
Dinosaurs	20	20	20
Elephants	15	16	16
Flowers	12	16	19
Horses	16	15	20
Mountain and glaciers	11	11	14
Food	16	19	19
Average	14.4	15.8	17.7

V. CONCLUSION

This paper proposed a content-based image retrieval system which is based on the combination multi-features with a Wang image database. The color feature is extracted by quantifying the YUV color space. The color distribution features are the global characteristic of an image such as the mean value and the standard deviation, and the image bitmap represented the local feature to depict the local characteristics of the image. In addition, the entropy based on the gray level co-occurrence matrix and the edge histogram are representing a texture feature. There are also normalized by the normalized information distance and linear combination was used to combine to represent the similarity measurement of our retrieval scheme. The experimental results of the proposed approach have shown the percentage in retrieval performance and compared the efficiency and accuracy with other the retrieval scheme, which the our approach had better than the performance of other systems.

ACKNOWLEDGMENT

The authors would like to thank the AUN/SEED-Net for the financial support to Mr.MekchoneBounthanh to study at King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, THAILAND.

REFERENCES

[1] J. Han, K. N. Ngan, M. Li, and H. J. Zhang, "A memory learning framework for effective image retrieval, " *IEEE Trans. Image Process.*, vol. 14,no. 4, pp. 511-524, Apr, 2005.

[2] M. Antonelli, S. G. Dellepiane, and M. Goccia, " Design and implementation of Web-based systems for image segmentation and CBIR," *IEEE Trans. Instrum. Meas*, vol. 55, no. 6, pp. 1869–1877, Dec. 2006.

[3] W.M. Smeulders, M. Worring, A. Gupta, and R. Jain. "Content-based image retrieval at the end of the early years" *IEEE Trans. Pattern analysis*, vol. 22, no. 12, Dec.2000.

[4] T.Kato. "Database architecture for content-based image retrieval" *Proc. SPIE. Image Storage and Retrieval systems.* vol. 1662, USA, Feb 1992

[5] A. del Bimbo. *Visual Information Retrieval*. Morgan Kaufmann Publishers, San Francisco, CA, USA, 1999.

[6] R. Sudhir, Lt.Dr. S. Santhosh. " An efficeint CBIR technique with YUV color space and texture features" *ISSN Computer Eng and Intelligent Systems* vol. 2, no. 6, 2011.

[7] K. Kim, S. Jeong, *et al.* "Efficient video image retrieval by using local co-occurrence matrix texture feature and normalized correlation" *IEEE Tencon.* vol. 2, pp. 934-937, 1999.

[8] T. Dacheng, L. Xuelong, Y. Yuan, *et al.* "A set of novel textural features based on 3D co-occurrence matrix for content-based image retrieval" *ISIF.* vol. 2, pp. 1403-1407, 2002.

[9] J. Zhang, G. Li, H. Seok-wun. " Texture-based image retrieval by edge detection matching GLCM" *IEEE International Conference.* pp. 782-786, Oct 2008.

[10] Z. Yanhong, W. Beizhan, *et al.* "Image retrieval based on improved MCD and EHD of MPEG-7" *ICTM.* vol. 1, pp. 127-132, Dec 2009.

[11] T.-C. Lu and C.-C. Chang, "Color image retrieval technique based on color features and image bitmap," *Inf. Process. Manage.*, vol. 43, no. 2, pp. 461–472, Mar. 2007.

[12] A. Vadivel, S. Sural, and A. K. Majumdar, "An integrated color and intensity co-occurrence matrix," *Pattern Recognit. Lett.*, vol. 28, no. 8, pp. 974–983, June 2007.

[13] Guoping Qiu, "Color image indexing using BTC" *IEEE Trans. Image process.*, vol. 12, No. 1, Jan 2003.

[14] C. Chin Lai, Y. Chuan Chen, " A user-oriented image retrieval system based on interactive genetic algorithm," *IEEE Trans. Measurement.*, vol. 60, No. 10, Oct 2011.

[15] E. J. Delp and O. R. Mitchell, "Image coding using block truncation coding," *IEEE Trans. Commun.*, vol COM-27, No. 9, pp. 1335-1342, Sep 1979.

[16] M. Haralick, K. Shanmugam, H. Dinstein. "Textural features for image classification" *IEEE Trans.Systems*, vol. SMC-3, no. 6, pp. 610-621, Nov 1973.

[17] T. Sikora. "The MPEG-7 visual standard for content description an overview" *IEEE Trans. Systems for video technology*,vol. 11, no. 6, pp. 696-702, June 2001.

[18] K. Kaabneh, A. Abdullah, Z. Al-Halalemah. "Video classification using normalized information distance" *IEEE Computer*, pp. 34-40, Jul 2006.

[19] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantic sensitive integrated matching for picture libraries," *J. Inf. Tchnol. Appl.*, vol. 1, No. 3, pp. 205-214, Dec, 2006.

[20] M. Bounthanh, B. Attachoo, *et al.* "Content-based image retrieval system based on color and image bitmap features" *ICEAST. Inf. Engineering, Applied Science and Technology*, Nov 2012.

Content-Based Image Retrieval System Based on Combined and Weighted Multi-Features

Machine Bounthanh
International College
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, THAILAND
mekchone_b@fe-nuol.edu.la

Kazuhiko Hamamoto
Department of Information Media Technology,
School of Information and Telecommunication Engineering,
Tokai University, Tokyo 108-8619, JAPAN
hama@keyaki.ce.u-tokai.ac.jp

Boonwat Attachoo
Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, THAILAND
boonwat@kmitl.ac.th

Tha Bounthanh
Faculty of Engineering
National University of Laos
Vientiane 3166, LAOS
tha@fe-nuol.edu.la

***Abstract*—This paper, we proposed a novel framework for combining and weighting all of three i.e. color, shape and texture features to achieve higher retrieval efficiency. The color feature is extracted by quantifying the YUV color space and the color attributes like the mean value, the standard deviation, and the image bitmap of YUV color space is represented. The texture features are obtained by the entropy based on the gray level co-occurrence matrix and the edge histogram descriptor of an image. The shape feature descriptor is derived from Fourier descriptors (FDs) and the FDs derived from different signatures. When computing the similarity between the query image and target image in the database, normalization information distance is also used for adjusting distance values into the same level. And then the linear combination has used to combine the normalized distance of the color, shape and texture features to obtain the similarity as the indexing of image. Furthermore, an experimental results indicated, a weight variation to achieve higher retrieval efficiency and the proposed technique indeed outperforms other schemes in terms of the accuracy and efficiency.**

Keywords—Content-based image retrieval (CBIR); Image bitmap; Color distribution; YUV; Gray level co-occurrence matrix; Fourier descriptors.

I. INTRODUCTION

Owing to the rapid development of digital and information technologies, more multimedia information is generated and available in digital form from varieties of sources around the world. And the results to the traditional text-based retrievals based on keywords are not sufficient enough to resolve image retrieval.The content-based image retrieval (CBIR) can greatly enhance the accuracy and efficiency of retrieving and managing the huge number of image data. So, in the recent year, the CBIR system has been an active research area in image processing and attracted more and more attention of researchers.

The content based image retrieval has many application areas such as architectural design, education, commerce, military, medical diagnosis, biomedicine and web image classification. All most of The CBIR research is finding technique to measure the performance for retrieving more similar image from the image databases of retrieval scheme [1, 2, 3]. Some of them have been developed algorithms exploit the low-level features of the image such as color, texture, shape, and spatial descriptors of an object to help retrieve images. In this paper, we have been combined three features including color, shape and texture features and weight variation to achieve higher retrieval efficiency of retrieval scheme. In addition, we have been comparing the accuracy and efficiency of the scheme with other the retrieval scheme. In the next section, we provide a review of related works about CBIR. Then, the proposed method to measure the similarity of images is presented in section III. Section IV shows the experimental results and provides comparative performances. Finally, section V is a conclusion.

II. RELATED WORKS

This section introduces some important literatures review which content based image retrieval (CBIR) was first introduced by Kato in 1992 [4]. One of the most important issues in a content based image retrieval is the feature extraction process, where the visual content of the image is mapped into a new space, the feature space. The term of CBIR is widely used for retrieving desired images from a large collection, which is based on extracting features such as color, texture, edge and shapes from images themselves. For the color feature extraction, the YUV color space is widely used for CBIR scheme. In [5, 6] the YUV color space were extracted a characteristic of images by discrete wavelet transform (DWT) and compares the most significant wavelet coefficients of the YUV components of the query image and those of the images in the database to integrate similarity scores result from the matching processes. Some of experiment result of them were compares the performance for retrieving image of the YUV color space higher more than the RGB and HSV color space when has used the same work. In addition of the color space,

AUN/SEED-Net, JICA

texture and shape are also important features for retrieving image of the retrieval scheme. In [7], a texture was used to retrieve images from database, which is considered six different texture features produced from the co-occurrence matrix and used normalized correlation to adopt as a similarity function. In [8, 9], a texture also was used to retrieve image from database, but in these cases has found a novel technique of texture applied to the CBIR field. Some of them had a combined texture feature with edge histogram to get better the performance in the retrieving image of the retrieval scheme. Shape is one of most important low level image features due to that shape is a very important feature to human perception. Various shape representation methods, or shape descriptors, exist in the literature, these methods can be classified into two categories: region based versus contour based. Contour based shape representation only exploits shape boundary information, these are methods based Fourier descriptors to represent a shape descriptor [10]. The shape classification methods based on autoregressive modeling and Fourier descriptors of closed contours is carried out [11]. "Wang Beizhen [12]" have improved dominant color histogram and edge histogram and evaluates the retrieving result with the criteria provided by MPEG-7. Although, the visual feature is important in the retrieving scheme, but the feature extraction and the method to extract the characteristic is also helping to improve the performance of the retrieval scheme. So, in [13], a CBIR scheme based on the global and local color distribution and in an image is presented. The scheme had used their method with RGB color space to query image and saved a total memory space. "Vadivel [14]" have introduced an integrated approach for capturing spatial variation of both color and intensity levels and shows its usefulness in image retrieval application. In addition, a color space features has been used conjunction with many methods for relevant image query from the large database image. In [15], the Block Truncation Coding (BTC) represented a technique to retrieve images from databases. The result showed that BTC could not be used for color image compression only, BTC also could be used for content base image retrieval.

III. The Proposed Method

Most of the important issues in the image retrieval system is the choice of appropriate the feature extraction method and corresponding similarity measure. In This section, we propose the feature extraction approach of the low-level visual features.

A. The color distribution feature

In the proposed approach, the RGB images are first transformed to the YUV color space for two reasons: 1) efficiency and 2) ease of extracting the feature based on the color tones. It is well-know that the Y component of an image is much more significant than the U and V component can be much less than those kept for the Y component without sacrificing the accuracy rate. Moreover, compared with the features derived from the RGB color space, the features derived from the YUV color space are more suitable for the purpose of screening the image based on a certain color tone by weighting the corresponding features. So, we had choices the YUV color space apply to this research and it can be defined in [5, 6].

To obtain the color features, The mean value and the standard deviation were used to extract the global feature from an image that depicts the global characteristics of the images. And the image bitmap as the local to describe the local characteristics of the image. There are can be defined in [16, 17].

B. Texture descriptor feature

Texture is also a visual feature that refers to innate surface properties of an object and their relationship to the surrounding environment. In this case, we are using gray-level co-occurrence matrix (GLCM) and edge histogram for extracting texture features of images [18], [19], which is a simple and effective method for representing texture. Co-occurrence matrices display the gray level spatial dependency along different angular relationships, horizontal, vertical, and two diagonal directions, in a block or an image. The GLCM is specified by the relative frequencies $P(i,j,d,\theta)$ with which two pixels, separated by a distance d, occur in a texture along the direction of angle , one with gray level I and the other with gray level j, The GLCM is mathematically defined as follows:

$$P(i,j,d,\theta) = \#\{(x_1,y_1)(x_2,y_2)|g(x_1,y_1) = i, g(x_2,y_2) = j, |(x_1,y_1) - (x_2,y_2)| = d, \angle((x_1,y_1)(x_2,y_2)) = \theta\} \quad (1)$$

Where # denotes the number of occurrences inside of the window, with i and j being the intensity levels of the first pixel and the second pixel at positions (x_1,y_1) and (x_2,y_2), respectively.

In order to simplify and reduce the computational effort, we computed the GLCM according to one direction (i.e., = 0°) with a given distance d (= 1) and only select one of the features as retrieval feature: Entropy. The entropy (E) measures the disorder of an image and it achieves its largest value when all elements in P matrix and are defined as follows:

$$E = -\sum_{i,j} P_{i,j} \log P_{i,j} \quad (2)$$

Edge in an image is an important low-level feature, it can describe both shape and texture features, which are also essential elements for content-based image retrieval. One way to represent such an important edge that is used a histogram. The edge histogram in the image space represents the frequency and the directionality of the brightness change in the image. The edge histogram descriptor (EHD) [20] represents the local edge distribution in the image, it's useful in image matching when the texture is not homogeneous.

C. Shape descriptor features

In general, Fourier descriptors (FDs) are obtained by applying Fourier transform on a shape signature, the normalized Fourier transformed coefficients are called the

Fourier descriptors of the shape. The shape signature is any one-dimensional function representing two-dimensional areas. Tree shape signatures are considered in our case, these are centroid distance, complex coordinates, and curvature signature which is derived from shape boundary coordinates. In the following, we assume the shape boundary coordinates $(x(t), y(t)), t = 0, 1, \ldots, L-1$, have been extracted in the preprocessing stage [10], [11].

A complex coordinates function is simply the complex number generated from the boundary coordinate:

$$z(t) = x(t) + iy(t) \tag{3}$$

In order to eliminate the effect of bias, we use the shifted coordinate function:

$$z(t) = [x(t) - x_c] + i[y(t) - y_c] \tag{4}$$

Where (x_c, y_c) is the centroid of the shape, which is the average of the boundary coordinates

$$x_c = \frac{1}{L}\sum_{t=0}^{L-1} x(t), y_c = \sum_{t=0}^{L-1} y(t) \tag{5}$$

This shift makes the shape representation invariant to translation.

The centroid distance function is expressed by the distance of the boundary points from the centroid of the shape

$$r(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{1/2} \tag{6}$$

Due to the subtraction of centroid, which represents the position of the shape, from boundary coordinates, the centroid distance representation is also invariant to translation.

A curvature function is a function of angular changes of a boundary tangent. The curvature at a boundary can be defined as the differentiation of successive boundary angles calculated in window w:

$$K(t) = \theta(t) - \theta(t-1) \tag{7}$$

Where

$$\theta(t) = \tan^{-1}\frac{y(t) - y(t-w)}{x(t) - x(t-w)} \tag{8}$$

The Fourier transformation on shape signatures is widely used for shape analysis. In this case, we described the shape signatures by the discrete Fourier transform (DFT), let's *s(t), t=0, 1, ..., L,* is a sampling stage. The discrete Fourier transform of *s(t)* is given by

$$u_n = \frac{1}{N}\sum_{t=0}^{N-1} s(t) exp\left(\frac{-j2\pi nt}{N}\right), n = 0, 1, \ldots, N-1 \tag{9}$$

The coefficients $u_n, n = 0, 1, \ldots, N-1$, are usually called Fourier descriptors (FD) of the shape, denoted as $FD_n, n = 0, 1, \ldots, N-1$.

Before applying Fourier transform on the shape signature, shape is first sampled by equal arc length sampling to fix the number of points [21]. The equal arc length sampling method selects candidate points spaced at equal arc length along the shape boundary. The space between two consecutive candidate points is given by *P/K* , where *P* is the perimeter of the shape boundary and *K* is the total number of candidate points. For each shape, we select $N = 2^8$ candidate points with equal arc length space between them. After that, the DC component was used to describe each Fourier descriptors to represent shape invariant [22].

D. The similarity measure of the features

In the experimental system, we use linear combinations of three feature distance measurements to evaluate the similarity which is the distance of color features, the distance of texture feature, and the distance of the shape descriptor feature. When computing the similarity of each feature between the query image and the target image in the database. The normalized information distance (NID) [23] was used to normalize the value of each distance adjusted to the same level: the distance has a value between 0 and 1. For the distance of the color features have described in the term of global features and the local features. The local feature is used the Hamming distance to evaluate the distance between the two binary image bitmaps. On the other hand, the global features, the mean value and standard deviation were used the Euclidean distance to calculate the similarity, the Euclidean distance between the entropy and edge histogram to represent the distance of texture feature [24], and the distance of shape feature derived from the normalized information distance of the Fourier descriptors.

$$d_{Shape}(d, q) = \frac{fd(d, q) - min\{fd(d, q)\}}{max\{fd(d, q)\}} \tag{10}$$

Where fd represent the shape similarity between the image in the database d and the query image q respectively. The similarity measurement is obtained by linearly combining and weight these of all similarity values.

$$SIM(d, q) = w_c \times d_{color} + w_t \times d_{texture} + w_s \times d_{shape} \tag{11}$$

IV. Experimental Result

In this experiment, we used the database of the SIMPLIcity project covering a wide range of semantic categories of natural scenes with artificial objects for experiments [25]. The database is partitioned into ten categories, including African people and village, beach, building, buses, dinosaurs, elephants, flowers, horses, mountains and glaciers, food, etc., and each category contain 100 images (Fig. 1). All the images are in RGB color space. They are in the JPEG format with size in 256×384 and 384×256 pixels.

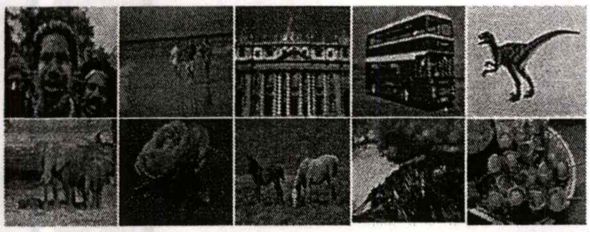

Fig. 1. Sample images of each category of the image database.

A. *The retrieval accuracy and efficiency*

To evaluate The performance of proposed CBIR methods, we examined by using the evaluated point value of the precision rate and recall rate for considering the number of relevant images to query the retrieval [26]. The precision gives information about the relative effectiveness of the retrieval scheme and the recall giving information about the absolute accuracy of the retrieval scheme

In each experiment, evaluation of the retrieval precision is performed ten category images were randomly selected from each specific category in the database as query images and count the precision and recall separately after getting the return result. The responding of image, we always used the top 20 retrieved images to compute the precision and recall. For the efficiency and accuracy of our retrieval scheme is shown on the Fig. 2, which is weight variation depends on the best weight of the color feature. In this case, we try given weight for the experimental to get the best weight. It is clear that the efficiency and accuracy of the retrieval scheme depended on the weight of the color feature, the best weight of the color feature is 70% . After that, the weight variation of the texture and shape features were varied seven times to achieve higher performance of the scheme.

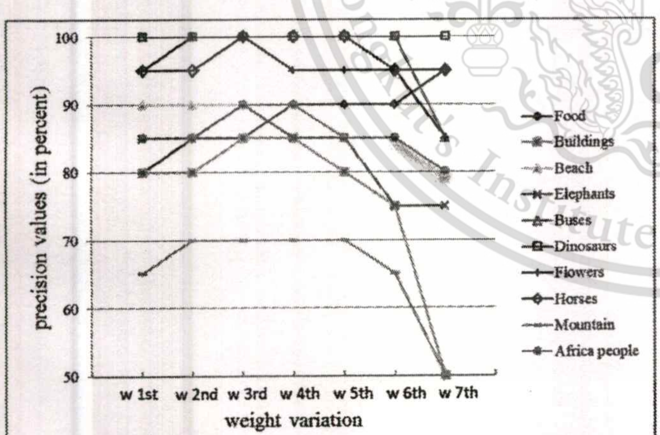

Fig. 2. Performance precision of retrieving images in the each image category.

Fig.2 gives crossover points of precision value of CBIR based on our method, it depends on weight variation. The weight of the color feature distance was constant at 70% and varied the weight of the texture and shape feature distances from 30% to 0% and 0% to 30% respectively. Based on that way, the higher performance was achieved when we set $w_c = 70\%, w_t = 20\%$ and $w_s = 10\%$, there are the weight of the color, texture and shape feature distances. The weight factor of color feature distance is higher than the weight factor of texture feature and shape feature distance because our database consists of most natural image.

B. *Comparison with other method*

To evaluate the superiority of our approach, we had compared our experimental result with other retrieval systems. The first system is a user-oriented image retrieval system based on an interactive genetic algorithm [16], which the color attributes, the entropy based on the gray level co-occurrence matrix and the edge histogram of an image are combined to represent the similarity measure and applied the interactive genetic algorithm to achieve better retrieval efficiency of the retrieval scheme. Second one is effective content-based image retrieval: combination of quantized histogram texture feature in the DCT domain [27]. The statistical texture features are extracted from the quantized histogram in the DCT domain using only the DC and first tree AC coefficients of the DCT blocks of an image having more significant information were proposed. And the third one is content based image retrieval using color, texture and shape features [26], which the local color, the texture descriptors in a coarse segmentation framework of grids, and the shape descriptors in terms invariant moments computed on the edge image were used to achieve the performance of the scheme. The comparison is considered the performance during the term of the precision and recall, there are shown in TABLE I and TABLE II respectively. Actually, the propose approach of our scheme has achieved better than the other systems. Exceptionally, in the case of the "beach" and "food" categories, the performance of our approach is slightly inferior to that obtained in [16]. This is because their feature sets derived from the interactive genetic algorithm to achieve better performance, which lets the user evaluate the retrieved images as more or less relevant to the query one. For our approach also used the same low-level visual features, but some of the features, we used different technique to extract feature and the YUV color space is suitable for image retrieval more than the other color space in the same work (certify in [5, 6, 20]). So, our proposed approach scheme had a higher performance more than the other systems illustrated.

TABLE I. AVERAGE OF PRECISION VALUE (IN PERCENTAGE)

Category	P [16]	P [26]	P [27]	Ours
Africa people	61	48	92	90
Beach	93	34	86	90
Buildings	85	36	68	85
Buses	71	61	89	100
Dinosaurs	100	95	100	100
Elephants	80	48	76	90
Flowers	82	65	100	100
Horses	79	70	93	100
Mountain	56	40	47	70
Food	99	50	66	85
Average	80.6	54.7	81.7	91.0

TABLE II. AVERAGE OF RECALL VALUE

Category	P [16]	P [26]	P [27]	Ours
Africa people	12	10	18	18
Beach	18	6	17	18
Buildings	17	8	16	19
Buses	14	12	18	20
Dinosaurs	20	19	20	20
Elephants	16	10	15	18
Flowers	16	13	20	20
Horses	15	14	18	20
Mountain	11	8	9	14
Food	19	10	13	17
Average	15.8	11.0	16.4	18.4

V. CONCLUSION

This paper a content-based image retrieval system is proposed in a novel framework for combining and weighting all of three features to retrieve images with a Wang image database. The color features are extracted from the global characteristic of an image such as the mean value and the standard deviation, and the image bitmap represented the local feature, there are obtained by quantifying the YUV color space. The texture features are extracted from the entropy based on the gray level co-occurrence matrix and the edge histogram. And the Fourier descriptors derived from difference signatures are represented shape descriptors. There are also normalized by the normalized information distance. The linear combination was used to combine all of three features to obtain the similarity measurement of our retrieval scheme. The experimental results of the proposed approach have shown weight variation to achieve higher retrieval efficiency: the efficiency can be achieved in 91% by weight $w_c = 70\%$, $w_t = 20\%$ and $w_s = 10\%$. In addition, we had compared the efficiency and accuracy with other the retrieval scheme, which the our approach had better than the performance of other systems.

ACKNOWLEDGMENT

The authors would like to thank the AUN/SEED-Net for the financial support to Mr. Mekchone Bounthanh to study at King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, THAILAND.

REFERENCES

[1] J. Han, K. N. Ngan, M. Li, and H. J. Zhang, "A memory learning framework for effective image retrieval, " IEEE Trans. Image Process, vol. 14, no. 4, pp. 511-524, Apr 2005.

[2] M. Antonelli, S. G. Dellepiane, and M. Goccia, " Design and implementation of Web-based systems for image segmentation and CBIR," IEEE Trans. Instrum. Meas, vol. 55, no. 6, pp. 1869–1877, Dec. 2006.

[3] W.M. Smeulders, M. Worring, A. Gupta, and R. Jain. "Content-based image retrieval at the end of the early years,"IEEE Trans. Pattern analysis, vol. 22, no. 12, Dec 2000.

[4] T.Kato. "Database architecture for content-based image retrieval" Proc. SPIE. Image Storage and Retrieval systems, vol. 1662, USA, Feb 1992

[5] A. del Bimbo. Visual Information Retrieval. Morgan Kaufmann Publishers, San Francisco, CA, USA, 1999.

[6] R. Sudhir, Lt.Dr. S. Santhosh. " An efficeint CBIR technique with YUV color space and texture features," ISSN Computer Eng and Intelligent Systems, vol. 2, no. 6, 2011.

[7] K. Kim, S. Jeong, et al. "Efficient video image retrieval by using local co-occurrence matrix texture feature and normalized correlation," IEEE Tencon. vol. 2, pp. 934-937, 1999.

[8] T. Dacheng, L. Xuelong, Y. Yuan, et al. "A set of novel textural features based on 3D co-occurrence matrix for content-based image retrieval," ISIF. vol. 2, pp. 1403-1407, 2002.

[9] J. Zhang, G. Li, H. Seok-wun. " Texture-based image retrieval by edge detection matching GLCM," IEEE International Conference. pp. 782-786, Oct 2008.

[10] D.S. Zhang , G Lu. "A comparative study of Fourier descriptors for shape representation and retrieval," ACCV02. Asian Conference on Computer Vision, pp. 646-651, Australia, Jan 2002.

[11] H. Kauppinen, T. Seppanen. "An experiment comparison of autoregressive and Fourier-based descriptors in 2D shape classification," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 17, pp. 201-207, 1995.

[12] Z. Yanhong, W. Beizhan, et al. "Image retrieval based on improved MCD and EHD of MPEG-7" ICTM. vol. 1, pp. 127-132, Dec 2009.

[13] T.-C. Lu and C.-C. Chang, "Color image retrieval technique based on color features and image bitmap," Inf. Process. Manage., vol. 43, no. 2, pp. 461–472, Mar 2007.

[14] A. Vadivel, S. Sural, and A. K. Majumdar, "An integrated color and intensity co-occurrence matrix," Pattern Recognit. Lett, vol. 28, no. 8, pp. 974–983, June 2007.

[15] Guoping Qiu, "Color image indexing using BTC"IEEE Trans. Image process, vol. 12, No. 1, Jan 2003.

[16] C. Chin Lai, Y. Chuan Chen, " A user-oriented image retrieval system based on interactive genetic algorithm," IEEE Trans. Measurement, vol. 60, No. 10, Oct 2011.

[17] E. J. Delp and O. R. Mitchell, "Image coding using block truncation coding," IEEE Trans. Commun, vol COM-27, No. 9, pp. 1335-1342, Sep 1979.

[18] M. Haralick, K. Shanmugam, H. Dinstein. "Textural features for image classification" IEEE Trans.Systems, vol. SMC-3, no. 6, pp. 610-621, Nov 1973.

[19] J.R. Carr, F. Pellon. "The semivariogram in comparison to the Co-Occurrence matrix for classification of image texture," IEEE Trans. Geoscience and Remote Sensing, vol. 36, no. 6, Nov 1998.

[20] T. Sikora. "The MPEG-7 visual standard for content description an overview" IEEE Trans. Systems for video technology, vol. 11, no. 6, pp. 696-702, June 2001.

[21] J. Peter, V. Otterloo. "A contour-oriented approach to shape analysis," Prentice Hall International (UK) Ltd. C1991.

[22] D. Zhang, G. Lu. "A comparative study Content-based shape retrieval using different shape descriptors: A comparative study," IEEE International Conference, pp. 1139-1142, Aug 2001.

[23] K. Kaabneh, A. Abdullah, Z. Al-Halalemah. "Video classification using normalized information distance" IEEE Computer, pp. 34-40, Jul 2006.

[24] M. Bounthanh, B. Attachoo. "Content based image retrieval system based on combination multi-features" ICESIT. International Conference on Embedded system and Intelligent Technology, pp. 132-136, Jan 2013.

[25] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantic sensitive integrated matching for picture libraries," J. Inf. Tchnol. Appl, vol. 1, No. 3, pp. 205-214, Dec 2006.

[26] M. Bounthanh, B. Attachoo, et al. "Content-based image retrieval system based on color and image bitmap features," ICEAST. Inf. Engineering, Applied Science and Technology, Nov 2012.

[27] F. Malik, B. Baharudin "Effective content-based image retrieval: combination of quantized histogram features in the DCT domain," ICCIS. International Conference on Computer on Information Science, vol. 1, pp. 425-430, June 2012.