CONTENT-BASED IMAGE CLASSIFICATION AND RETRIEVAL USING SUPPORT VECTOR MACHINE

CHU THINZAR

M.C.Sc. **MARCH 2019**

CONTENT-BASED IMAGE CLASSIFICATION AND RETRIEVAL USING SUPPORT VECTOR MACHINE

By

Chu Thinzar
B.C.Sc.

A Dissertation Submitted in Partial Fulfilment of the Requirements for the Degree of

Master of Computer Science
(M.C.Sc.)

University of Computer Studies, Yangon
March 2019

ACKNOWLDGEMENTS

First of all, I would like to express my deepest gratitude and sincere appreciation to the following persons who have contributed directly or indirectly towards the completion of this thesis and helped me make this dissertation possible.

I would like to express my respectful thanks to **Prof. Dr. Mie Mie Thet Thwin**, Rector of the University of Computer Studies, Yangon, for her kind permission to submit this thesis.

My sincere thanks and regards to **Dr. Thi Thi Soe Nyunt**, Professor and Head of Faculty of Computer Science, the University of Computer Studies, Yangon, for her administrative supports and encouragement throughout the completion of this thesis.

I would like to express my sincere gratitude to my supervisor, **Dr. Hlaing Htake Khaung Tin**, Associate Professor, Faculty of Information Science, the University of Computer Studies, Yangon, for her valuable guidance, encouragement and helpful suggestions throughout the progress of the research and in the preparation of this thesis.

My gratitude also goes to **Daw Sandar**, Lecturer, Language Department of the University of Computer Studies, Yangon, for her editing this thesis from the language point of view.

I especially give thanks to my parents, all of my colleagues, and friends, for their encouragement and help in writing this thesis.

Finally, I would like to express special thanks to all of the teachers, staffs from the University of Computer Studies, Yangon, for their support.

STATEMENT OF ORIGINALITY

I hereby certify that the work embodies in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

---------------------------------- -------------------------------

Date Chu Thinzar

ABSTRACT

Image search is more efficient for managing a wide range of image databases. Content-based image retrieval (CBIR) is one of the image retrieval techniques in which users use the visual characteristics of images such as color, shape and texture, etc. It permits the end user to give a query image in order to retrieve the image stored in the database based on the similarity to the query image. The system extracts the features of the query image, searches the database for images with similar features, and exhibits relevant images to the user in order of similarity to the query. Many CBIR systems have been developed to compare, analyze, and search images based on one or more of these features. This system is implemented as an image retrieval system combining visual content features and a support vector machine (SVM) classification.

First, the system extracts the features of images from dataset with color auto-correlogram, color moment and gabor wavelet for the training phrase. When the user input query image, the system extracts features with these feature extraction methods in the testing phrase. And then, the system applies support vector machine (SVM) classifier to classify the image. After that, the system compares feature vectors between the query image and image dataset. Finally, the system retrieves the relevant image with query image. The applied system uses Wang dataset for the purpose of training and testing the system. And other 100 images that are not from dataset is also used for testing system. The overall accuracy of the system is over 80% for all classes. The system is implemented with MATLAB programming language on window platform.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1

INTRODUCTION

1.1 Introduction to Image Retrieval System

Number of digital images have increased dramatically due to the use of services of multimedia, digital cameras and the fact that the facade has become increasingly popular on the Internet. In many areas, the use of digital images is increasing. Given this situation, it is necessary to classify images into significant categories and objects such as tree, dog, humans, and planes to manage and organize the images in the database.

Digital images are used in multiple applications, e.g. Geography, medicine, architecture, advertising, military design and album. However, it is difficult for users to find and manage the maximum number of images in the database. In general, image retrieval system is divided into two methods such as Text-Based Image Retrieval (TBIR) and Content-Based Image Retrieval (CBIR).

The first method is text-based image retrieval system. This is easy to implement and the conventional database query technique where textual metadata to each image is added to retrieve image by keywords. This requires a manual annotation of the database images and consumes huge time and also a complex one. Moreover, the annotation process is not an effective format since users generally are not interested to make it in systematic way. This is because there are diverse users who tend to use different words to describe a same image characteristic and also more than one object can be referred by the same word, inefficiency. Hence, this method loses its popularity and performance due to lack of systematization.

The second one is content-based image retrieval system. Text-based image retrieval has some disadvantages, such as incompetence, data loss, time-consuming processing, and more complex tasks. This problem is solved using content-based image retrieval. “Content based” means searching and analyzing content of images that are not metadata [3]. Metadata is a keyword, tag, or description related to an image. The term "content" refers to color, shape, texture or other information that can be obtained from an image. CBIR is desirable because most web search engines use purely metadata information and result in large amounts of garbage. Also having humans manually enter image keywords in a large database, it may be inefficient, costly, and may not capture every keywords that describes the image. As a result, systems capable of filtering images based on content will allow better indexing and more accurate results.

There is a growing interest in CBIR because of the limitations inherent in metadata-based systems as well as the large range of possible uses for efficient image retrieval. The capability of present CBIR systems has been limited by their use of only primitive features, so they cannot satisfy most semantic-level query demands. Primitive features denote some general visual characteristic including color, shape, texture, and spatial relationships among objects, and these features can be used in most CBIR applications. The color feature which is widely used in CBIR systems.

The CBIR system extracts the stored images from the database by comparing the features of the query image with the images in the collection. First, it will extract and store the features of the query image, then it will transmit all the images in the database and extract the features of each image. The result of the system is the image that the features are most similar to the image query.

There are many ways to extract images with different features such as colors, shapes and textures from the content. In general, colors, textures and shapes are used to extract similar images from an image database. Some CBIR techniques use different techniques. The authors in [1, 3] chose two visual features such as textures and colors. Texture analysis is generally a time consuming process. The texture and color features are extracted by wavelet transformation and color histogram which resist resizing and translation of objects into images in [4, 7].

Shape is also a discrete feature of the image retrieval system. The shape of the image must be extracted from the image according to the segmentation and classification of the forms in [4, 6]. Authors in [7] used three types of features, such as colors, shapes and textures for image retrieval, to enhance image retrieval. The texture analysis algorithm uses the gray level co-occurrence matrices (GLCM) to account for orientation and multiple scale in [2]. The authors in [8, 13] presented the retrieval of the image such as color, texture, and histogram are used, called Wavelet Based Color Histogram (WBCHIR) retrieval. The texture and color features are extracted by wavelet and color histogram combination of these features is robust to scale and translation of object in the image.

CBIR is an important and challenging research topic. The system is used three types of image feature such as color auto-correlogram, color moments, gabor wavelet. And then the system introduces a technique to retrieve images by classifying it on the basis of the features and characteristics it contains using Support Vector Machine (SVM). The dataset of images is created which is used for feature matching purpose by SVM to find similar images from the database and based on user requirements images are retrieved.

1.2 Objectives of the Thesis

The objectives of the thesis are as follow:

- to understand the image processing technique and the content-based image retrieval
- to learn about the integration of the visual content features of image such as color auto-correlogram, color moments, gabor wavelet and SVM classifier
- to access the stored images in an efficient way
- to implement content-based image retrieval system to get more relevant results

1.3 Motivation of the System

The problem of finding images is difficult to deal with the image than text. Text documents are "one dimension" (an array of words), while digital images are "two dimensions". The size of the image data is much larger than the text. The most important difference between the text and the image has to be processed and interpreted in order to extract the sense of perception, which is the work to be done in computer vision and image understanding. There are some problems (i) because the image data contains very rich data, it is very difficult to capture the content of the image using only a few key words, not to mention the tedious work associated with the manual annotation process and (ii) manual annotation process is rather subjective, ambiguous and incomplete. If the query refers to the content of the image that is not annotated at the beginning, or if the user uses different words to describe the same image content, the text retrieval system will fail.

Textual features about images could be easily retrieved using existing technology. However, it requires humans explain each image of the database manually. This is not practical for large databases or auto-generated images, such as CCTV. It is also possible to get rid of images with different synonyms in the description. Systems based on categorizing images in semantic classes like "cat" as a subclass of "animal" avoid this problem but still face the same scaling.

To overcome the problem of text-based approach, the content-based image retrieval (CBIR) approach uses the visual appearance features of image for retrieving system. CBIR system protects the disadvantages of the traditional way of retrieving (text-based or description-based) approaches such as, time consuming and expensive to retrieve the similar image from image database. To produce more accurate and efficient compared result, this system combines the visual features of images and Support Vector Machine (SVM) classifier.

1.4 Organization of the Thesis

The organization of the thesis is as follows:

Chapter 1 includes the introduction to the system, objectives of the system, motivation, and organization of the thesis.

Chapter 2 describes the basic concept of digital image processing, the features of image and how to extract image features for the image retrieval system.

Chapter 3 discusses about the color auto-correlogram, color moments, gabor wavelet features and the types of classification techniques (Classifiers).

Chapter 4 presents the implementation of the system describing system design, process flows, and experimental results of the system.

The overview of the proposed system, limitations and further extensions are presented in Chapter 5.

CHAPTER 2
BACKGORUND THEORY

2.1 Digital Image Processing

An image can be defined as a two-dimensional function, where x and y are the spatial (the plane) the coordinates and the amplitude of *f* where any pair of coordinates (x, y) is called intensity or gray level of the image at this point, when x, y and the intensity of f are all discrete numbers, we call them digital images. Digital image processing means processing digital image with digital computers. Remember that digital images consist of a limited number of elements, each with a unique position and value. These elements are called image elements, picture elements and pixels. Pixels are the term widely used to represent digital picture elements.

Visibility is the most advanced sensation, so it's not surprising that the image play the most important role in human perception. However, unlike humans, who are only limited to the electromagnetic spectrum (EM), the spectrum of images covers almost all EM spectrum, from gamma to radio. They can work on images generated by sources that humans are not accustomed to associate with images. Linked to these images, including ultrasound, electron microscopy and computer generated images. Thus, digital image processing covers a wide range of applications.

Image processing uses techniques to modify or interpret existing images, such as photos and TV scans. Two principle applications of image processing are improve the quality of the image and machine perception of visual information.

When using image processing methods, the system will first convert the photo or other image into an image file. Then, use digital methods to rearrange parts of the image, to increase color separation, or improve shading quality. These techniques are widely used in commercial art in relation to the decoration and arrangement of parts of photographs and other works of art. This similar method is used to analyze satellite images of the world and galactic photographs. Medical applications also use extended image processing techniques to improve tomography and simulation images. Tomography is an X-ray imaging technique that allows you to see the cross section of the physiological system. Image processing and computer vision are often combined in many applications.

The ultimate goal of computer vision is to use computers to mimic human vision, including learning and the ability to make inferences and act upon image capture. This area is a branch of artificial intelligence (AI) that aims to imitate human intelligence. Image analysis area (also known as visual perception) lies between image processing and computer vision.

There is no clear limit to the continuity of image processing from one side to the other of the computer view. However, a useful paradigm is to consider three types of computerized processes in this continuous process: low-, mid-, and high-level process. Low-level processes use traditional features such as advanced image processing to reduce noise, contrast enhancement, and image sharpening. Low-level processes have the fact that both input and output are images. Mid-level image processing involves tasks such as segmentation (division of images into regions or objects), explanation of these objects to reduce the size to the appropriate format for the computer processing and classification of each object. The mid-level process is characterized by the fact that the input is generally an image, but the output is a separate feature of these images (like the edge, shape and identity of each object). Finally, high-level processing involves the "making sense" of groups of accepted objects, as well as in the analysis of images and at the end of the continuity of cognitive functions related to vision.

Thus, digital image processing encompasses processes whose inputs and outputs are images and, in addition, encompasses processes that extract attributes from images, up to and including the recognition of individual objects [13]. The processes of acquiring an image for the area contain the text, preprocessing that image, extracting (segmenting) the individual characters, describing the characters in a form suitable for computer processing, and recognizing those individual characters are called digital image processing.

2.2 Applications of Digital Image Processing

There are many image processing applications in a wide range of human activities from remote rendering of scenes to biomedical image translation.

2.2.1 Automatic Visual Inspection System

The automatic vision control system that is necessary to upgrade the efficiency and value of products in production [6]. In the vision-based automated inspection system, filament binary image layer is made from which the silhouette filaments form is made. Analyze this form to show the non-uniformity of the shape of the filament geometry in the object.

2.2.2 Remotely Sensed Scene Interpretation

The facts on natural resources such as agriculture, hydrology and minerals, forests, geological sources, and others. They can be extracted from remote image analysis. For long-range scenes analysis, the surface image of the earth is detected by sensors in the satellite, which is displayed remotely or transmitted to the ground station by a multi-spectrum scanner placed on the plane. Regional interpretation techniques and satellite imagery for urban planning, resource organization, flood protection, agricultural production monitoring, and so on.

2.2.3 Biomedical Imaging Techniques

Different types of imaging devices, such as X-ray pictures, computerized X-ray images (CT), ultrasound, and so on, they are used for massive medical purposes [9] - [11]. Some biomedical imaging applications include:

(i) **Identification of pulmonary disease:** on chest X-rays, dark air structure, while solid tissue is lighter and bones are brighter than soft tissue. The anatomy structure which is evident in chest x-ray typically consists of ribs, pectoral breastbone, heart and diaphragm that separates the thoracic cavity from the abdomen .These fields on the chest radiography are checked for abnormalities with relevant segments.

(ii) **Identification of heart disease:** quantitative measures such as the size and shape of the heart are significant diagnosis features in the classification of heart disease. You can use image analysis techniques for radiography to improve the diagnosis of heart disease.

(iii) **Digital mammograms:** these are very suitable for identifying properties (for example, small calcification) for the diagnosis of breast tumors Image processing techniques such as increased contrast, segmentation, extraction features, shape analysis, and so on. The uniformity of the tumor shape determines whether the tumor is benign or malignant.

2.2.4 Content-Based Image Retrieval (CBIR)

Image retrieval that is user-defined image from a large image database is a critical image processing technique. The method for large collections of multimedia and digital libraries has created a great need for the development of search engines to index and retrieve data. There are now a lot of good search engines that can extract texts in a format that the device can read. However, there is no quick tool to extract intensity and color images. The traditional way to search and index image is slow and expensive. Therefore, there is an urgent require to progress an algorithm to retrieve embedded content images. The digital image features (shape, texture, color, object structure, etc.) can be used as an index button to find and extract images from a large image database. Image retrieval based on the content of the image is called the content-based image retrieval [12, l3]. Content-based Image Retrieval or CBIR involves two steps:

Feature Extraction: The first step in the feature extraction process is to extract the features of the image at different levels. These features are defined as one or more measurement methods, each specifying certain quantitative properties of the object and being calculated from the quantitative properties of the object. The extraction of features is a special form of dimensionality reduction.

Matching: The second step is to match these properties to achieve similar results.

2.2.5 Moving-object Tracking

Tracing moving objects to measure movement parameters and acquiring images of recorded objects is an important area for image processing. [14, 15] Typically, object tracking has two distinct ways to detect objects:

1. Recognition-based tracking
2. Motion-based tracking

The rapid tracking systems (such as military bomber, missiles, etc.) have been established based on motion-based prediction techniques, for example Kalman filter, Kalman filtration particle, etc. The objective is that enters the cavity of the vision sensor is automatically received without human interruption. While tracking of region-based, objects are recognized in the continuous frame and tracking is performed using their location information.

2.2.6 Image and Video Compression

This method is an active image processing application [13, 15]. The development of image and video compression technology continues to act as an essential part in the advance of communication and multimedia applications. Over the last two decades, while the cost of storage has dropped seriously, the need for video and image storage is increasing rapidly. Digital radiography 36 cm x 44 cm at 70 post-mortem needs approximately 45 MB of storage space. Likewise, the storage requirement for high-definition television with a resolution of 1280 x 720 at 12 frames per second exceeds 1250 megabits per second.

Direct transmission of non-compressed video images from real-time communication channels is now a crucial proposition. Significantly, the stills and video images contain a large bulk of information that is visually exaggerated in the canonical representation. The outrage stays in the fact that adjacent pixels in a smooth and homogeneous region of natural imagery have very small variations in their values that cannot be seen by human observers. Likewise, the next image is somewhat similar to the slow motion video sequence and has combined the surplus over time. Image and video compression techniques essentially reduce this visual redundancy in data presentation to show frames with smaller number of bits, thereby reducing the chaos of communication memory and broadband effectiveness.

2.3 Types of Images

Four types of images are: Intensity images, Binary images, Indexed images, and RGB (Red, Green and Blue) images [3].

2.3.1 Intensity Images

This image is a data matrix whose values are ranged to express intensities. If the elements of an intensity image belong to class uint8 or to class uint16, they have integer values in the length of (0.255) and (0, 65535), for each. If the image has a double class, the values are floating-point numbers. The miniature value, twice the intensity of the class, is in range (0, 1) depending on the pattern.

2.3.2 Binary Images

A binary image is a logical matrix 0s and 1s. An array of 0s and 1s whose data class value is uint8 is not considered a binary image in MATLAB. A digital matrix is converted to binary using logical functions.

2.3.3 Indexed Images

These images consist of two parts: the matrix data of the X integers and the map color matrix map. The matrix map is the m * 3 matrix of double class with a decimal value in between (0, 1). The length of the map corresponds to the specified number of colors. Each map line identifies the red, green and blue part of a single color. Indexed images use "direct mapping" of pixel intensity value and color mapping values. The color of each pixel is determined using the corresponding amount of the matrix. Integer X is a pointer on the map. If X is a class, all components less than or equal to one point in the first row of the map are doubled. It costs 2 points in the second line and so on. If X is a uint8 or uint16 class, all elements with value 0 point to the first set of map. This go to the second row and so on.

2.3.4 RGB Images

The RGB color image is the matrix of the M * N * 3 matrix, where each pixel color is a triplet that matches the red, green, and blue portion of the RGB image at a particular location spatial. The RGB image is seen as a "stack" of three gray images inserted in red, green and blue items on the screen. Creating a color image is called red, green, and blue. The part image data class specifies the set of values. If the RGB image has a double class, the set of values is (0, 1). Also, the set of values (0, 255) or (0, 65535) for the RGB image is uint8 or uint16, respectively. The number of bits used to display the pixel image value of the part refers to the bit depth of the RGB image.

2.4 Types of Image Features

In image processing and pattern recognition, the extraction of features is an important step that constitutes a special form of dimensionality reduction. When the input data is too large to process and is suspected to be duplicated, the data will be changed to a reduced set of features. The process of modifying input data into a set of properties is called feature extraction. Features typically include information related to the characteristics of digital images.

2.4.1 Low-level Image Features

Most of the recommended CBIR methods are based on pre-processing procedures for feature extraction, in order to extract the appropriate image features (descriptors, properties) that contain enough information to extract relevant images from database containing thousands of images [14]. Low-level features depend on color, texture, and shape, which are related easy to calculate and can be relevant to the low-level human perception.

2.4.2 Higher-level Image Features

High level functions are designed to show concepts that have meaning in the image (such as activities appearing in images or objects seen in image) that are of greater interest to humans [14]. The higher level features in images are not extracted easily from pixels.

2.5 Image mining and Content-based Image Retrieval

Many visual information in the form of video and image data is scattered around the world and other irrelevant information such as numbers, numerical data, voice, audio, text, learning etc. Data mining get the important information from large data sets through data embedded models and knowledge to be discovered. Data mining involves the retrieval of beneficial data embedded in large images and videos. Most of the jobs in this area are limited especially for the improvement of content-based image retrieval systems (CBIRs). Retrieving images from a large image database is a vital task in terms of image processing and artificial vision.

Image retrieval with similarities is an elegant method applied developing CBIR. Basically, the CBIR system ought to automatically capture important data about the image for certain applications. The content of the image is a natural image and the interpretation of the information passed through the image is subjective and depended upon the basic human image system. Image data is applied for artificial vision banked on the desired image properties and interpreting these features for specific applications.

Most image retrieval activities are made up of searches and retrieval of images based on image equivalency analysis or features of the image database. The image retrieval system could be divided into two categories by the search form.

For first group, image is explained based on user-defined text [4, 7]. The image is indexed and extracted according to the basic description; size, type, date and time of acquisition, owner's identity information, keyword or textual description. For this reason, it is often referred to as image retrieval depend on the description or text. The image index is predefined based on descriptions and searches for these indices when the query "Search Images within the database is stored according to the specified set of descriptions".

The text-based description is usually manually printed for each image by human operators, as creating automatic keywords for this image is difficult without having to combine visual information and extracting features. For this case, a process of producing energy that is not practical at the age of multimedia information. Because the image description is very subjective, the automotive process of creating a description of the text used to index the image can also be false and incomplete.

For second group, the query can point in following: "Search image is similar to the image of the specified query image". The second group of an image retrieval process with similarity is CBIR [2-6]. In the CBIR system, searches and captures images depended on the content of these images and the desired image functions can be retrieved and used as a search for indexes or bases. Basically, the image recovery system includes three primitive content such as: visual content or feature extraction, multidimensional indexing and retrieval.

The images in the database are indexed according to extract sections visual content (or features) such as color, texture, pattern, structure of the image, the shape of the object, the layout and position in the image, etc. The image can be acted as a multidimensional vector of features retrieved from the image. The feature vector perform by image signature. This feature vector can be assumed to be associated to a point in the multidimensional space. For example, the images can be displayed with the N-Dimensional feature vector, where the first element n1 can perform the color of the component. The following n2 can represent the shape of the component. The following n3 can represent the topology of some images and finally the element n4 can represent the surface of the images to contain the components N =n1+n2+n3+n4.

Image queries can be analyzed to distinguish visible and comparable functions to find matches in image indexes stored in databases. The extracted image function is saved as metadata and images indexed depended on this metadata. This metadata contains several dimensions of the extracted image function. The feature vector is to measure several different image functions. The feature vector with the same image is summarized in N-dimension space. The retrieval of an image similar to query image then decreases to find the index of the image in the search space. N-dimensions with vector features in N-dimension space.

2.6 Architecture of a Content-based Image Retrieval System

The architecture for a possible content-based image retrieval system is shown in Figure 2.1.

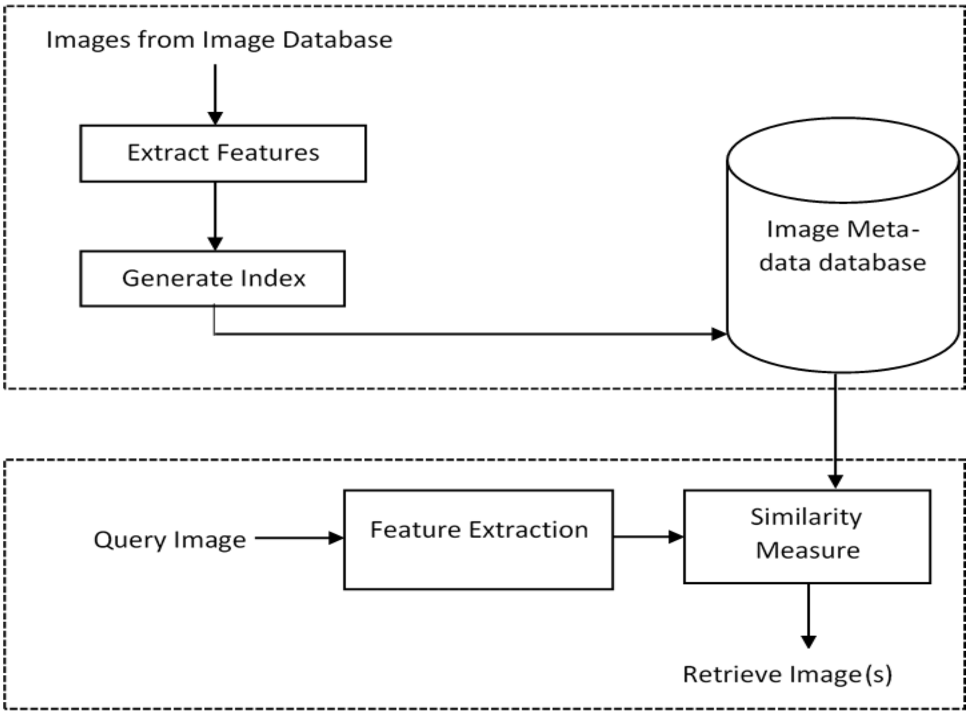


Figure 2.1 Architecture of a Content-based Image Retrieval System

Architecture for the CBIR system is separated into two parts. In the first part, images from the database are proceed offline. Each image features in the image database will be extracted to create image metadata describing the image using the image content features. This function then indexes the image and is stored with the image in the metadata database. The second part describes the image retrieval process. Image query is analyzed to retrieve visible features. These features take same images from the image database. Instead of directly compare both images, the visual properties of the query images are measured by the characteristics of image stored in the meta database. Often, the similarity of both images is determined by calculating the distance between the vector vectors of two images. The system takes the image by returning the first image k, which has the distance from the same request image or less than the specified threshold.

Many image features can index images based on content. The most popular image retrieval system between them is color, texture and shape, image topology, color arrangement, interest region and so on.

2.7 Image Features for Retrieval and Mining

Image features affect all aspects of the CBIR system. It is therefore important to choose the appropriate image properties for each CBIR system.

2.7.1 Color Features

Color is one of the most widely used visual features in content-based image retrieval [12-15].During users may identify a restricted bulk of grayscale, but their eyes can identify thousands of colors and computers can display millions of different colors in the field. The colors used to properly extract images that have a very strong relationship with the underlying object in the image. In addition, the properties of color resistant complications, perspective, attenuation (scaling), alignment (orientation) and image size.

Although the system can use any color space to calculate color histograms in HSV (hue, saturation, value), HLS (hue, brightness, saturation) and CIE color space (such as CIELAB, CIELUV), giving better results than RGB places . Since the color space becomes visual (or viewed) from RGB, they are more effective in measuring the color equation between the images.

2.7.1.1 RGB Color Model

In the RGB model, each color appears in the main components of the spectrum red, green and blue. The RGB format depends on the Cartesian coordinate system. The color sub-space where the RGB core value is at three corners; the secondary colors cyan, magenta and yellow are three corners; black is the origin and the white color is in the farthest corner of the source.

In the RGB model, the gray scale (points with equal RGB values) will extend from black to white along the lines connected to these two points. The color difference in the RGB model is that the top or the inside of the cube is determined by an extended vector from the origin. For convenience, assume that all color values are normalized so that the cube is a unit cube. In other words, all values of red, green and blue are considered in the range [0, 1]. RGB images in which each red, green and blue image is an 8-bit format. The term full-color image is 24-bit format.

2.7.1.2 CMYK (Cyan, Magenta, Yellow, Black)

One more attraction color model uses CMYK (cyan, magenta, yellow and black). This model is useful for color printers. Most output devices, including printers or color copiers, use CMY color templates and additives. The main colors are red, green and blue, while the primary colors of the pigments are magenta, cyan and yellow and the secondary colors are red, green and blue. The conversion from RGB to CMY can be done in case of R, G, and B which shows the normal color value in the range from 0 to1.

You can see from above that cyan coated surfaces do not have red or dark skin turning from magenta to green. It can be seen that the same number of pigments as primary colors (such as cyan, magenta and yellow) produce black. Therefore, the four-color system is cyan (C), magenta (M), yellow (Y), and black (B) in the four-color model.

2.7.1.3 HSI (Hue, Saturation, Intensity)

HSI is a color space that describes colors seen by humans. HSI (or HSV) means Tone (H), Saturation (S) and Intensity (I) (or V Value). Hue have been mentioned as chromatic light characteristics. It can be regarded as a surface property that is reflected or transmitted by light. For example, the blue car reflects the blue hue. It is also a feature of human perception. Hue, which is an important part of our perception, can be seen as a weal or hue or strong hue. The color intensity is explained by the saturation component. For example, the color of a single monochromatic light source that produces only one color of the wavelength is very saturated, while the colors consisting of different wavelengths have color intensity less chroma and less saturation. The gray color has no hues and therefore zero saturation or no saturation. Therefore, the saturation is the size of the color or whiteness in color perception. The brightness (L) or intensity (I) or value (V) basically indicates the measurement of color brightness. This allows you to measure the amount of light shown by the object or the amount of light emitted from the region. It is proportional to the electromagnetic energy the object emits. Brightness (or intensity) helps the human eye recognize the color. The colorful objects in the dark do not appear in color at all.

The hue component describe their own colors as angles between (0, 360) degrees. 0 means red, 120 degrees means green, 240 means blue, 60 degrees is yellow, and 300 degrees is magenta. The saturation component indicates the amount of color contaminated by the white. The range of elements S is (0, 1). The intensity range is between (0, 1) and 0 means black, 1 means white.

2.7.1.4 HSV (Hue, Saturation, Values)

HSV color space (color, saturation, value) is often used by users who choose color (like color or ink) color wheel or color palette because it reflects the way people see colors over RGB color space. Touch the color space. Since this color has a value between 0 and 1.0, the same color varies from red to yellow, green, blue, cyan, magenta and back to red, so there are actually red values 0 and 1.0. When color saturation ranges from 0 to 1.0, the relevant colors (hue) range from unsaturated color (gray level) to fully saturated color (no white level). Since the value or brightness varies between 0 and 1.0, the same color always becomes progressively lighter.

Hue, the type of color (such as red, blue or yellow) can be measured between 0 and 360 by central tendency of the wavelength, the saturation of the color and the intensity of the color (or how much greyness is present) 0-100% of the value for the amplitude of the wavelength, the brightness of the color, measured in the value 0-100% by the diffusion of the wavelengths.

The value of hue is 0, which means red, green, the color is the value corresponding to 120 and the blue color is the value corresponding to 240.The horizontal plane through the hexagonal cone, the primary and the secondary (red, yellow, green, blue, cyan and magenta) occurs at the top of a hexagon. The two saturation components describe the intensity of the color. The saturation value of the color 0 (in the center of the hexagon) means that the color is "colorless" (gray), the highest saturation value (on the outer edge of the hexagon) means that the color is in the "colorfulness" for that angle, and brightness.

The value component (in the HSV space) and the lightness component (in the HLS space) describe the brightness or luminance. HLS (hue, lightness, saturation) similar to HSI: this word uses lightness instead of intensity. In the HSV and HLS color spaces, the value 0 means black. In the HSV space, the maximum value means that the color is the brightest .In the HLS space, the maximum brightness value means that the color is white, regardless of the current value of the hue and the saturation components. The brightest and most intense color in the HLS space appears at half of the highest value.

The difference between HIS and HSV is to calculate the brightness component (I or V), which determines the distribution and dynamics of brightness (I or V) and saturation (S). The HIS color space is best for the traditional images processing function such as convolution, equalization, histograms and others that work by manipulating the brightness, because I depend on R, G and B. The HSV color space is preferred for the manipulation of hue and saturation (to change the color or adjust the number of colors) because it offers a wider dynamic saturation range.

2.7.2 Texture Features

The texture is a pattern of repetitive pattern or a type of structure having a normal intervals. In the general sense, the texture refers to the surface characteristics and appearances of an object determined by size, shape, density, arrangement, proportion of basic elements. The basic procedure for collecting these features through the texture analysis process is called as texture feature extraction. Due to the definition of texture information, texture feature is the main function of several image processing applications, such as remote sensing, medical imaging and content-based image retrieval.

The texture is a very interesting visual feature used for image characterization with applications for retrieval images from the content. There are no formal definitions of textures in the literature. However, the main feature of the texture is the repetition of patterns or patterns over the region of the image. The composition of the pattern is sometimes called textons. Size, shape, color and orientation of textons may vary by region. The difference between the two textures can be at the level of texton changes. This can also be caused by the spatial statistical distribution of textons in image. The texture is a natural property of almost all surfaces such as brick, fabric, wood, carpets, clouds, trees, lands, skin, etc. It contains important information about the layout of the surface structure of the image. While a tiny area of the image has a variety of distinct tonal features, the texture is a dominant feature of that area. In contrast, the gray tone is a major property while the tiny area of image have discrete color tone features.

There are many techniques to measure the similarity of the texture. The Gray level co-occurrence matrix represents the features of the texture in the image. In this way, the co-occurrence matrix is created according to the orientation and the distance between the pixels of the image. Significant statistics of this co-occurrence matrix representing the texture are obtained. Since the basic pattern of the texture is controlled by the appearance of certain gray levels, the presence of the gray level in a predefined relative position may be a reasonable measure of the existence of the texture and periodicity of the patterns.

Many features of the texture, such as entropy, energy, contrast, and homogeneity, can be extracted from the co-occurrence matrix that occurs at the gray level of the image. The gray-level co-occurrence matrix C (ij) is defined by first specifying the vector displacement $dx, y = (\delta x, \delta y),$ then all the pixels separated by $d_{x,y}$ are counted, and there possesses the scaling matrices of gray I and j. The matrix C (ij) is normalized by dividing each element of the matrix by the total number of pixels.

These texture features let to extract images based on a large amount of content. Famous signal processing systems are applied in the analysis of texture and in extraction of texture features of the image. Wavelet transform is used in the texture analysis and the image classification according to the decomposition of multiple points of the images and show the textures at several levels.

The features of the texture can be separated using several methods such as statistical, structural and model-based and transform information.

2.7.2.1 Structural Based Feature Extraction

Structural methods represent well-defined primitive texture and the hierarchy of traditional spatial arrangements. The description of the texture requires the original definition. The advantage of structural method according to the feature extraction is that the method provides a good description of the symbol of the image. However, this feature is useful for image synthesis instead of analysis work. This method is not suitable for natural surfaces due to the variety of micro and macro-texture.

2.7.2.2 Statistical Based Feature Extraction

Statistical methods characterize the texture indirectly as a function of undefined properties that manage the relationship between the gray levels of the images. Statistical methods are used to analyze the spatial distribution of gray values by calculating the local features at each point of the image and to obtain a set of statistics on the distribution of local features. Statistical methods can be divided into first order (one pixel), statistics of the second order (pair of pixels) and higher (three pixels or more) statistics. Statistics of the first order make it possible to estimate the properties (as well as the mean and variance) of each pixel value, with the exception of spatial interactions between pixels in the image. Second order statistics and higher order statics measure the nature of two or more pixel values occurring in a particular position relative to one another. The second most common statistical feature used for texture analysis comes from the co-occurrence matrix.

2.7.2.3 Model Based Feature Extraction

The analysis of the texture of the model, that is fractal models and Markov model, depends on image structure which can be applied to describe the texture and the synthesis. This method determines the image as a probability model or a linear combination of basic function sets. Fractal models are useful for creating natural textures models with statistical roughness quality at different scales and their own similarities, as well as for texture analysis and discrimination. There are techniques to extract the features according to different models, depending on the neighboring system and the noise source. The different ways are moving average (MA), auto regressive (AR), auto regressive moving average (ARMA) and one-dimensional time series models. Random field model analyzes spatial variation in two dimensions. The global random field model considers all images as a realization of random fields and random field patterns at this location, taking into account the intensity ratio in small neighborhoods. Markov models, which has the probability of having the condition of the specified pixel intensity, depends only on the intensity of the pixel in that area.

2.7.2.4 Transform Based Feature Extraction

Transform methods, such as Fourier transform and wavelet transform, represent images in space for which the coordinate system has interpretations that is closely related to the characteristics of the texture. Fourier transform methods have weaknesses in spatial localization so these things do not work well. Gabor filter provides methods for better spatial localization but its advantages are limited in practice because there is usually no single filter resolution to limit the spatial structure on the natural textures. These methods involve converting the original image by using filters and calculating the energy of the transformed image. This depends on the complete image process, which is not suitable for some applications based on part of the input image. Between the various wavelet filters, Gabor filters are very efficient for the analysis of texture.

2.7.3 Shape Features

The other image feature in the CBIR is shape. This is a description of the object of rejection of position, orientation and size. Therefore, the feature of the shape are not different in terms of translation, rotation and scaling for CBIR which is effective when the order of the object is unknown first. To apply shapes as a feature of image, you need to segment the image segment to recognize the region or object. That is challenge. Form characterization system can be broken into two types. The first type is the edge that uses the contour outside the shape of the object. The second type depends on the region where all the forms of objects are used. The most important representation in both categories is the Fourier descriptors [13] and the moment variants [14]. The main concept at the base of Fourier's descriptors is the use of the Fourier boundary changing the object as a characteristic form, while the temporary variant concept is the use of geometric moment based on the region that has changed translations and rotation.

2.7.4 Edge and Boundary Features

The edges of the image are areas of high intensity and a jump the intensity from one pixel to another. A pixel can make a significant difference in the quality of the image. Edge detection of the image significantly reduces the amount of data and filters out non-critical information while preserving the important features of the image. The edges are scale-dependent and the edges may have other edges, but to some extent the edges are without width. If precise edges are specified, all objects will be placed and basic properties such as area, perimeter and shape can be easily measured. Therefore, the edges are used to estimate the boundaries and segmentation of the scene.

2.7.4.1 Sobel Technique

The Sobel edge detection technique consists of 3 x 3 convolution kernels, a kernel is rotated at 90 °, as shown in Figure 2.2. These kernels are designed to respond to vertically and horizontally maximum in relation to the pixel grid of the images, which is a kernel for both perpendicular orientations. The kernels can be used separately with the input images to create measurements, separating the gradient component in each orientation. These things can be combined to find the exact magnitude of the gradient at each point and the orientation of the gradient.

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

G_x

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

G_y

Figure 2.2 Masks Used for Sobel Operator

2.7.4.2 Robert Technique

Robert cross operator performs simple and fast, 2-D spatial measurements in images. The pixel value at each point of the output represents the approximate absolute size of the spatial gradient of the input image at that point. The operator contains 2 x 2 convolution kernels as shown in Figure 2.3. One kernel is simply that the other rotates at 90 °. This is very similar to the Sobel operator.

+1	0
0	-1

G_x

0	+1
-1	0

G_y

Figure 2.3 Masks Used for Robert Operator

2.7.4.3 Prewitt Technique

The Prewitt operator is similar to the Sobel operator and can detect vertical and horizontal edges of images. Prewitt operators measure two components. The vertical edge element is calculated with the kernel Gx and the horizontal edge component with the kernel Gy, as shown in Figure 2.4 | Gx | + | Gy | provides an indication of the gradient intensity in the current pixel.

-1	0	+1
-1	0	+1
-1	0	+1

G_x

+1	0	-1
+1	0	-1
+1	0	-1

G_y

Figure 2.4 Masks Used for Prewitt Operator

2.7.4.4 Canny Technique

Canny's edge detection algorithm is widely recognized as the most appropriate edge detector. The Canny algorithm uses the most appropriate edge detector based on the set of criteria, which includes looking the most edges by reducing the error rate, making the edges as much as possible to the actual edges to maximize localization, and making the edges only one when a single edge exits for minimal response. According to Canny, to obtain the least possible response, the best filter that meets the three criteria can be estimated efficiently using the first derivative of the Gaussian function. The first step is to smooth the image using Gaussian filter. Next, look for the gradient of the image by providing a smoothed image through the convolution operation with Gaussian derivatives both vertically and horizontally. This process reduces the problems associated with edge discontinuity by identifying the strong edges, and associated weak edge processing, in addition to maintaining noise reduction levels. Eventually, hysteresis is work to prevent the band. Streaking is a breakdown of the edge contour caused by the shocking operator's expenses above and below the threshold.

2.7.5 Multidimensional Indexing

This is a significant part of retrieving images from content. The improvement for the indexing system is applied in database management, calculation geometry and model recognition. But, indexing concepts differ slightly in different cooperation. The concept of multimedia data navigation indexing and content-based image retrieval differ from the concept of a traditional database management system. In the traditional DBMS (especially for contact databases), indexing refers to the accessibility of database files in terms of organizational records. Index is determined by one or more attributes to process queries based on this attribute. The structure of this file and record is structured and aided by access structures such as hashing, B-tree and so on. For community of IR (information retrieval), the indexing mechanism refers to an associated term process (or expression or keyword or description) on the document to be obtained under such circumstances. Indexing in the retrieval of images, applied content or mining multimedia information is similar to the idea used to obtain information. The main purpose of indexing is to determine the exact description of the data to determine the contents of the data. Description of multimedia information is retrieved corresponding to some features or data feature vectors. Then describes this content to classify it as the appropriate access structure for recovery. The main problems with indexing content retrieval are: high dimension reduction for feature vectors, find the effective data structure for indexing and find the appropriate action.

In CBIR, function vector sizes are often very high. Before indexing, it is important to reduce the size of the feature vector. After extracting the features related to the query, it is important to use the appropriate equality steps to extract data from the database. Steps of similarity based on analysis of statistical are presented in terms of CBIR measures such as Euclidean distance, Mahalanobis distance, Manhattan distance and other same methods. Measuring similarities distance and histogram intersection methods were applied for this reasion, especially with color properties.

CHAPTER 3

THE PROPOSED SYSTEM METHODOLOGY

3.1 Color Auto-Correlogram

One of the most significant techniques for content-based image retrieval is color histogram. It is effective for calculation and good for searching results .For an m*n image I, the colors of the image is quantized to C_1, C_2… C_k. The color histogram H (I) = {h_1, h_2 ... h_k}, where the number of pixels is C_i.

$$\Pr(P \in C_i) = \frac{h_i}{m*n} \tag{3.1}$$

The color histogram also shows the possibility of any pixel in image I, which is color Ci. The weakness of the histogram method is that there is no space information in the color histogram

Color auto-correlogram techniques have been proposed to combine spatial information with color histograms. Auto-correlogram of image I for color C_i can be determined with distance k

$$\gamma_{C_i}^{(k)} = \Pr[|p_1 - p_2| = k, p_2 \epsilon I_{C_i|} p_1 \epsilon I_{C_i|}] \tag{3.2}$$

The correlogram feature of image represent in figure 3.1 how the spatial autocorrelation of color changes with distance.

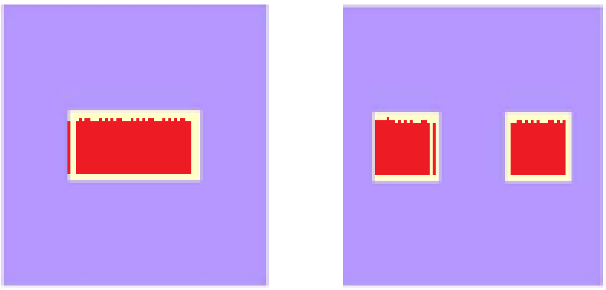


Figure 3.1 Sample color auto-correlogram feature

Auto-correlogram combines information about colors and spaces. For each pixel in the image, the auto-correlogram method must cross all the neighbors' pixel. Correlogram method is more stable in color change than in histogram.

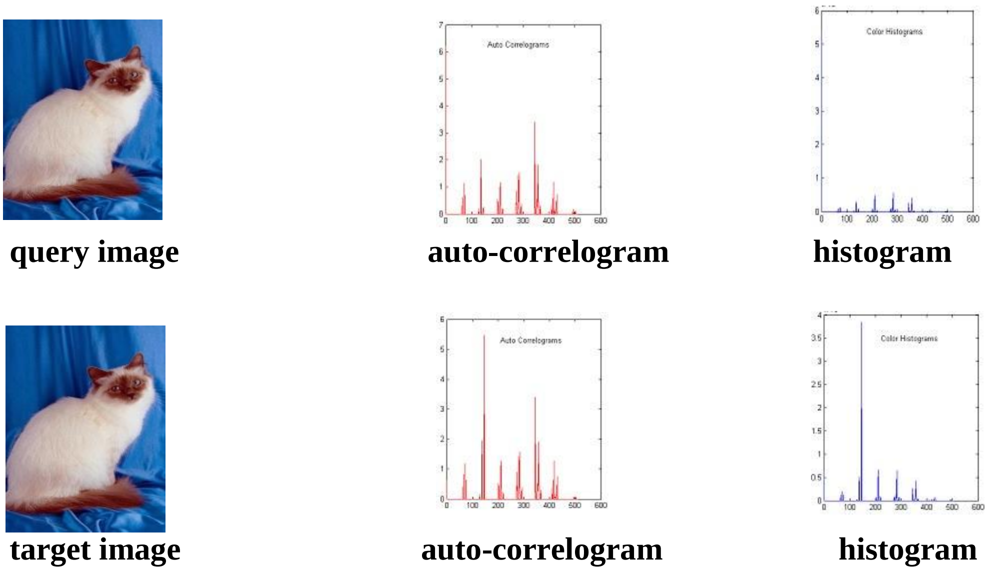


Figure 3.2 Sample color auto-correlogram feature of query and target images

3.2 Color Moments

Color moment is a measurement of the color distribution in an image. It is primarily used for color indexing purposes because of the features of image retrieval applications to compare how similar two images are based on color. It encodes both shape and color information which is a good feature to be used under changing lighting conditions and can be calculated for any color model. This system uses two types of color moments. These are:

(i) Mean

The first color moment can be interpreted as the average color in the image, and it can be calculated by using the following formula:

$$E_i = \sum_{j=1}^{N} \frac{1}{N} p_{ij} \tag{3.3}$$

where N is the number of pixels in the image and pij is the value of the j-th pixel of the image at the i-th color channel

(ii) Standard Deviation

The second color moment is the standard deviation, which is obtained by looking for the square root of the variance of the color distribution.

$$\sigma_i = \sqrt{\frac{1}{N}\left(\sum_{j=1}^{N}(p_{ij} - E_i)^2\right)} \tag{3.4}$$

where E_i is the mean value, or first color moment, for the i-th color channel of the image.

3.3 Gabor Wavelets

This plot is defined as a surface structure formed by repeating certain elements in different relative spatial positions. Repetition includes a local variation of scale, orientation of the elements. Image textures are defined as images of natural textured surfaces. It contains important information about the structural arrangement of the surface: clouds, leaves, bricks, etc.; and also describes the relationship between the surface and the surrounding environment. This feature describes the distinctive physical composition of a surface. Gabor wavelet is widely used to extract texture of the images. Each wavelet capturing energy at a specific frequency and specific orientation. Gabor filter can be shown with the equation 3.5.

$$\Psi(x, y, \lambda, \theta) = \frac{1}{2\pi S_x S_y} e^{-\frac{1}{2}\left(\frac{x_1^2}{S_x^2} + \frac{y_1^2}{S_y^2}\right)} e^{j\frac{2\pi x_1}{\lambda}} \tag{3.5}$$

where (x, y), the pixel position in the spatial domain

- λ, Wavelength of frequency of pixels
- θ, Orientation of a gabor filter
- S_x ,S_y, Standard deviation of the x & y directions

3.4 Types of Classifiers

Image classification is also a machine learning field that uses algorithms mapping all attributes, variables or inputs - function X space - for the definition of class labeled Y. This algorithm is called the classifier. Basically what a classifier does assign a pre-defined class label to a sample .Figure 3.1 display the regular architecture of the classification system.

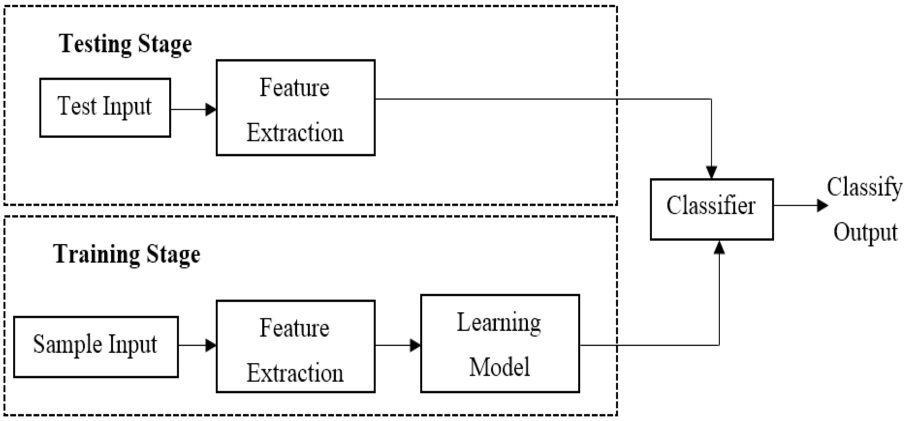


Figure 3.3 Architecture of classification system

Classification is the problem of determining which a set of categories (sub-populations) a new observation are based on all set of training data observed (or instances) whose category membership is known. The purpose of classification is to analyze the input data and to develop an correct description or model for each class using the features present in the data for its most effective and efficient use.

There are two major steps in the classification system such as training step and testing step.

Training defines criteria based on recognized features. In this process, this is getting its individual rules for the classification of the set of training. In the training stage, images are took and saved in the database. And then this is the process of extracting features. As mentioned in Chapter 2, images are represented by several descriptors that construct a feature vector structure. This feature vector is considered an input variable and is included in the learning component. Output is a label associated with the class (e.g. plane, face and flower).

In component learning, this system is created by differentiating and situational models. The first model sends input to the output variable for classification. For field generative models, the distribution of attributes and learning depends on the availability of information. During the test, the query / test image functions as an input. The classification sets the foundation for the learning model with its own classification rules, the feature vector class.

Classification algorithm is separated into supervised and unsupervised algorithms. In the supervised classification, a labeled set of training examples to “train” the algorithm will be used, whereas in the case of unsupervised classifications, the data is collected into a particular cluster without using a training set. Parametric and non-parametric classification is another way of classifying the classification algorithm. The functional form of densities may be the feature vectors of each class known in the parametric method. Otherwise, it is not parametric. On the other hand, no particular form of function is assumed in advance, otherwise the probability density is estimated locally according to training data.

3.4.1 Naïve Bayes Classifier

Naive Bayes is one of the easiest ways to determine the density that can be achieved. Naive Bates classifiers classify patterns in the class primarily on the basis of prior knowledge.

This classifier can control the numbers representing independent variables, whether continuous or categorical.

Determine a set of feature vectors, = $\{X_{1,}\ X_{2,\dots\dots\dots,}X_n\}$, the goal is to create the posterior probability for the class C_j among a set of possible results set of classes C = $\{C_{1,}\,C_{2,\,\dots\dots.,}\,C_m\}$.

Using the Bayes theorem, the posterior probability of class C_j that corresponds to X is written as equation number 3.6.

$$p\left(C_j \middle| X_{1,}\,\dots\dots, X_n\right) = \frac{p\left(C_j\right)p\left(X_1, \dots., X_n \middle| C_j\right)}{p(X_1, \dots, X_n)} \qquad (3.6)$$

where $p\left(C_j\right)$ is the prior probability of class C_j , $p\left(X_1, \dots., X_n \middle| C_j\right)$ is the likelihood of X given class C_j and $p\left(X_1, \dots., X_n \middle| C_j\right)$ is the evidence.

In practice, only the numerator of this fraction is important in equation 1, since the denominator does not belong to C and the value of the X character is indicated, the designation is sincere. The numerator corresponds to the typical probabilistic model that can be recorded handling conditional decision-making applications with repeated conditions, as equation 3.7.

$$\begin{aligned} &P\left(C_j, X_1, \dots., X_n\right) = p\left(C_j\right)p\left(X_1, \dots., X_n | C_j\right) = \\ &p\left(C_j\right)p\left(X_1 | C_j\right)p\left(X_2, \dots., X_n | C_j, X_1\right) == \cdots = \cdots = \\ &p\left(C_j\right)p\left(X_1 | C_j\right)p\left(X_2 | C_j, X_1\right)\dots\ p\left(X_n | C_j, X_1, X_2, \dots., X_{n-1}\right) \end{aligned} \qquad (3.7)$$

The independence nave conditional assumptions ensures that each X_i feature vector is independent of the conditions of the other feature vector X_k for $k \neq i$ at equation 3.8.

(3.8)

$$p(X_i | C_j, X_k) = p(X_i | \mathrm{C})$$

The Naive Bayes classifier combines this model with decision rules. A general rule is to choose the most likely hypothesis. This is called the maximum a posterior (MAP) decision rule.

In image classification, many complex applications using the Naïve Bayes classifier have been successfully implemented. They proposed a hierarchical classification method where they took images indoor or outdoor at first. Outdoor images are classified into images of the city or the landscape. In the end, the landscape image was classified as sunset, forest and a mountain.

3.4.2 The K-Nearest Neighbor Classifier

The K-nearest neighbor classifier is an example of a classifier that does not use parameters. The basic algorithm of these classifiers are simple. For each input feature vector to be classified, the search is performed to find the location of the K-nearest training sample, and then assign the input to the class having the largest members in this location. Euclidean distance is often used as the metric to measure neighborhoods. In the particular case of K=1 the system will receive the nearest neighbor classifier, which will assigns the input feature vector to the same class as the nearest training vector. The Euclidean distance between feature vectors X = {x_1, x_2,......, x_n} and Y= {y_1, y_2, ….., y_n } are obtained from equation 3.9.

$$\mathrm{d} = \sqrt{\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{x_i} - \mathrm{y_i})} \qquad (3.9)$$

However, some elements must be taken into account when using the KNN classifier. Euclidean distance measurements are often used in KNN algorithms. In some cases, using this metric may result in undesirable results. For example, in the case of several sets of features (the feature set is large enough), used as input combined with the KNN classifier, KNN will have a higher value than terrible performance the possible way to avoid this problem is to make the set of normal functions.

The KNN algorithm is as simple as described above. However, there are things must be taken into account when using KNN classifiers. The Euclidean distance measurements are often used in the KNN algorithms. In some problems, the use of this metric can lead to undesirable results. For example, in the cases of several sets of features (the feature set is large enough), used as input combined with KNN classifier, KNN is affected by a larger amount. This leads to a poor performance. The possible way to avoid this problem is to make the set of normal functions.

In Figure 3.4, show examples of three class assignments. The goal is to apply the KNN classifier to find the class of unknown features X. As shown in the figure, the nearest neighbor's image (K = 5 neighbors) is four in class a and there is only one class b, so X is assigned to class a.

The disadvantages of the nearest K classifier include:

- Need all the vector features of all training data when the new feature of the vector. It will be classified and the need for large data storage.
- Longer classification times analyze with another classifiers. The K nearest neighbor classifier has some important qualities.
- Does not require training, which is especially useful when new data on training is added.
- Use local information and learn complex functions without having to specify them clearly.

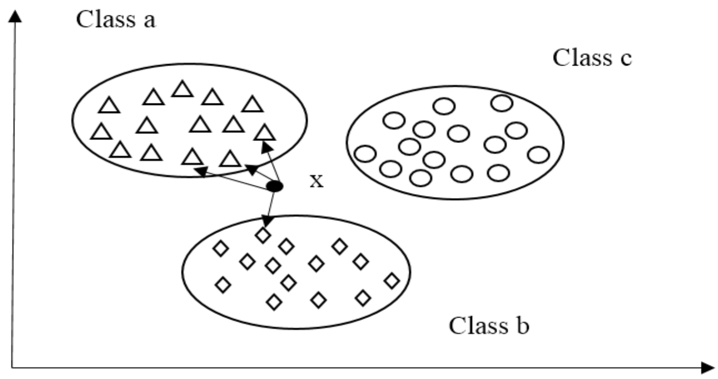


Figure 3.4 The K-nearest Neighborhood Rule (K=5)

3.4.3 Artificial Neural Networks

The Artificial Neural Network (ANN) is a group of interconnected artificial neurons that use mathematical models or computation models for information processing using the connection to calculations method. Neural networks have at least two physical components: the processing elements and the connections between them. A processing element is called neurons and connections between neurons, also called links. Each link has associated weight parameters. The weight of these links represents the knowledge of the network. The knowledge is expressed in artificial neural networks by the form of connection between the processing elements and by the adjustable weight of these links. Each neuron is stimulated by neighboring neurons connected to it, processing the information and generating an output. Neurons that receive stimuli from outside the network (that is, not from the network neurons), called input neurons [14]. Neurons used externally are called output neurons. Neurons that are stimulated by other neurons and by other neurons in the neural network are called hidden neurons. Several methods make it possible to process data by neurons and different ways of connecting the neurons to each other. Different neural network structures can be created using different processing elements and the specific manner in which they are connected. In many networks, neurons are organized in multiple layers. The neural network has one or more layers of neurons, followed by output neurons. Artificial neural networks have more than one class of neurons are sometimes called multilayer neural networks. Several neural network structures have been developed for signal processing, pattern recognition, data classification, and so on.

3.4.4 Rough Sets Classifier

The biggest advantage of an approximate set is the excellent ability to calculate the reduction of information systems. In information systems, some attributes may not be related to the objective concept (decision function) and some redundant attributes. Reduce the need to create a simple and useful knowledge from it. Reduction is an important part of the information system. It is the smallest subset of conditions attributes related to the decision attributes. When there is an incoming e-mail, the system must first select the attributes best suited to the classification. Then, the input dataset become a decision system, which will be divided into training datasets (TRs) and testing datasets (TEs). Classifiers will be induced by TR and applied to TE to obtain a performance evaluation.

3.4.5 Support Vector Machines

A supervised learning model involving learning algorithms that analyzing data. It works by accepting the complete set and then reading it so that each of the relevant input, relevant outputs are extracted. The whole process is considered a classification. Classifying the data by looking for the best hyperplane separates all the data points of a class from the other classes.

Support vector machines depend on the concept of the decision plan that determines the boundaries of the decision. The decision plan is a plan that separates sets of objects with different members in the class. Support vector machine learning algorithm is used to produce the classification parameters according to the calculation feature.

Results can be received in two discrete or continuous formats. The classifier assigns the input space and the feature space. Feature spaces are defined as stored space to calculate similarities using kernel functions.

Support vector machine (SVM) formerly separated binary classes (k = 2) with the maximum margin criteria. However, problems in the real world often require more than two categories of discrimination. Therefore, the recognition of multi-class has a wide range of applications, including optical character recognition, inclusion detection, speech recognition and bioinformatics. In practice, the classification problems of multi classes (k> 2) are generally divided into a set of binary problems, so that the standard SVM can be directly used. The set of two schemes are one-versus-rest (1VR) and one-to-one (1V1) guidelines.

1. One-Versus-Rest Approach

One-versus-one (1VR) method creates a different binary classifier k for k-class classifications. The m-th binary classifier is trained using the m-th class data as a positive example and the remaining k-1 classes is a negative example. During the test, the class label is determined by the binary classifier that gives the highest output value. Imbalanced training is a major problem of 1VR approach. Suppose that every class has an equal training size. The proportion of positive and negative samples in each individual classifier is k-1. In this case, the symmetry of the original problem is lost.

2. One-Versus-One Approach

Another classical method of multi-class classification is a one-to-one d (1V1) or pairwise decomposition. It evaluate all possible pairwise classifiers and thus induces k(k-1)/2 individual binary classifiers. Applying each classifier to a test example would give one vote to the winning class. A test example is labeled to the class with the most votes. The size of classifiers created by the one-versus-one approach is much larger than that of the one-versus-rest approach. However, the size of Quadratic Programming (QP) in each classifier is smaller, which makes it possible to train fast. In addition, compared with the one-versus-rest approach, the one-versus-one method is more symmetric.

SVM Classifier Algorithm

1. Training vectors: x_i, i = 1. . . L
2. Consider a simple case with two classes:
3. Define a vector y:

$$y_i = \begin{cases} 1 \;\; if \;\; x_i \; in \; class \; 1 \\ -1 \; if \; x_i \; in \; class \;\; 2, \end{cases}$$

4. A hyperplane which separates all data and separating hyperplane with:

$$w^T x + b = 0 \qquad (3.10)$$

$$(w^T x_i) + b > 0 \qquad if \; y_i = 1$$

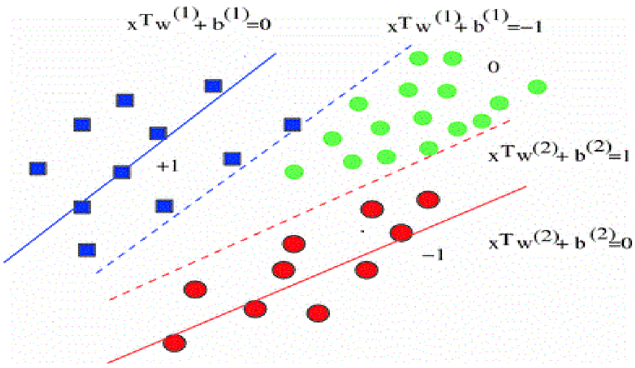
$$(w^T x_i) + b < 0 \qquad if \; y_i = -1$$

5. Decision function $\quad f(x) = sign(w^T x + b), x$:test data $\qquad$ (3.11)

Variables: w and b are coefficients of a plane

6. Select w, b with the maximal margin.

Maximal distance between $\quad w^T x + b = \pm 1$



3.5 Cross Validation

Cross-validation is a better way to evaluate models than others. The problem with the remaining assessments is that they do not say what the leaner will make when asked to make new predictions for data that they have not yet seen. One way to solve this problem is not to use all the date set when it comes to training learners. Some data will be deleted before the training begins. After the training is completed, the data removal can be used to control the performance of the model that learns about "new" data. This is a primary concept for all model evaluation methods, cross validation.

3.5.1 Holdout Method

The holdout method is the simplest method of cross validation. The datasets are divided into two sets, called training sets and testing sets. Functionality calculation is suitable for use only through training set. Next, the approximator estimate the output value of the test set (which never saw the value of these outputs). The error continues to accumulate in advance, giving a real error test used to evaluate the model. The advantage of this is that it is often better than the rest and does not have to be counted. However, the evaluation can vary widely, the assessment may depend on the data points in the training module and finally the test set. Therefore, the assessment can be very different depending on the segmentation method.

3.5.2 K-fold Cross Validation

The accuracy of cross-validation with the k-fold is one way to improve over the holdout methods. The dataset is divided into subsets and the holdout method will be repeated k times. A set of k subsets will be used as a test set and the other k-1 subsets will be combined to create training sets. Then, calculate the average error in all k trials. The advantage of this method is that it is less important than how the data is divided. Each data point must be included once in the test set and must be included in the training set k-1 times. The variance of the outcome estimate will be reduced as k increases. The disadvantage of this method is that the training algorithm must be returned from the scratch k times, which means that it takes as long as k times to compute. The change of this approach involves dividing data into test groups and training set k at different times. This advantage is to choose the degree of independence of each set of tests and the average number of trials.

3.5.3 Leave-one-out Cross Validation

Leave-one-out cross validation is K-fold cross validation taken to its logical extreme by K equals N is the number of data points in the set. This means that the division times N, the approximation of the function is practiced with all data except one point and there is a prediction for this point. Just as we used to, we calculated the average error and we used the model evaluation. The leave-one-out cross-validation (LOO-XVE) evaluation is good, but when it first occurs, the calculation seems to be very expensive. Fortunately, leaners with local weight in the area can easily predict the LOO as they normally do. This means that calculating LOO-XVE does not take more time than calculating the remaining errors and is a better way to evaluate the model. In this thesis, the system uses the K-fold cross validation method to evaluate the efficiency of the system.

CHAPTER 4

DESIGN AND IMPLEMETATION OF THE SYSTEM

This chapter describes the design and implementation of image retrieval and classification with a combination of visual content features, such as Color Moment, Color Auto-Correlogram and Gabor Wavelet features. When a new query image enters to the system as an input, it search and retrieving the relevant images from the image database. The system uses a combination of three types of features: Color Moment, Color Auto-Correlogram and Gabor Wavelet features to provide similar images of the user's query images. In order to improve the performance and accuracy of image retrieval, the Support Vector Machine (SVM) classifier is used with visual features. The experience of the system has shown that the CBIR using the SVM classifier with Color Moment, Color Auto-Correlogram and Gabor Wavelet features produced better results than the CBIR based on these features.

4.1 The Overview of the System Architecture

The architecture of the content based image retrieval system with SVM classifier is described in Figure 4.1. The system accepts input images from users. Then the Color Moment, Color Auto-correlogram and Gabor Wavelet features will be extracted from the query image, and then the system will extract the image by comparing the similarity of the measurement between the query image and the images of the database stored with the SVM classifier to efficient query results. Finally, the system will compare the similarities values of these features with the features of the image database. The same value for each feature for query image, and database images are calculated by Euclidean distance.

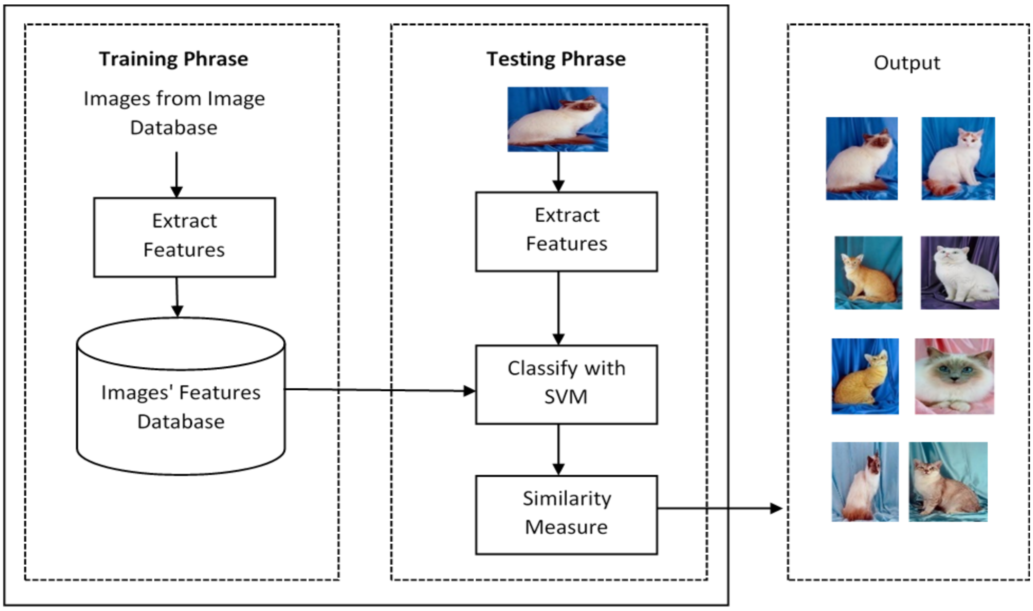


Figure 4.1 Content-based Image Retrieval Architecture

4.2. System Design

Figure 4.2 shows the process flow of the Contents Based Image Retrieval System (CBIR). The system has the three main processes. The first process is the creation of images features database. The system extracts the Color Moment, Color Auto-correlogram and Gabor Wavelet features of each image from images database to get the characteristics of images. The second is image retrieval based on the three types of visual features. In these stages, the feature of query image which is the user input extracts to get the features of image. And then, the system continues to match the feature values of query image with the values from image database by using the fusion of features. The final one of these stages is the retrieval of similar images from database by matching the feature of query image with similarity value. The third step is the combination of classification and image retrieval with SVM to get better results of the CBIR. SVM classifier will be used Color Moment, Color Auto-correlogram and Gabor Wavelet features as attributes of classifier to generate retrieval results.

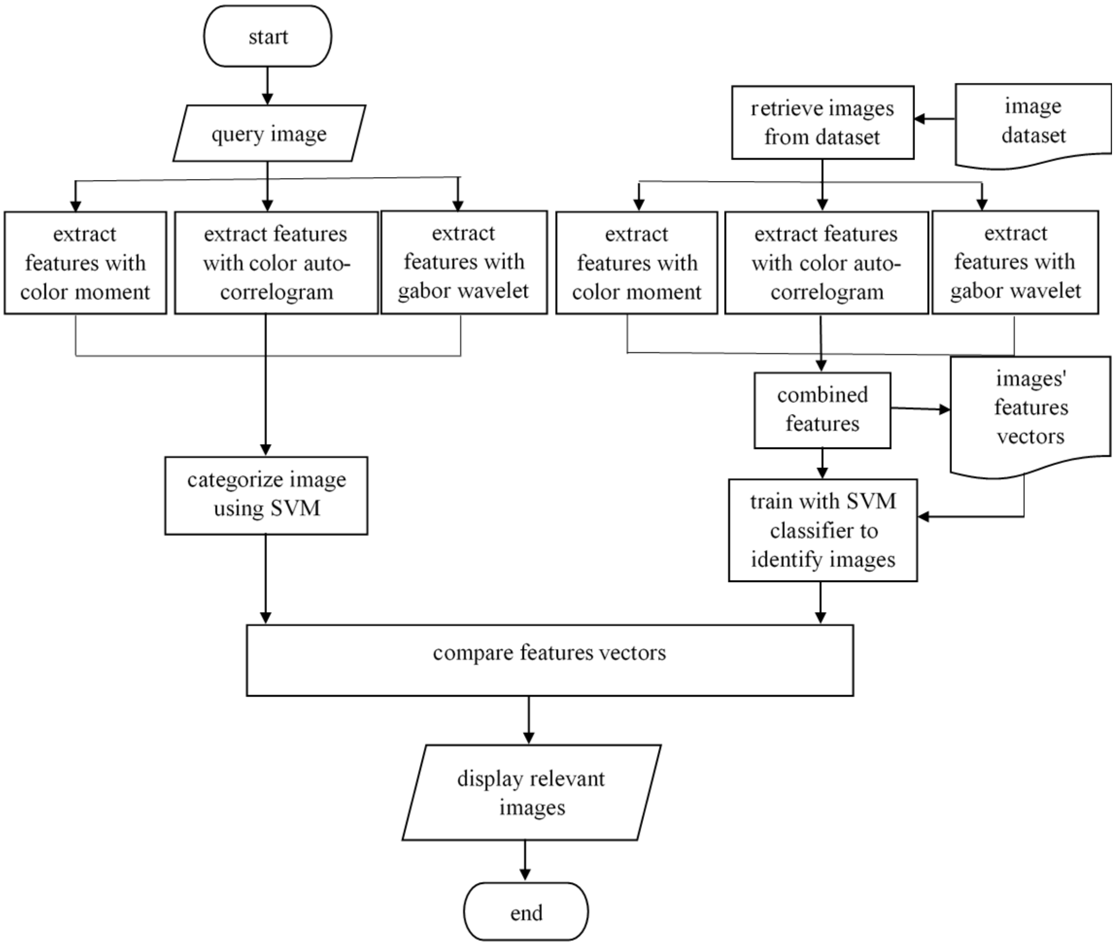


Figure 4.2 Process Flow of the System

The system used Holdout Method cross validation model (section 3.5.1) to divide training and testing dataset and model evaluation. The input dataset is partitioned into 2 subsets for training and testing purpose using Holdout Method cross validation method. The training algorithm runs 2 times. On the first training run, it uses the part of the training data to train as shown in Figure 4.3, and then compute the performance of model using the next part of remaining data to classify the category of images with SVM classifier as shown in Figure 4.4. After finished training and testing cycles, the system generates the retrieval results of query images and class of image. And then, the system calculates the estimate of the accuracy of the model for target images and output images.

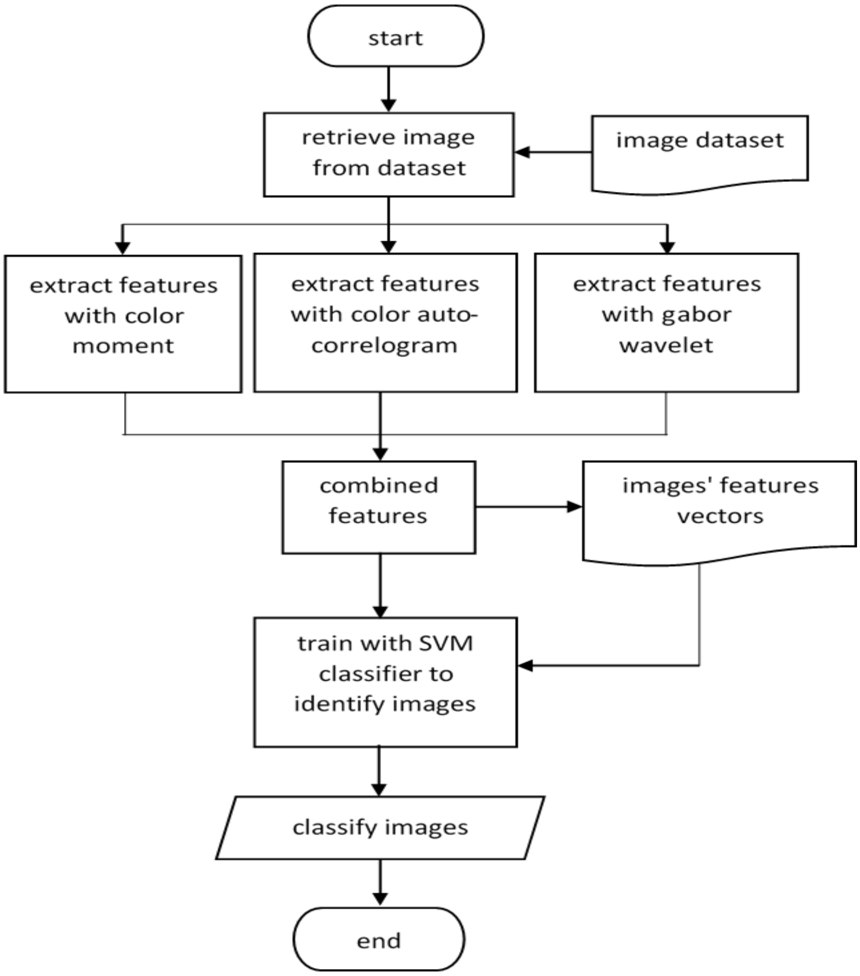


Figure 4.3 The Training Part of SVM Classifier

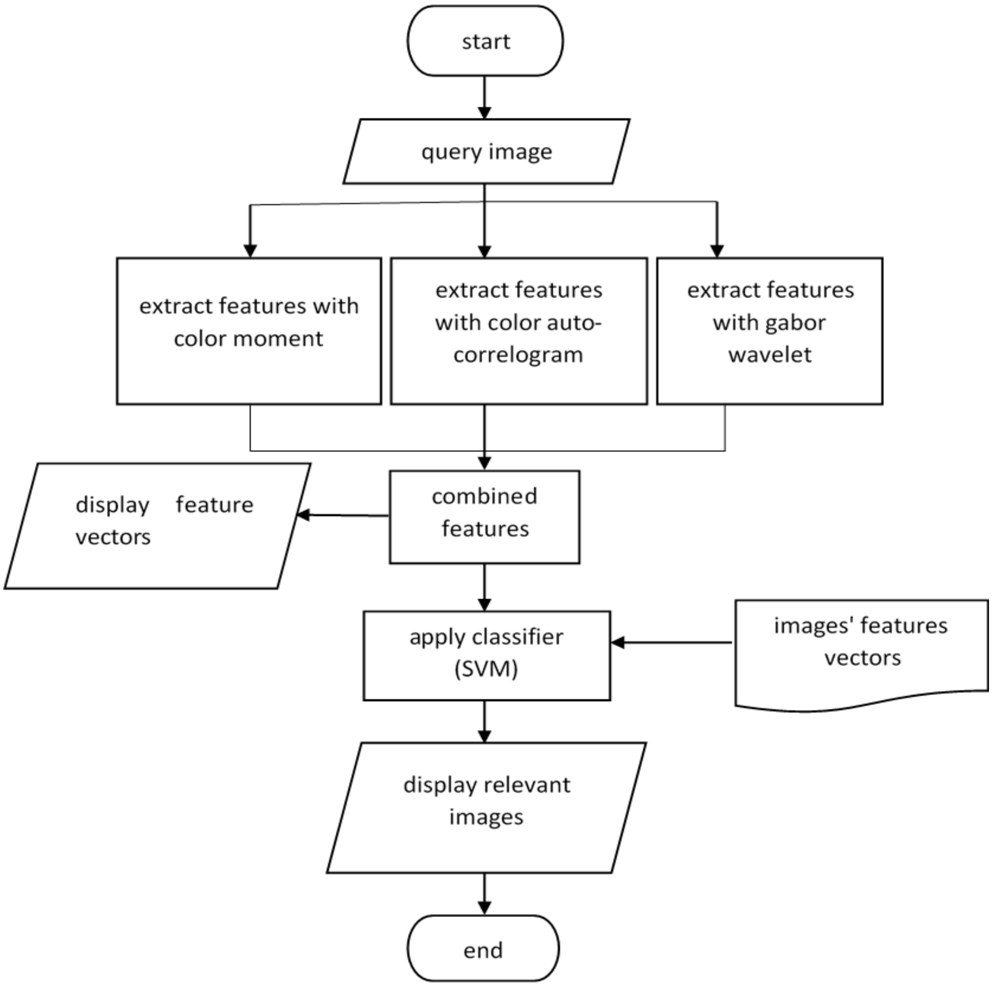


Figure 4.4 The Testing Part of SVM Classifier

In the classification of images, images are classified according to the content of the image. For example, whether it contain rose or cat. An important application is to extract similar images: search through image datasets to receive (or extract) images with specific visual content.

4.2.1 Matching Similarity Measures

After extracting the relevant features of the query image, the system goes on to measure the similarity of the matches to extract similar images from the image database. The measurements for similarity stand on statistical analysis are essential in CBIR. Similarity measurements are made by Manhattan or Euclidean distances for the feature vector values of combined features.

4.2.2 Extraction and Classification of Similar Images

The system divides two parts which are image extraction without SVM classifier and image extraction with SVM classifier. To extract the similar images from database, the user provides the query image to the system as an input.

And then retrieve the query image features and change into the internal representation of feature vectors. The similarities/distances between the feature vectors of the query image and those of images in the database are then calculated and the top similar images are retrieved.

The second part used the features of images as an attributes of SVM classifier to obtain the results of query image and classify the type of query image. The similarity images are extracted from database by filtering SVM classifier.

The system developes the main two types of techniques to retrieve the content based images retrieval. They are as follows:

(i) Image retrieval based on combined the three types of features

(ii) Image retrieval based on combined the three types of features with SVM

4.3 Implementation of the System

The system is implemented with MATLAB programming language on window platform according to the processes of system design. The system accepts the input color image (RGB) with .jpeg extension. According to the feature extraction techniques, the color image is converted to grey scale and HSV color space.

This system has been tested with Corel Image Database, which is available for researchers [14] and user-defined images and no trained images. The database contains 1000 images. There are 10 types, each with 100 images. These categories are 'Dinosaurs', 'Beach', 'Tree_and_Leaves', 'Drink', 'Arabian Horses', 'Moment', 'dogs', 'cats', 'Roses' and 'buses’. And other no trained images consist of 100 images with 10 categories each of which has 10 images. All of these are tested for experimental results. All images in database are in RGB color space which have the same size and the same file format.

Firstly, the system has to create image features database for interesting image folder and keeps the image features database to use for image retrieval. Then, it extracted the three types of visual features of query images which is put from user.

After extracting of relevant features, it calculates similarity distance of the feature values of query image and the images from database. The system gives the topmost similar images from database depended on retrieval method as shown in Figure 4.5 and Figure 4.6 respectively. And then, the system display the better results and classification of query image in Figure 4.7 and Figure 4.8.

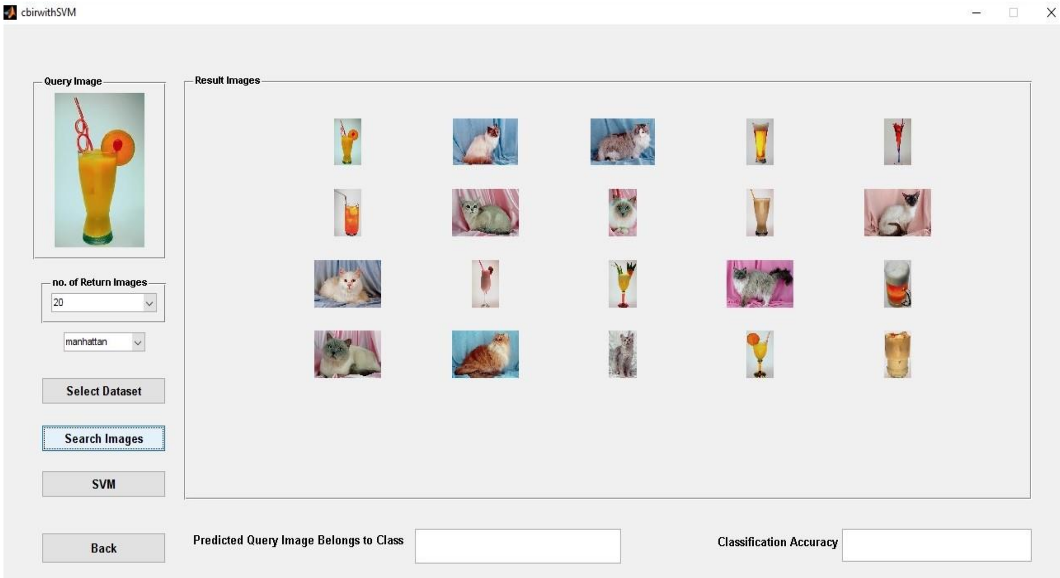


Figure 4.5 Content-based Image Retrieval System with Only Visuals Features for Trained Image

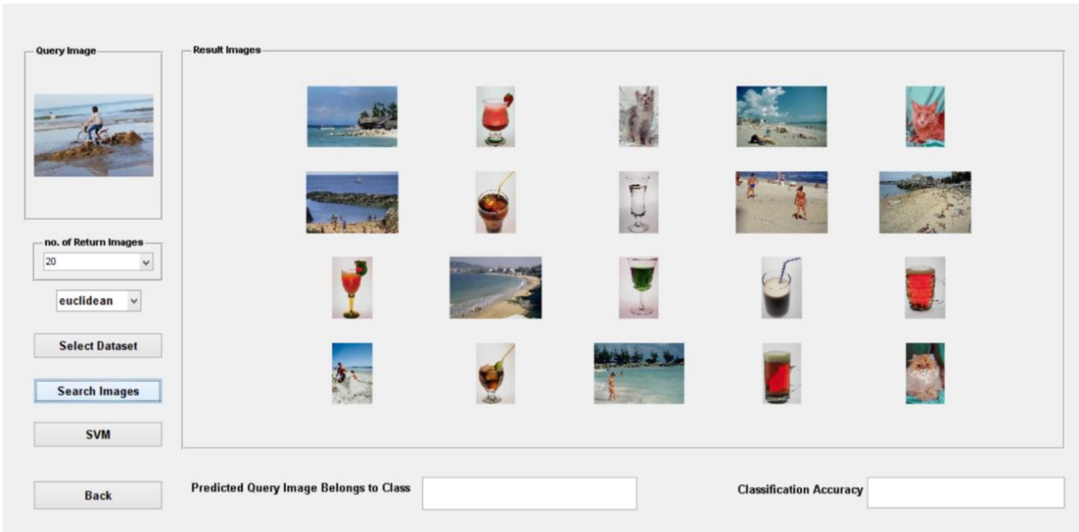


Figure 4.6 Content-based Image Retrieval System with Only Visuals Features for No Trained Image

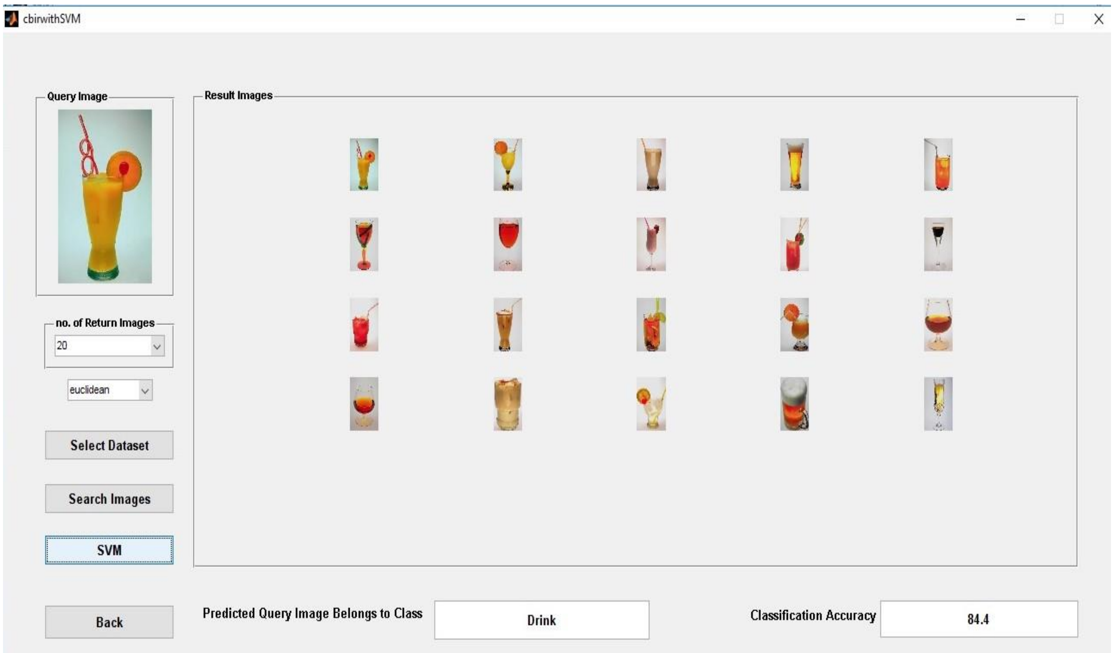


Figure 4.7. Content-based Image Retrieval System with SVM for Trained Image

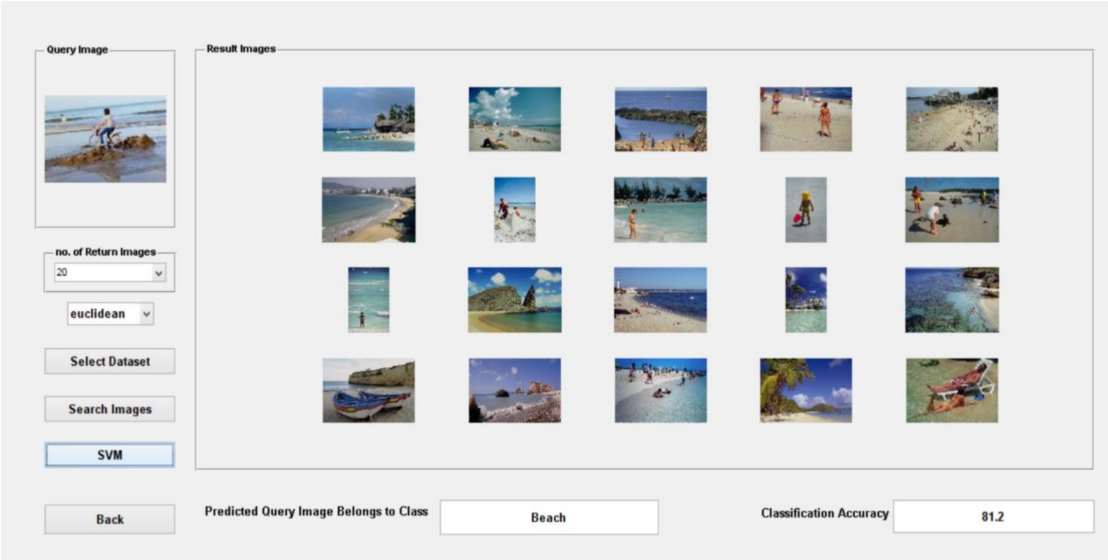


Figure 4.8 Content-based Image Retrieval System with SVM for No Trained Image

4.3.1 Extraction of the Three Types of Features of Query Image

When the user puts the desire query image, the system extracts the color moment features of image. The basis of color moments lays in the assumption that the distribution of color in an image can be interpreted as a probability distribution. Probability distributions are characterized by a number of unique moments. A color can be defined by 3 values. The system restricts to use RGB color model and the first two types of moments are calculated for each of these channels in an image. An image therefore is characterized by 6 moments which means 6x1 vectors containing the first two color moments from each R, G, B channel as shown in Figure 4.9 and Figure 4.10.

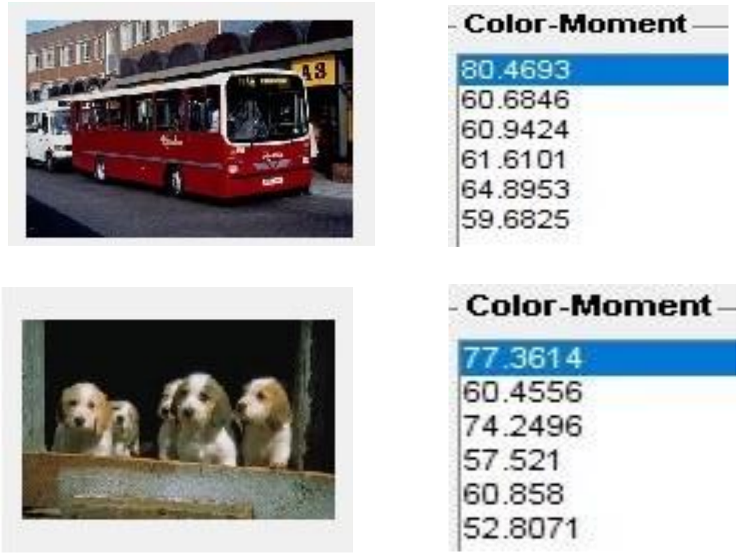


Figure 4.9 The Color Moments Values of Query Image for Trained Image

Figure 4.10 The Color Moments Values of Query Image for No Trained Image

Then, the system continues to extract auto-correlogram feature of query image. An auto-correlogram of image I give the probability of finding identical at distance D. It makes assumptions about the pixels p1 and p2, d away from each other. To extract the color auto correlogram, the system quantizes the input image into 64 colors = 4 x 4 x 4 in RGB color space and generates 64 feature vectors containing the color auto correlogram values. Figure 4.11 and Figure 4.12 is shown the auto correlogram values of input image.

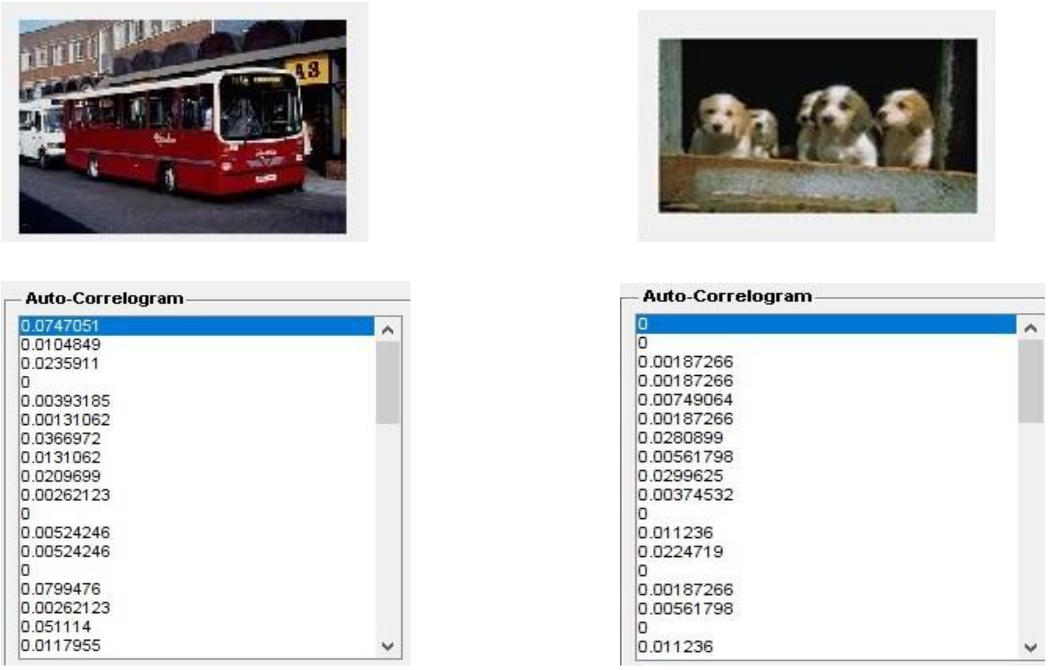


Figure 4.11 The Color auto-correlogram Values of Query Image for Trained Image

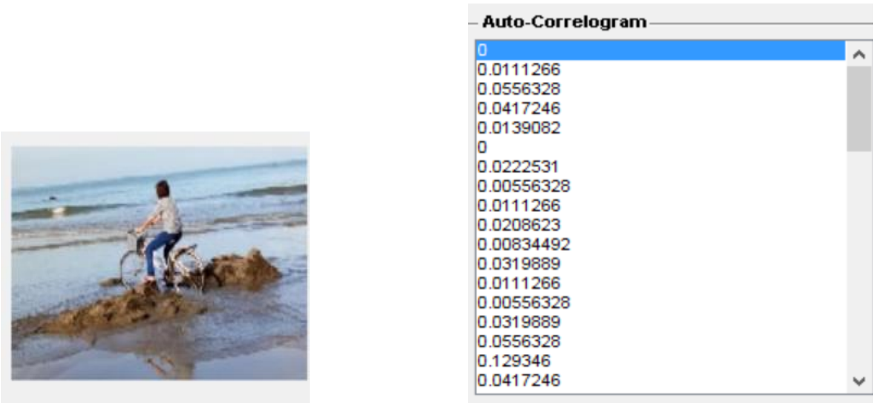


Figure 4.12 The Color auto-correlogram Values of Query Image for No Trained Image

The gabor wavelet features of query image also have to calculate for image retrieval. Basically, Gabor filters are a group of wavelets, with each wavelet capturing energy at a specific frequency and specific orientation. The scale and orientation tunable property of Gabor filter makes it especially useful for texture analysis. This function calculates gabor features. Mean-squared energy and mean Amplitude for each scale and orientation is returned. The system used the number of scales is 4 and the number of orientations 6. The values of Energy and Amplitude values are shown in Figure 4.13 and Figure 4.14.

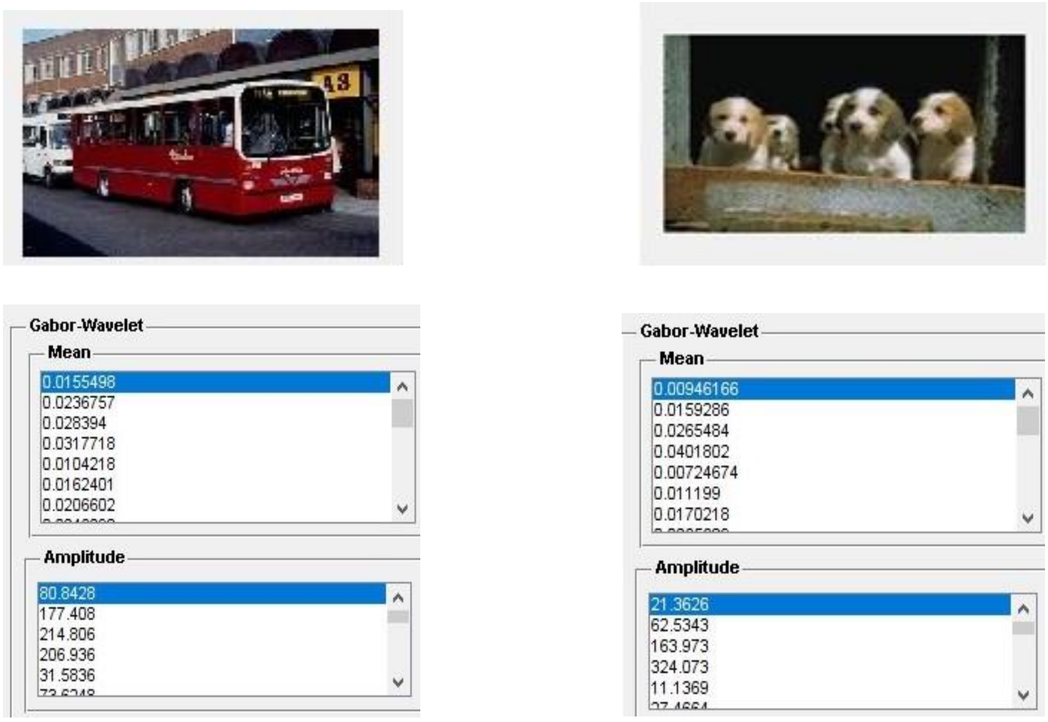


Figure 4.13 The Gabor Wavelet Values of Query Image for Trained Image

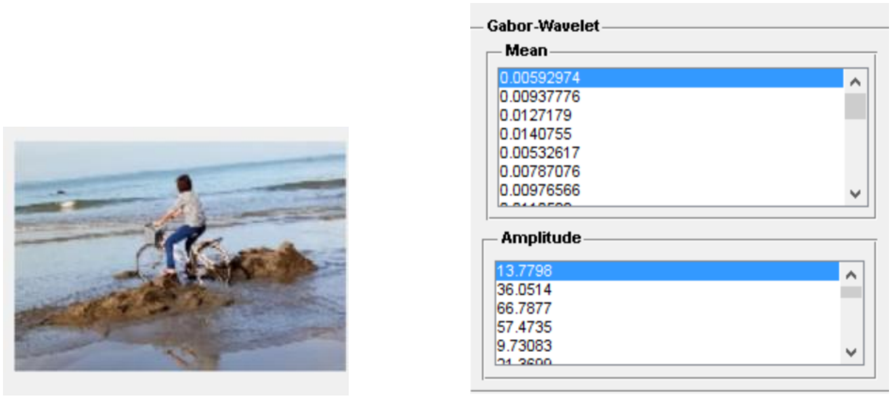


Figure 4.14 The Gabor Wavelet Values of Query Image for No Trained Image

4.3.2 Image Retrieval with Combined the three Visual Features

After creating the features vectors of image database and query image, the system search the similar image of query image from images databases with similarity measures. Results for combined the three visual features based CBIR is shown in Figure 4.15. Visual features are literally defined as consistency of a substance or a surface or vision of images. Technically, it is the visual pattern of information or arrangement of structure found in an image. All of the retrieved images are not red colored or buses which includes different color and images. Figure 4.16 also includes the difference color and images for query image of dogs. The images with similar visual pattern or surface of query image are shown by the following figures. Figure 4.17 also includes the different color and images for query image of beach. So, the obtained result with three visual features is not satisfactory.

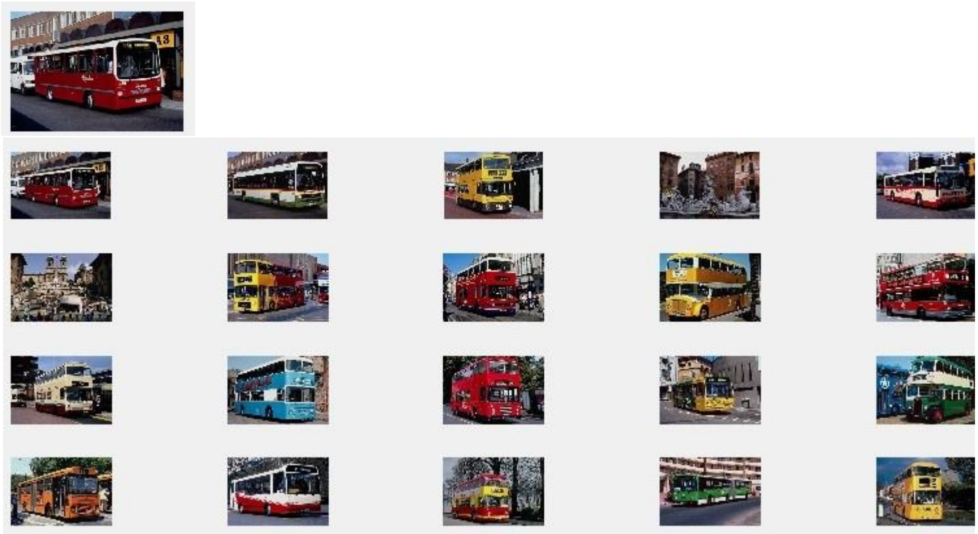


Figure 4.15 Image Retrieval of Buses with Three Visual Features for Trained Image

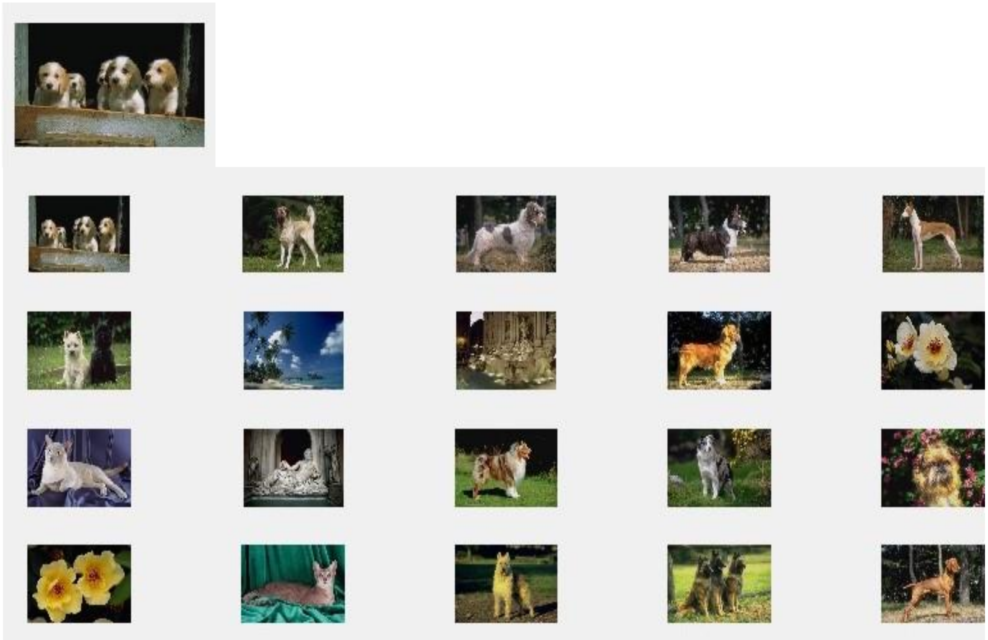


Figure 4.16 Image Retrieval of Dogs with Three Visual Features for No Trained Image

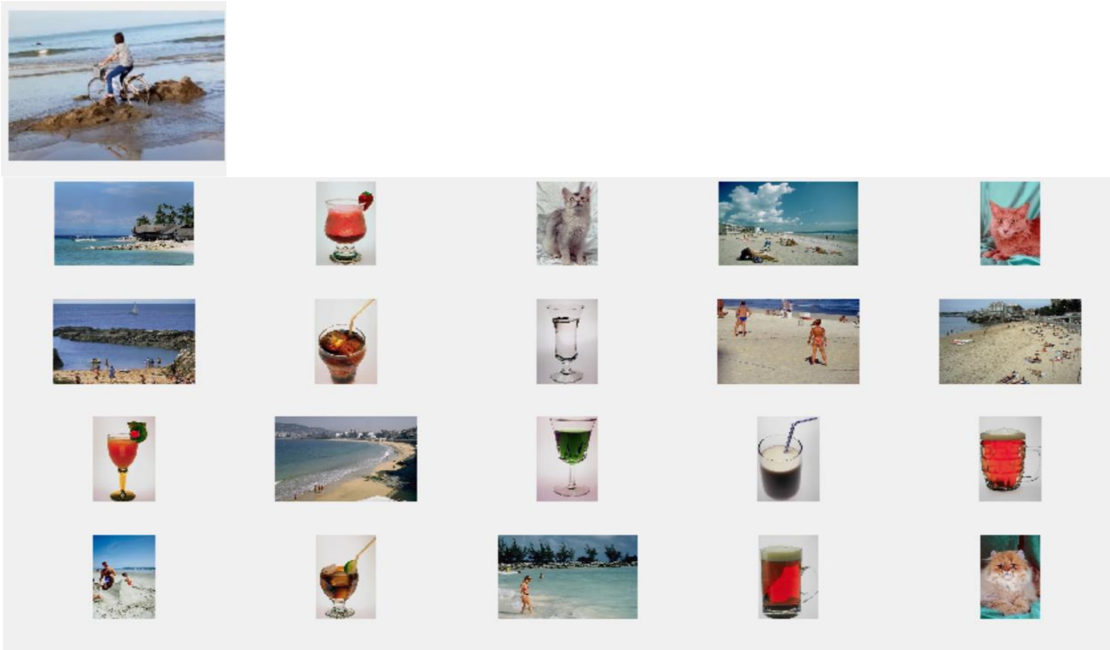


Figure 4.17 Image Retrieval of Beach with Three Visual Features for No Trained Image

4.3.3 Image Retrieval with Combined Visual Features and SVM

When the system combines the Color Auto-correlogram, color moments and auto-correlogram features with SVM classifiers, the results are improved compared to other CBIRs based on three types of features such as Color Auto-correlogram, color moments and auto-correlogram. The query image is a red colored bus and in the results the system receives on all images are buses and bus, most of which are red, as shown in Figure 4.18. The user query image is a dog and in results the system got all the images are dogs and most of dogs have different color as shown in Figure 4.19. The query image is beach that is not trained image and the results got all the images are beaches as shown in Figure 4.20. Table 4.1 represents the values of bias for each class that is used to separate the best hyperplane.

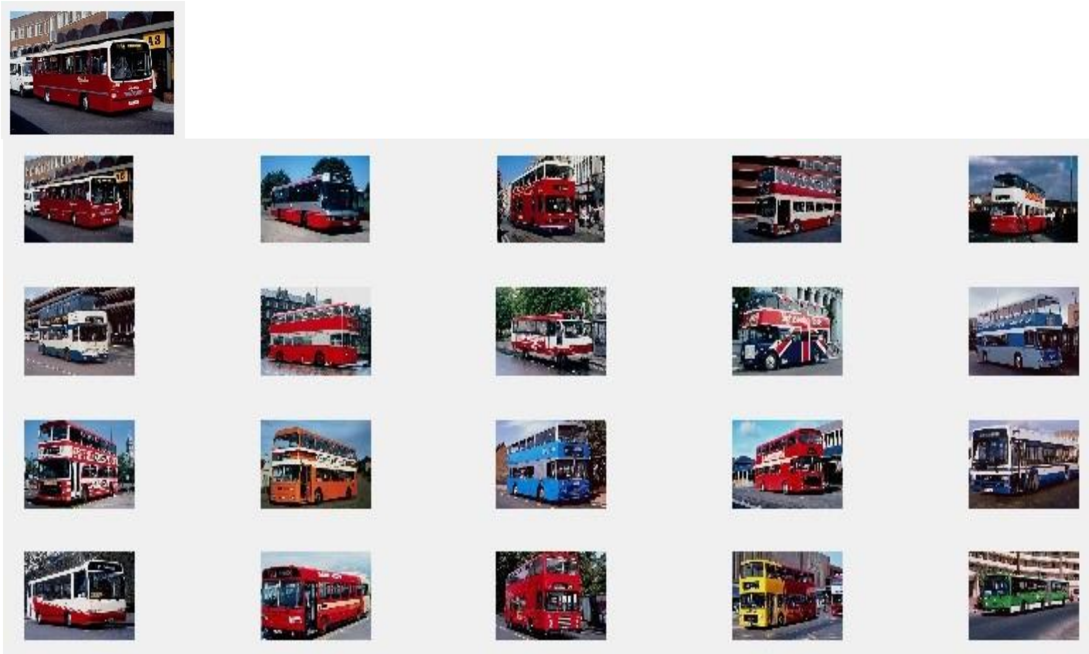


Figure 4.18 Image Retrieval with Combined SVM and Visual Features of Buses for Trained Image

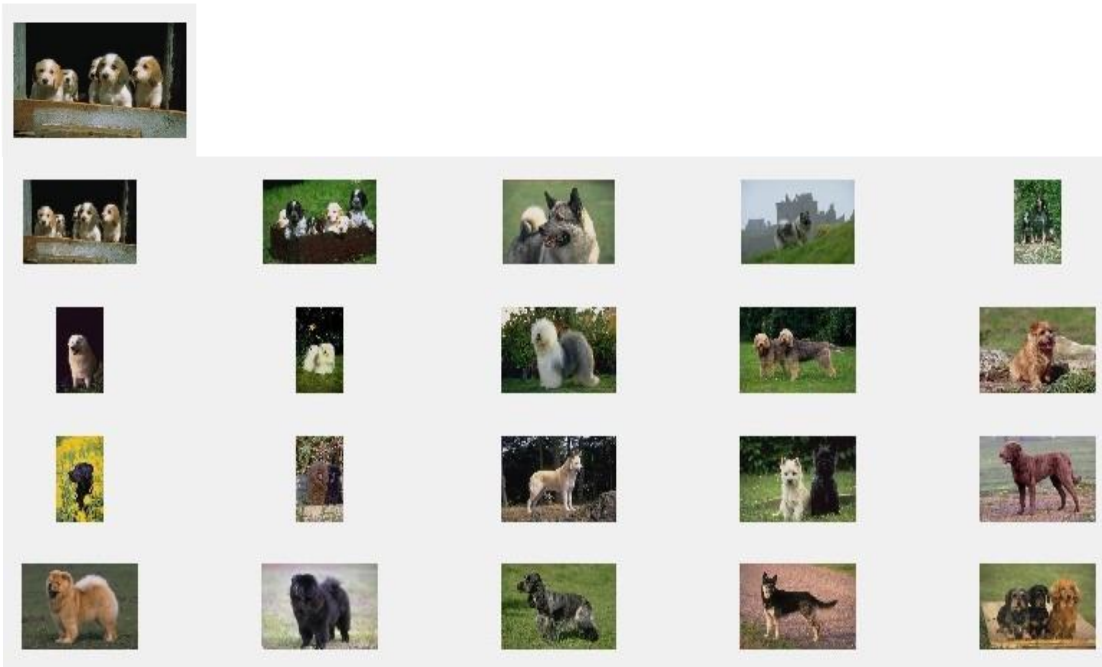


Figure 4.19 Image Retrieval with Combined SVM and Visual Features of Dogs for Trained Image

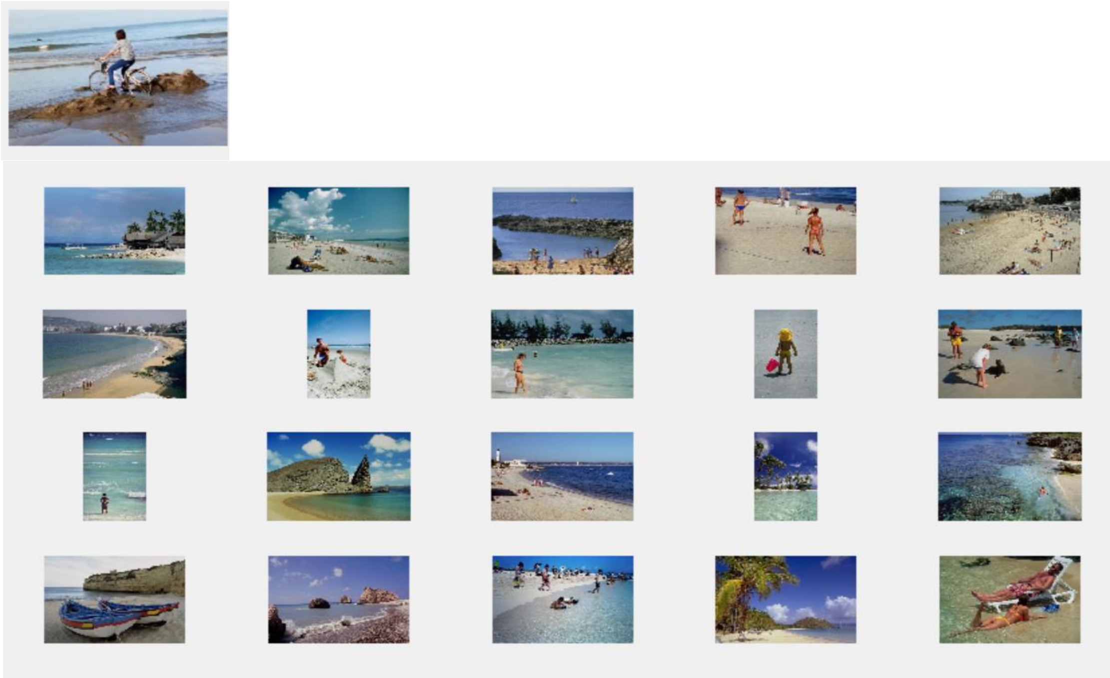


Figure 4.20 Image Retrieval with Combined SVM and Visual Features of Beaches No Trained Image

Table 4.1 The values of bias for each class

Class of data	Bias(b)
Drink	**-0.3710**
Bus	**-0.6630**
Cat	**-0.5411**
Dog	**-0.6012**
Moment	**-0.5490**
Rose	**-0.5183**
Beach	**-0.6797**
Dinosaurs	**-0.8274**
Trees and Leaves	**-0.8629**
Arabian_Horses	**-0.5665**

4.4 Evaluation of the System Performance

SVM classifier is used to classify the type of query image and evaluate the performance of retrieval of the system. The performance of retrieval of the system can be measured in terms of its accuracy. System accuracy is to measure the related images to the query in the entire image of the retrieval image. The system will calculate the accuracy values of each class in images database.

To evaluate the system, it has collected 1000 images which contain the images of Arabian Horses, buses, castles, Roses, etc.; database. Then, it retrieves 20, 40, 60, 80, and 100 numbers of images respectively by using SVM classifier. The system uses the holdout method to evaluate the performance. The dataset is divided into two sets: training set and testing set. The accuracy is calculated by the following equation 4.1:

$$Accuracy = \frac{TP+TN}{Total\ Number\ of\ Test\ data} \quad (4.1)$$

Where,

TP = True Positive, TN = True Negative

Table 4.2 shows the average accuracy and error values of an image regarding SVM classifier with three types of combine features of image retrieval method.

Table 4.2 The average accuracy and error values of the classifier

Class of Data	Accuracy for Trained Images (%)	Error for Trained Images (%)	Accuracy for No Trained Images (%)	Error for No Trained Images (%)
Drink	84.4	15.6	83.8	16.2
Buses	83.8	16.2	80.2	19.8
Cats	83	17	83.4	16.6
Dogs	83.4	16.6	81.8	18.2
Moment	84.2	15.8	81.2	18.8
Roses	83.8	16.2	82.8	17.2

Beach	83.8	16.2	80	20
Dinosaurs	85	15	82.4	17.2
Trees and Leaves	82.8	17.2	82	18
Arabian_Horses	83.8	16.2	80.4	19.6

CHAPTER 5
CONCLUSION

5.1 Overview of the Proposed System

The concepts of image classification and Content Based Image Retrieval have been combined in the proposed system is to classify the images to identify and retrieve the formation of query image. Classifier is generally used to classify the category of each image. CBIR is the process of searching for relevant images in image database when the user sets a new image or query image. The proposed system presented: the fusion of auto-correlogram, color moments, gabor wavelet features and the combination of these three features and support vector machine technique, in similarity image retrieval system.

SVM classifiers can be learned from relevant and irrelevant user-generated image for training data. The experimental results demonstrate that there is considerable increase in retrieval efficiency when the three types of visual features and SVM classifier are combined.

5.2 Limitations and Further Extensions

There are some limitations in the proposed system. The images in the image database must be the file format which is the same. The goal of system is to support image retrieval based on content properties (color, texture), usually encoded into feature vectors. It is used by combining the content properties of the image. The more combination of the contents, the better result of the system will come out. The system must use proper contents extraction techniques and classification techniques to improve the accuracy of system results.

The system will be also considered to further improving the system by fusing more features of useful visual information and help to overcome the shortcomings of only using color, texture and shape features. The implemented system only used RGB and HSV color model for color feature extraction. Other color models such as HIS, CMYK, CMY, etc.; can also be incorporated to extract color feature for image retrieval to improve the accuracy of the system. The system can also be used other improved classifier to generate better results and other visual contents can also be used by combing the contents information. A combinational approach can be used to achieve higher system performance.

AUTHOR'S PUBLICATION

[1] Chu Thinzar, Hlaing Htake Khaung Tin, "Content-Based Image Classification And Retrieval Using Support Vector Machine", to be published in the Journal of Parallel and Soft Computing (PSC 2019), Yangon, Myanmar, 2019.

REFERENCES

[1] Amit Singla1, Meenakshi Garg, "*CBIR Approach Based On Combined HSV, Autocorrelogram, Color Moments and Gabor Wavelet*," International Journal of Engineering and Computer Science ISSN: 2319-7242 Volume 3, Issue 10 October, 2014 pp. 9007-9012.

[2] C. S. Gode, A. N. Ganar,"*Image Retrieval by Using Colour, Texture and Shape Features*", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering , Vol. 3, Issue 4, April 2014.

[3] Emile Karamutsa, Cheruiyot W.K, Anthony luvanda, "*Similarity Image Retrieval using Enhanced Color Histogram Based on Support Vector Machine*," International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 10, October 2015.

[4] H. Mohamadi, A. Shahbahrami and J. Akbari, "*Image retrieval using the combination of text-based and content-based algorithms*," Journal of AI and Data Mining, Published online: 20 February 2013.

[5] K.Valli Madhavi, R.Tamilkodi, R.Bala Dinakar, K.JayaSudha,"*An Innovative Technique for Content Based Image Retrieval Using Color and Texture Features*," International Journal of Innovative Research in Computer and Communication Engineering,Vol. 1, Issue 5, July 2013.

[6] M.C. Hingane, Satish B. Matkar, "*Classification of MRI Brain Image using SVM Classifier*," International Journal of Science Technology & Engineering | Volume 1, Issue 9, March 2015.

[7] Mohammed Alkhawlani, Mohammed Elmogy, Hazem Elbakry, "*Content-Based Image Retrieval using Local Features Descriptors and Bag-of-Visual Words*," International Journal of Advanced Computer Science and Applications, Vol. 6, No. 9, 2015.

[8] R.Senthil Kumar, M.Senthilmurugan, "*Content-Based Image Retrieval System in Medical Applications*," International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 3, March – 2013.

[9] Ramesh K Lingadalli, N.Ramesh, "*Content Based Image Retrieval using Color, Shape and Texture*," International Advanced Research Journal in Science, Engineering and Technology Vol. 2, Issue 6, June 2015.

[10] S.Mangijao Singh , K. Hemachandran,"*Contents Based Image Retrieval using Color Moment and Gabor Texture Feature*," International Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, September 2012.

[11] S.Mangijao Singh, K.Hemachandran, "*Image Retrieval based on the Combination of Color Histogram and Color Moment*," International Journal of Computer Applications Volume 58– No.3, November 2012.

[12] S.Pavani, Shivani, T.Venkat Narayana Rao, Deva Shekar," *Similarity Analysis Of Images Using Content Based Image Retrieval System*," International Journal of Engineering and Computer Science Volume 2, Issue 1, January 2013, pp.251-258.

[13] Sanmukh, N.M. Tejaswini, M.L."*Color histogram features for image retrieval systems*," International Journal of Innovative Research in Science, Engineering and Technology, Vol.3, Issue 4, March 2014.

[14] Su Myat Hlaing, "*Efficient Image Retrieval with Combination of Visual Contents Features*," Master Thesis, University of Computer Studies, Yangon, March 2017.

[15] Suchismita Das, Shruti Garg, G. Sahoo, "*Comparison of Content Based Image Retrieval Systems Using Wavelet and Curvelet Transform*", International Journal of Multimedia & Its Applications, Vol.4, No.4, August 2012.

[16] X.-Y. Wang, B.-B. Zhang, H. Y. Yang, "*Content based image retrieval by integrating color and texture features*," Springer journal of Multimedia Tools and Applications, Volume 68, Issue 3, February 2014, pp. 545–569.