1. INTRODUCTION

Now-a-days, the database frameworks are becoming more famous. The examples include University students’ detail information, library book information, livestock market information and grocery information, etc. These all examples provide the facilities of querying and retrieving the data. Every database has a query facility as a fundamental feature.

We must be able to retrieve better data from a database on the basis of data members’ entities. If we take an example of a library books are differentiated from each other by providing a unique indexing number to every book. People generally search book’s using one of the book’s properties. If the indexing of the book contains one or more feature of the book, the system will sort the books which similar properties.

The same process can be used to develop an image database system. Many areas are related to image database systems, such as weather forecast system, interior designing system, space exploration, biometric system, Global Positioning system, medical imaging system, etc. These areas can be benefited in terms of finance and efficiency by using image database systems. As we know that Indexing is the most important factor in database systems. It determines the way to find and access the data from the database.

An index of images is not similar as an index of a book. Providing an image a name and using it as an index is not the process we desire. We want to atomize the process of indexing and retrieval of images.

If we search an image on the basis of a particular color or image in a database the problem persist as we cannot compare every image and their pixel with the query image and the same time the result of it is equal images but not the similar images. This is due to the identical images with different perspectives in an image data set. We need to use features of an image as its index. Every image has featured in terms of shape, texture, and color. These shape, texture, and color of an image can be mixed to use as an index.
Image features such as shape, texture, and color can be mixed and can be used as an index.

The benefits of this process are as follows:

1. Instead of storing whole image only feature of images are stored such as shape, texture, color as an index.
2. It saves data storage space.
3. It saves time of storage and retrieval.
4. The use of features as an indexing will increase the accuracy.

Thus, by using image features as an index, we get more proficient and effective indexing system.

If the image data sets we have are separated by size, we should make size as an index rather than color, shape or texture. Else if we use color, shape or texture as index in such case it caused the inappropriate result at the time of retrieval.

### 1.1 Content Based Image Retrieval

T. Kato [6] used the term CBIR (Content-Based Image Retrieval) in 1992, to explain the usual retrieval of images from an image database (texture, shape, color) within an image. Since that time, we use this term to explain the process of retrieving preferred images from a huge collection of an image data set on the basis of syntactical image features. Content based indexing & retrieval tools, techniques and algorithms are originating from areas such as pattern recognition, statistics, image processing etc.

Image queries are categorized into three levels:

1. Primitive features such as texture, color, shape, etc.
2. Logical features such as the distinctiveness of the objects shown.
3. Abstract features such as the importance of the scenes depicted [17].

Whilst Content- Based Image Retrieval systems [1] operate most efficiently at the lowest of these levels, most users require higher levels of retrieval, challenging research to reduce the gap between the usefulness of these features and the requirements of the user.

Usually in Content-Based Image Retrieval [19], comparisons are made from a query image and prospective images from the image database by using an image distance technique [18].

An image distance technique compares the match of two images in a range of dimensions such as texture, shape, color, etc. A distance of zero (0) indicates an exact match with the query image. While the value greater than zero (0) indicates a range of
similarities between the images. Then the resulted images can be sorted based on their distance with the queried image. For each image computing distance measures are most often computed by a color histogram. Color histogram identifies the ratio of pixels contained in an image holding specific values [15].

Current research is an attempt to Segment:

1. Color percentage by region.
2. The spatial relationship among several color regions.

The color image analysis is one of the most commonly used techniques. It brings into consideration the color of the image and do not depend on image orientation or size. Color histogram is also one of the techniques used for searching an image based on the color. Color histogram is also one of the techniques uses the color of the image for searching images in an image database.

Texture measures consider visual patterns in an image and find out how they are spatially defined. Texture is a complicated concept to represent and recognition of specific textures in an image. It generally accomplishes by texture, modeling as a two dimensional grey level difference. A comparative brightness of pairs of pixels is calculated on the basis of coarseness, regularity, directionality and contrast.

When we talk about the shape of an image, we do not consider the dimension of the image, but the shape of object resides in that image.

Shapes may be determined

1. By applying edge detection or segmentation to an image.
2. By applying shape filters to recognize given shapes of an image.

In some instances of shape detection, methods like segmentation are very hard to complete automatically and needs human intervention. As we want a fully automated classification system, the shapes cannot be included as the feature descriptors.

Content based search means analyzing the contents of the image instead the metadata such as descriptions, tags, keywords or associated with the picture. The term "content" here refers to shapes, colors, textures, or any other information obtained from the image. The other searches that depends purely on metadata are reliant on annotation quality and completeness and hence Content Based Image Retrieval (CBIR) is desired.

Many times manually (user) annotate images by entering metadata or keywords in a big image data set is found to be very time consuming and may not capture the preferred keywords describing the image. The benefits of keyword image
search are subjective and still not been well-defined. Similar challenges are with CBIR systems.

In the early method of CBIR global feature extraction was used to obtain the image descriptor. For instance, The IBM Almaden Research Center developed QBIC (Flickner, 1995) [13] that extracted several features from each image, namely color, texture and physical body characteristics. These descriptors are obtained globally by extracting

1. Global texture information on contrast, coarseness, and direction.
2. Shape features about the moment invariants, curvature, eccentricity and circularity.
3. Information on the means of color histograms for color features.

Also global features (to represent image semantics) were used by the Photobook system (Pentland et al, 1996), Virage (Gupta and Jain, 1997), VisualSeek (Smith and Chang 1997) [14]. But, these global approaches are not sufficient to solve

1. The image search queries where specific objects in an image having the specific texture and colors are present.
2. The shift/scale invariant queries, where the dimension and the position of the query objects may not relevant.

For example, suppose in one image there are two balls with different colors: green and red. The global features describe the image as the global average color which is black. But in this case, black is certainly not a feature of the image. This shows the weakness of the global features technique.

Region-based retrieval systems [16] try to overcome shortcomings of global based retrieval systems. It represents images as collections of regions that may correspond to objects such as car, airplane, faces, trees, skies, flowers, etc. According to Shi and Malik, 1997 region based retrieval systems apply image segmentation to decompose an image into regions, which correspond to physical objects like car, airplane, faces, trees, skies, flowers. Here region based feature descriptors are extracted on each object instead of global image. For each pixel that belongs to object its features like shape, color, texture is extracted and the average value of pixel feature describe a related object.

So, CBIR is the process of selecting a subset of an image database corresponding to a description given by the query. The goal of this chapter is to give an insight on how
this description can be formulated, how relevant images can be selected and how the data allowing this task can be extracted from the image database.

CBIR is the image retrieval process based on their visual features such as shape, color and texture [3]. It was Kato, who used the term content based image retrieval first time. In his experiments he uses color, shape feature to retrieve images from the image database. After that the term CBIR has become popular, for the process of retrieving images from a large collection of images. For many image database applications, CBIR systems provide a dependable instrument.

There are several benefits of CBIR techniques over other text-based retrieval techniques. We use CBIR system for image data management systems in medical imagery, satellite imagery, and criminology for a better solution.

Since digitalized images consist of an array of pixel intensities, with no inherent meaning, CBIR differs from traditional information retrieval. One of the major issues with every type of image processing is the requirement to extract useful data from the raw data present in the form of specific shape or texture [4].

Many mainstream image processing and retrieval topics come under the scope of research and development part of CBIR.

Few important issues are [1]:

✓ Understanding image users’ needs and information-seeking behavior.
✓ Identification of suitable ways of describing image content.
✓ Extracting such features from raw images.
✓ Providing compact storage for large image databases.
✓ Matching query and stored images in a way that reflects human similarity judgments.
✓ Efficiently accessing stored images by content.
✓ Providing functional human interfaces to CBIR systems. Content based image retrieval is applied in many fields like computer vision and image processing.

It is a myth that CBIR is a subset of image processing and computer vision, but it differs from image processing and computer vision. CBIR is a technique used for retrieving an image from a large image database based on some of the desired characteristics. An image processing covers a broad area such as compression, restoration, image enhancement and transmission.
For example, if the police forces want to recognize some faces for suspects, they may compare the image individually with each single image in the database to verify the identity. In this scenario, there will be a comparison between two images at a time. If they use the database to find the nearest similar images, this will be a typical application of CBIR apply the contents of an image to identify and retrieve. CBIR systems extract features based on the value of the pixels (image) from the database. The features are lesser than the image size and stored in feature database. The theatre database holds a dense form of the image in the database (image). Every image is symbolized by a dense representation of its contents (texture, shape, spatial information and color) in the form of a real-valued and multi-component feature vectors or signature and called off-line feature retrieval [5].

When a query is submitted in the CBIR system, the system automatically extracts the content feature of that image as it does in the image database.

The feature vector of the query image and the one stored in the feature database is computed and the distance or similarity between them is recognized. The system will retrieve the best identical images according to their similar values and called on-line image retrieval [5].

The benefit of Content Based Image Retrieval System is that it utilizes image features in its place of using the image itself. So, content based image retrieval is cheap, fast and efficient then another image search methods.

1.2 Importance of CBIR (Content Based Image Retrieval)

If we use some text and keyword, that explain the image to search and retrieve it from a given image data base. This method is not a good way for image retrieving. The reason is that in this method we must have a detailed description of every image and then the words we used to search, must match. Unfortunately, the image databases are huge in size and it is very difficult to describe every image in the database with a detailed and complete description. In this case when we will retrieve the images, the system will miss some images and will retrieve images that don't queried. So, we need to find a technique to retrieve images depending on its content, not on its description [2]. To solve the problem we will use the content of the image to search and retrieve it rather than searching for an image using text. A CBIR system processes the detail information stored in the image data set and creates an abstraction of its content in terms of visual characterization. These characteristics are texture, color, and shape. Any query process deal with this abstraction instead of with the image itself. Here
every image put in the image database is examined and a dense representation of its content is stored in the form of signature or feature vector [5]. To get desired image, the query image must evaluate with other images in the database for similarity. The similarity assessment uses the image representation. Representation of an image contains extracted features. Features extracted from an image can be shape, color or texture. The similarity is to analyze the difference between two images’ features. From this point of view Content based image retrieval has several advantages compared to other approaches.

1.3 Application scenarios of CBIR
A broad range of potential application of CBIR technology has been identified. Some potential areas are as under.

1.3.1 Crime Prevention & Law Enforcement
Many law enforcement agencies maintain large archives of image evidences, including fingerprint, suspects facial photographs, shoe prints and tire treads, etc. Investigating a criminal incidence, they may compare evidences from the scene of the crime with previous records stored in their archives for similarities [118].

Content based image retrieval has been useful for many reasons in crime prevention and law enforcement, such as Face recognition and fingerprint recognition [25], DNA matching, surveillance systems and shoe sole impressions. WWW image retrieval can be applied in additional fields of law enforcement. Many criminal uses the internet as a way of promoting their illegitimate goods, illegal weapons, drugs, etc. it has been found that many such websites contain less text and more visuals & image. Relying entirely on text retrieval technique is often ineffective in identifying them. Using both text based image retrieval and content based is thus much vital in order to locate such web sites. Other types of unlawful and socially harmful operations include the use of web sites to call for Nazism, violence and racism. Tools for content query based image retrieval from the web helps police to locate these websites in fighting such operation [109].

1.3.2 The Military
The application of imaging technology in the armed forces is probably the best developed utilized but least publicized. The example includes Recognition of identification of targets form satellite photographs, provision of guidance system for cruise missiles; enemy aircraft from radar screens are known examples. Many of the surveillance techniques are relevant to the military field [109].
1.3.3 Intellectual property

One of the major applications of CBIR is Trademark image registration. Here a new trademark marks are compared with existing trademarks to ensure dissimilarity among them. Copyright protection is also a potentially important application area[118,121].

Image copyright protection one of the application of web image retrieval to intellectual property. Image copyright protection on the web is very important, in particular because unlawful copies of images can easily be broadcast over the internet. A watermark is one of the existing techniques for image copyright. This involves adding the distributor or creators identify on each image. But using a large variety of watermark method increase detection time exponentially. Content based image retrieval (CBIR) for the web can be recommended as an excellent alternative for solving this kind of trouble. For an example, in such cases where we want to find out whether the data has been illegally used on the internet, we can use CBIR to identify it. The application of a CBIR system designed to identify replicated images on the World Wide Web is also one of the examples.

1.3.4 Architectural and Engineering Design

Architectural and engineering design contributes a number of general features. The capacity to search design collection for previous cases which are in some sort of similar way or meet specified appropriate criteria may be valuable[112].

1.3.5 Fashion, Architecture and Engineering Design

Similarity can also be honored in the design process in other subject areas, including fashion and home design. The capability to search a collection of fabric to find a peculiar combination of texture or color is increasingly being acknowledged as a useful aid in the development process [107].

Fashion and industrial designer reprocess images of previous blueprints as a source of previous plans as an origin of motivation. For the similar basis, engineers and architects may take to understand designs, drawing image of machines and new projects. With the dawn of the web, these experts are no longer limited to their confined collections in their search for inspiration. The web (internet) allows them to access further images that may demonstrate similar or diverse styles, providing better-off inspiration for their imagination. Web (internet) image retrieval systems and surfing catalogs are so useful in all of these areas [113].
1.3.6 Advertising & Journalism
Both stock shot agencies and newspapers maintain archives of still image to illustrate an advertising or copy article. These image archives can often be millions of images and impressively very costly to maintain if detailed keyword indexing is confirmed. Broadcasting companies are faced with an even larger problem, having to deal with millions of hours of archival image, video footage, which are approximately impossible to annotate without some degree of automatic support. Advertising & Journalism area is probably one of the major users of CBIR technology [122].

1.3.7 Medical Diagnosis
The increasing confidence of recent medicine on diagnostic techniques like automated tomography, radiology and histopathology has resulted in an unexpected increase in the number and significance of medical images now stored by nearly all hospitals. While the major necessity for medical imaging systems is to be competent to display images concerning to a named patient there is growing significance in the use of CBIR technique to support the diagnosis by recognizing similar previous cases [26].

1.3.8 Remote sensing & Geographical Information Systems (GIS)
Although not strictly a case of image retrieval managers responsible for marketing, planning and distribution in big companies needs to be able to search by spatial characteristic (e.g. to search the 11 retail outlets nearby to a given storehouse) [27]. And the armed forces are not the only group interested in investigating satellite images. Physical geographers and agriculturalists use such images extensively, both in practice and research designs, such as identifying areas where crops are lacking in nutrients or diseased or alerting governments to farmers budding crops on territory they have been paid to leave lying fallow[122].

1.3.9 Art & Historical Research
Art galleries deal with intrinsically image objects. The capability to recognize objects sharing some portion of visual similarity can be helpful both to researchers trying to trace historical influences and to art lovers seeking for further examples of sculptures or painting interesting to their taste [28,122]. History sociologists and archeologists use image data as a resource of information or to help in their research. When access to the novel work of artistic creation is confined (e.g. Due to ownership restrictions or geographic distance, the physical condition) researchers can use surrogates which may be set up on the web in the form of images
of the object to ward off this trouble. A web (internet) based image retrieval engine can save their time in searching such material [109].

1.3.10 Training & Education
It is very often complex to identify excellent teaching material to demonstrate key points in a lecture or in self-study module. The accessibility of searchable collections of image and video clips providing examples of (say) avalanches for a lecture on traffic jams for a course on town planning could reduce training time and lead to enhanced teaching excellence. In some cases like complex diagnostic and repair procedures such images and videos might even substitute a human instructor [100].

Training and Education are among the most significant applications of image retrieval in general and of image retrieval from the Web (internet) in particular. The scholar is regularly asked to do research on a particular area. They require images for at least two reasons: as a resource of information or to demonstrate their ideas. These images and videos can also be exercised by teachers in preparing their subjects, providing them with teaching notes, assignment to demonstrate and explain their ideas. Many such images and video are available on the Web (internet) for free access, but locating them is sometimes very lengthy. Having an image (web) search engines and navigation catalogs can give this procedure much easier going [109].

1.3.11 Home Entertainment
Home entertainment includes an image or video-based, home videos and scenes from favorite films or TV programmers and holiday snapshots. These are the areas where a market for CBIR technology could develop [122].

Many people use the web for home entertainment purpose. Visual information used for entertainment includes jokes, images, comic strips movies shots, music clips, and caricaturist. However, people often have trouble finding the images or visual media they are seeking for. Web video and image retrieval tools for such types of data appear indispensable in assisting web users locate the sought images [23,24]. In addition, if the visual image or video content of the web is categorized in an index structure that users can use for browsing retrieved to such resources will be easier and navigated.

1.3.12 Web Searching
The well advertises difficulty of locating images on the web indicates that there is a clear need for image search tools of similar power. Many prototype systems for
content based image searching on the web have been exhibited over the last three to five years [109,122].

1.3.13 Filtering of adult content on the web

Now in these days a large number of internet websites containing pornographic images and video and those are available for free to download. There is a great concern of parents that is accessing of such videos and images of children. In addition to this, in some culture and society, such kind of images is not tolerated even for adult people. For this it's necessary to get into appropriate mature content filtering tools for the WWW. Previous solutions include adult content free websites, which try to stop children from accessing unwanted sites, and software filters such as Cyber Patrol, CyberSiter and NetNanny. Such instruments have proven ineffective because they spread over only a very little portion of the network. Furthermore, even though the fact that pornography sites often contains many images, video and short text, such tools, verify only the text or the IP address of the website. In current research work, researchers have discovered the use of computer graphics & vision techniques to automatically recognize pornography images and video as well as text. Fleck and Forsyth use an algorithm that seeks to identify nude people in the images. This algorithm uses a skin filter based on color and texture. Chang et al developed a new similar system that recognizes images of nude humans by limb shape and skin tone. Integrating such as tools in web image retrieval can help significantly in identifying doubtful websites and thus with filtering them. This can be done by proposing appropriate model queries to a query based retrieval system. That tries to recognize all web pages that keep up a correspondence with them. Wang et al developed a system called IBCOW that categorizes websites into objectionable and gentle base on image content [122,107].

1.3.14 Travel and Tourism

Prior to visiting a new tourist place, people always would like to know more about it. The data required includes maps of the country they will visit, as well as street and city maps. Which can give them an idea about essential conveniences, transportation networks, tourist attractions and monuments? A tourist may also concern with having a basic idea regarding the country’s way of life, characteristic, architecture such as its markets and traditional dress. How he can get such visual information with lesser time and effort? The World Wide Web is full of such information that is visually too, but
the same time it is scattered and needs automatic retrieval to help tourists and travelers. In addition to, categorized images which are available on the Web in a browsing structure can help tourists reach the sought information without much difficulty [95,107].

1.4 Existing Image Indexing System

Phillip & Nickolay and Washington [11,9] in their research paper mentioned regarding the use of algorithms to extract shape feature from an image. Phillip & Nickolay while working on medical image retrieval extracted shape feature. As an index in CBIR. Washington Mio used a Bayesian approach in research for recognition of an object. Phillip & Nickolay were focused on extracting a shape from the vertebrae and shell, for this they developed segmentation of image and convex hull models. Phillip & Nickolay [11] proposed the following formula to extract convex hull of the region.

\[ r_{j+1} = r_j + \delta i \alpha N_i p_j + V_j \] \hfill (1.4)

We may observe here that they constructed the process of extracting up to inserted for retrieval purpose.
In CBIR we use shape and color to retrieve images from the image database. To find out the feature similarity, Hore applied Euclidean distance measure. It is illustrated as follows [9]:

$$D_e(p, q) = \sqrt{\sum_{i=1}^{n}(Q_i - P_i)^2}$$

......................... (1)

Where,

$q =$ Query image.
$p =$ Database image.
$P_i =$ Database $i^{th}$ sample.
$n =$ Number of samples for this feature.
$Q_i =$ Query features $i^{th}$ sample.

He used a weight calculation method for the feature being extracted. The basic idea was to give more weight to the feature that uniquely or strongly represent the image.

Her researched on four formulas and find out the following one that is seen as the one which give better results in compared to other three formula:

![Figure 1.1 Query and construction procedures of content indexed image database [11]](image)
Andrew [10] suggested inserting the indexing keys inside the JPEG compression images for image retrieval. The main thought behind it is to minimize the storage space.

In his research, Wang [8] introduced Wavelet-Based Image Indexing and searching a new type of image indexing. In this image indexing method the algorithm partially sketches image for retrieval. He applied Daubechies' wavelets to develop multi-scale indexing & matching system [20].

Now we can further improve the algorithm to increase the searching accuracy and capability. E.g. while calculating the distance between two images we may adjust weight and use perceptually-comparable color space for different wavelet coefficients. Wang also used his preceding finding to enhance image retrieval As an image retrieval system, Wang proposed Semantics sensitive Integrated Matching for Picture Libraries (SIMPLIcity).

That applies a wavelet based approach (semantics classification techniques) for integrated region matching. That was based on image segmentation and feature extraction [21, 22]. The idea of classified images into semantic classes can reduce the undesirable result of

1. Incorrect segmentation.
2. Helps to clear up the semantics of particular parts.
3. Simpler querying interface for retrieval [8].

1.5 Existing CBIR Systems

Early CBIR systems as VisualSEEk (Smith and Chang, 1996) [14] and QBIC (Flickner et al., 1995) [13] were based on image colors represented by a kind of color histogram, which totally ignored structures of materials and object surfaces present in the scene. Visual appearances of such structured surfaces are commonly referred as textures and their characterization is essential for understanding of real scene images.

Later systems attempted to include some textural description, e.g. based on wavelets as CULE (Chen et al., 2005), IBM Video Retrieval System (Amir et al., 2005) or Gabor features as MediaMill (Snoek et al., 2008). MUFIN (Batko et al., 2010), which is focused on efficiency and scalability, includes a simple texture representation by
MPEG-7 descriptors. A CBIR system img (Anaktisi) (Chatzichristos et al., 2010) is aimed at a compact representation, which was extracted by fuzzy techniques applied to color features and wavelet based texture description. However, texture representations in these systems are more or less supplemental and the algorithms rely on color features. Although retrieval results look promising, they are often provided by enormous image databases than exact image indexing. It is quite simple to the entire first result page with very similar images from a large database (e.g. Sunsets, beaches, etc.), nevertheless, the lack of image understanding is revealed on further result pages.

In narrow image domains, CBIR systems are more successful e.g. trademark retrieval (Leung and Chen, 2002; Wei et al., 2009; Phan and Androutsos, 2010), drug pill retrieval (Lee et al., 2010) or face detection (Lew and Huijsmans, 1996) and similarity, which evolved into a separate field.

One of the reasons of disregarding textural features are that they are still immature for a reliable representation (Deselaers et al., 2008) and at least weak texture segmentation of images is required (Smeulders et al., 2000). If the segmentation is extracted, shape features and region relations can be employed (Datta et al., 2008), however, the reliable segmentation is a difficult problem on its own. Recent methods avoid the image segmentation by local descriptors as SIFT (Lowe, 2004), which were extended to color images and used for image indexing (van de Sande et al., 2010; Burghouts and Geusebroek, 2009a; Bosch et al., 2008). However, these key points based descriptors are more suitable for the description of objects without large textured faces than homogeneous texture areas.

The other reason for marginalizing textures is that a more precise description of textures also requires more attention to expected variations of acquisition conditions. Many existing systems do not care about such variations or they handle it in a very limited way. Recently, Shotton et al. (2009) demonstrated that textural features can be Successfully used for image understanding, if the variation of acquisition circumstances is considered.

1.6 Motivation for CBIR

As a result of recent advancements in digital storage technology, it is now possible to create large and extensive databases of digital imagery. These collections may contain millions of images and terabytes of data. For users to make the most of these databases effectively, efficient methods of searching must be devised. Prior to
automated indexing methods, image databases were indexed according to keywords that were both decided upon and entered by a human categorizer. Unfortunately, this practice comes with two very severe shortcomings. First, as a database becomes increasingly large the manpower required to index each image becomes less practical. Secondly, two different people, or even the same person, on two different days, may index similar images inconsistently. The result of these inefficiencies is a less than optimal search result for the end user of the system.

Having a computer does the indexing base on a CBIR scheme attempts to address the shortcomings of human-based indexing. Since a computer can process images at a much higher rate, while never tiring, the manpower issue is solved. Additionally, as long as the algorithms used in the indexing procedure are kept constant, all images will be indexed consistently, solving the inherent problems resulting from fallible human-based indexing.

It should also be noted that current CBIR methodologies are not without their limitations. For example, each CBIR system needs to be tuned for its particular use in order to give optimal results. A retrieval system designed for querying medical x-ray images will more than likely prove to be a poor system for retrieving satellite images of South American rain forests. In addition, presently employed algorithms cannot yet consistently extract abstract features of images, such as an emotional response, that would be relatively easy for a human to observe. For example, it would be rather difficult for current methods to classify images of heart shaped objects as well as two figures holding hands under a general category of “love” or “affection”.

1.7 Problem Definition

In this research work, the problem is to find out the best features or characteristics of an image as an index and combining those features for successful retrieval.

From the previous researches about image features for indexing. We select several features that can be considered suitable and can be implemented as given in figure. Those features are, as follows:

**Global feature** – Boolean Edge Density, Edge Density, Color Sigma, Edge Direction, Color Average.

**Region feature** – Moment Invariant, Grey Level, Region Area.
However the problem lies in combining features to make an effective index for each and every image.

There are still few question say

1. If we combine two or more different methods will it be more effective than using each method separately.
2. Can be improved retrieval by combining region and global feature.

In the past, there are many studies those are based on Tree structure and their use image features.

We required a tree structure to handle the image data structure features that should be suitable as an index. Same time for creating the index we should constraints on algorithm and tree structure. The study shows that a detailed and precise query requires to search through the least number of indexing entries.

For retrieval process, it is very difficult to match the features for similarity.
We need similarity method that should be easy to use and quick to compute. The comparison of the images, then considers the features similarity. The target images the query image are evaluated on the bases of their features.

1.8 Proposed solution

The aim of this research work is creating an image indexing system by identifying and explaining image features.

In this research work we are developing an image indexing algorithm. This algorithm based on the study of various image features of computation and their utilities. In the research work we study the different combination of image feature for successful retrievals of images.

In this research, we focus on selection of feature rather than the tree structure. To handle our data structure we use R-Tree for image indexing. We also test the image indexing by putting some queries on the indexed database for retrieval.

There are still few question say

1. If we combine two or more different methods it be more effective than using each method separately?
2. Can be improve retrieval by combining region and global feature?

To solve the problem, we develop a data structure which consist two parts

a. Global data structure consisting global features
b. Region data structures consisting of regions data.

The global data structure consists of Boolean Edge Density, Edge Density, Color Sigma, Edge Direction, and Color Average. Whereas the region data structure consists of Moment Invariant, Grey Level, Region Area.

Each image includes global feature and region feature. This shows each region data and global data are pointing to each other. R tree will store all region data structures. A rectangle will represent each region data structure in the R tree and used at the searching stage. We use 600 x 600 dimensions as a linear normalization size to overcome to possibilities of different sizes of images.
1.9 Summary of contributions
The contributions of this thesis in improving systems for retrieving, analyzing, and compressing images are as follows:

1. We have developed an indexing system that reduces the time to perform image retrieval and same time cut down the usage of storage space.
2. We have created a new indexing model using a threshold of original images, splitting the process into two parts i.e. Region part and Global part.
3. We have used a new approach to insert all region data structure that points to the corresponding global data structures into R-Tree.
4. In this research work we have investigated multiple combinations of available features and find out the best combination of the match and the new approach is shows that region features are much better than global.
5. Most of the system uses a grey scale image. If the image database or the query image is a color image they convert it to grey scale image, however, some system uses the color of the image as its feature they derive the histogram from the color space of the image. In this research work, we use color average, color sigma, edge density. Color average measure the average number of color layers in the image. Color sigma calculates the intensity variation in an image. Edge density enhances the pixel that belongs to edge and boundary by using a Sobel operator.
6. When CBIR system is ready for testing, the query image is compared with every target image in the database to find the best matching images. This is the linear search method. Which will take a long time if the database is large? This is not necessary to repeat the comparison between the query image and the database every time. In this research work, we create a 10 set of dataset of images. Each image belongs to one image data set. The system can select some image from different dataset and use them to compare with the query image. That can be done by dividing the image dataset into categories.
7. We are using region indexing that represents the number of pixels in a region. These region indexes, stored in R-Tree.
1.10 Outline of thesis

This thesis is structured as follows. Chapter 1 describes about Content Based Image Retrieval, importance of Content Based Image Retrieval, application scenarios of CBIR, the concept of Existing Image Indexing System, Existing CBIR Systems, and Motivation for CBIR, Problem Definition, Proposed Solution and summary of contributions. Chapter 2 reviews some related work, especially about the research issues like region features and global features and brief introduction of CBIR systems. Chapter 3 explains the methodology adopted to solve the problem. Chapter 4 contains an in-depth analysis and implementation of the system. Chapter 5 gives details of experimental results and discusses the performance of the system. Chapter 6 gives conclusions and future improvements which could be incorporated into the system. In chapter 7 reference and in chapter 8 Appendix-1 provides background and finally list of publication.