An Efficient Indexing Approach for Content based Image Retrieval

Shama P S
Assistant Professor
Dept.of.Computer Science
SJCIT, Chickaballapur

Badrinath K
Assistant Professor
Dept.of.Information Science
SJCIT, Chickaballapur

Anand Tilugul
Assistant Professor
Dept.of.Information Science
SJCIT, Chickaballapur

ABSTRACT

In this paper an efficient indexing and retrieval technique is proposed for identification of plant images. The plant images have been retrieved as herbs shrubs and trees based on color and texture features. Plant images have rich content of information such as leaves, bark and stem etc. This content can be extracted as various content features. Plant image database often represent the image objects as high-dimensional feature vectors and access them via the feature vectors and similarity measures. Hence in order to retrieve plant images an indexing method is proposed based image retrieval to improve the retrieval performance. In this work color and texture features are extracted for characterizing leafy mass and stem images of plants. There are 20 images of five plant species, amounting to total of 100 images, herbs, shrubs, and trees. Hence, totally there are 300 images of leafy mass and bark of 15 plant species. All the images are captured in natural environment using a digital camera. In the pre-processing stage, the images of leafy mass and bark with complex background information and noise is filtered out by using Unsharp filtering method. In the segmentation method 2D-OTSU threshold based segmentation technique is used to separate the object from the background For Ex: small leaves surrounded by bright sky, small openings among dense canopy, and mixed pixels. For feature extraction the Modified Color Co-occurrence Matrix (MCCM) and Gabour Filter is used to retrieve the color and texture of the images from the database. Usually the retrieval performance of an image retrieval system is greatly influenced by different similarities or distance measures. We have used Euclidean distance measures for similarity matching. Finally for retrieval of the images from the database Fast Indexing scheme for CBIR method is used. Then multi-dimensional indexing technique is employed using tree structures, R* tree. The effective indexing and fast searching of images on bases of visual features pose a significant issue in CBIR. We have used R*-Tree structure to achieve better performance and efficiency. A graph is plotted to compare the retrieval result of both technique. Indexing Technique gives better retrieval rate.

Keywords

Content Based Image retrieval(CBIR); Lab color space; color feature; texture feature; gabour feature; color tree image; image segmentation; image retrieval; Databases; search efficiency; Indexing; R*-Tree Structure.

1. INTRODUCTION

Indexing method for content-based image retrieval is proposed to improve the retrieval performance. Images have rich content. This content can be extracted as various content features. Image database often represent the image objects as high-dimensional feature vectors and access them via the feature vectors and similarity measures. Image retrieval is the field of study concerned with searching and browsing Digital Images from database collection. In CBIR systems index images using low-level features. Indexing is an integral part of designing a database system to reduce computational overhead and optimize retrieval.

In recent years, content-based image retrieval (CBIR), which aims at developing techniques that support effective searching and browsing of digital images from large database collection based on automatically derived imagery features. In CBIR systems index images using low-level features. A typical CBIR system for retrieving images from database based on their similarity to a query image consists of four main steps. First, extract the image features to convert the image from space based on pixel to that based on feature. The feature extraction is the basic process of a CBIR system. The system will select appropriate feature spaces and explore various visual features to characterize an image. Next, construct feature vectors based on selected features, by which images in the database are represented. Later on, compare the image feature vectors of a query image (or regions in query image, or many query images) with the target images in database by computing a similarity measure to search the "most similar" images in the database. Finally, output queried image results similar to query image based on the specific ranking method. Some different similarity measure algorithms are designed to search the similar image in the database.

Many approaches have been proposed such as the text- based retrieval and the content based image retrieval (CBIR). Text-based approach consists to attach keywords or labels to each item and then to perform searches based on these labels. CBIR approach extracts low-level features to index image such as color, texture and shape. However, these approaches are inefficient due to the gap between visual features and semantic concepts. However, the features from image data are low-level visual character, which have very limited ability in representing and analyzing the high-level semantic content of the image. This is referred to as the semantic gap between the low-level visual features and high-level semantic content. Several systems are proposed to improve the retrieval quality. Relevant feedback approach was used in text-based information retrieval and was introduced to CBIR to bring user in the retrieval process for reducing the semantic gap between what queries represent (low-level features) and what the user thinks. Indexing and retrieval of plant image database is process in which each individual plant is described by its feature vector and identified in large plant image database. Image database identification system requires a reduced search space for identification process. So that system can be retrieve image from large database in less time.

In this paper two approaches have been characterizing leafy mass and tree bark images of color and texture will be retrieved. The effective indexing and fast searching of images on based on visual features pose a significant issue in CBIR, Commonly a tree structure is utilized to store image information since it has high-dimensional metric space. The

majority of these multi-dimensions, Indexing methods perform significantly well for dimensions. We have used R*-Tree structure to achieve better performance and efficiency.

2. RELATED WORKS

Xiaosong Wang, Xinyuan Huang, was proposed segmentation method to extract object based on color and texture features of color tree images. Firstly, reduce noise of the color tree image by the use of anisotropic filter. Then, we select Lab color space as the space for image segmentation. And the color image of RGB space is transformed into Lab space. Next, due to the negative end of a-channel reflects the color feature of trees, the L, a, and b channels are split. Green is the main feature of tree images, so segmentation by two-dimension OTSU of automatic threshold in a-channel.

Seong-o shim,Tae-sun choi, was proposed with Modified Color Co-occurrence Matrix (MCCM) for image feature. First, CCM was simplified to account the number of certain colored pairs between all possible adjacent pixels in the image. For adjacency, four-connectedness was chosen. In this case, the diagonal matrix of CCM explains the color histogram of pixels that belong to the homogeneous regions. Experiment shows that this diagonal matrix follows the same trends as the color histogram of entire image. Non diagonal elements contain the shape information because the color changes between adjacent pixels imply the possible existence of object edges.

Suresh Pabboju, A. Venugopal Reddy, presented an elegant and effective system for content-based image indexing and retrieval. The system exploits the global and regional features of the images for indexing and fractional distance measure as similarity measure for retrieval. The images are quantized before extracting the global features. We have also presented a novel approach for image segmentation to extract the region features effectively.

Shyi-Chyi Cheng, Tian-Luu Wu, presented a fast indexing scheme for content-based image Retrieval based on the principal axis analysis. Image databases often represent the image Objects as high-dimensional feature vectors and access them via the feature vectors and Similarity measure. A similarity measure similar to the quadratic histogram distance measure is defined for this indexing method. The experimental results demonstrate that the proposed Method outperforms the compared methods in retrieval accuracy and execution speed.

Jiatao Sonl, Zheru Chi, presented a texture based bark classification method is presented. Our method uses two types of texture features, one is the gray-level co-occurrence matrix metrics and the other is the binary texture feature, named Long Connection Length Emphasis (LCLE. This method is easy to implement and has low computational complexity. Many researches also show that this method can achieve better texture analysis performance than other commonly used approaches.

M.V.Sudhamani, Dr.C.R.Venugopal, The new algorithm cluster-based R*-tree indexing is proposed and compared the efficiency with R* -tree and sequential search. The accuracy of the retrieval system is measured and compared the results. Experimental results show the high performance of the proposed method. We have implemented the technique of retrieval of images from databases based on color feature extracted from each region/regions of an image represented by dominant or representative color descriptor.

Suihua Wang, Ailing Zhao, presents a reasoning-based image indexing and retrieval framework which uses content-based image index and retrieval techniques to extract low-level characteristics of images and identify their semantic contents, in which the low-level image features and the complex objects could be linked together.

Content-based image retrieval techniques rely on the color, shape, texture and other low-level characteristics of image to index and retrieve them.

Mohamad Obeid, Bruno Jedynak, presented visual information retrieval systems use low-level features such as color, texture and shape for image queries. Using low-level features to correspond to high-level abstractions is one aspect of the semantic gap. We introduce intermediate features. These are low level "semantic features" and "high level image" features. That is, in one hand, they can be arranged to produce high level concept and in another hand, they can be learned from a small annotated database. These features can then be used in an image retrieval system.

Kian-Lee Tan, proposed a color-spatial retrieval technique, the color information is integrated with the knowledge of the colors' spatial distribution to facilitate content-based image retrieval. Content based image retrieval system that accesses images based on the content of the image such as color, texture and shape. In a content-based image retrieval system, the content of the database images are (semi)automatically generated and stored. A query image's content is extracted during runtime and used to match against those in the database. The result of the query is a set of images that are similar to the query image, rather than an exact match.

Jongan Park, we proposed color histogram for different spatial resolutions.. The distance between different histograms of the query image with the corresponding histograms of database images are calculated. The final distance is then weighted sum of the distances of different resolution histograms. The proposed algorithm is tested on a database comprising a large number of images and compared the results with the single resolution color histogram.

A.Drimbarean, P.F.Whelan, focus on the classification of colour texture images. The main objective is to determine the contribution of colour information to the overall classification performance. The evaluation criteria are the classification accuracy using a neural network classifier based on learning vector quantization. Experimental results indicate that the incorporation of color information enhances the performance of the texture analysis techniques examined.

Christoph Palm, Thomas M. Lehmann, Proposes a novel approach to Gabor filtering of color textures is introduced. Additionally, a novel Gabor texture feature for the grayscale as well as the color domain is proposed. The color features show significantly better results than the grayscale features. The RGB based features, where inherently color and intensity work together, performed best. Especially the local phase change measures supplement the known amplitude measure appropriately.

Sang Hyun Bae, Dae Young Lee, proposed feature-based similarity retrieval has become an important research issue in multimedia database systems. The performance of conventional multidimensional data structures (e g; R-Tree family K-D-B tree, grid file, TV-tree) tends to deteriorate as the number of dimensions of feature vectors increases. We experimentally compare the retrieval time cost of a SOM-based R*-tree with that of an SOM and an R*-tree using color

feature vectors extracted from 40,000 images. The results show that the SOM-based R*-tree outperforms both the SOM and R*-tree due to the reduction of the number of nodes required to build R*-tree and retrieval time costs.

The main methodology used in this project is retrieve and Indexing the leafy mass and Bark images based on color and texture feature extraction. The fig.1 gives the proposed model with different modules used in this project. At the starting input image is filtered using unsharp filtering method, and then apply segmentation using OTSU method, the texture feature extraction module using Gabour filter and MCCM Method. Similarity between images is computed. Finally image indexing method apply on clustered image database using R* Tree.

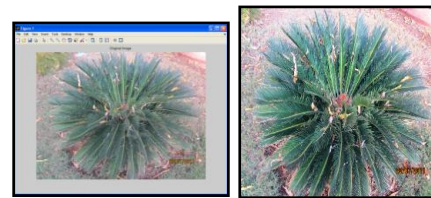# 3. CONTENT BASED IMAGE RETRIEVAL METHOD

The problem of searching identical images from large image repositories on basis of their visual contents is called Content-Based Image Retrieval. The term 'content' in CBIR refers to colors, shapes, textures, or any other information that can be possibly obtained from the image itself and 'Content Based' denotes that the search will consider the concrete contents of the image. Indexing remarkably affects the speed of data access besides supporting the accuracy for retrieval process and thus is a significant factor in image database systems.

3.1 Pre-processing Method

In the image processing and pattern recognition system, the input original image must be pre-conducted. Because the plant in the image which is collected in nature by digital equipment

chronically exposure to the natural world, there are all kinds of noise leaving in the crown, trunk, and the background. Before the image segmentation, we must remove the noise in the plant image. First of all is to deal with plant image de-noising. In the image demising, how to retain the important feature of the image (edge, texture, and line, etc.) has been a focus problem in image processing work. The selection of filter is very important, because it is determine whether as much as possible to keep the plants edge and the texture feature of the canopy, and at the same time it can de-noising fast.

In the first pre-processing stage, the images of leafy mass and bark from the camera having a lot of background information and noise that will be first filtered out by using unsharp filtering method. The "unsharp" of the name derives from the fact that the technique uses a blurred, or "unsharp," positive to create a "mask" of the original image. The unsharped mask is then combined with the negative, creating the illusion that the resulting image is sharper than the original. An unsharp filter is an operator used to sharpen images body.

(a) Input Image   (b) Sharpened Image

Fig 2: Preprocessing results for leafy mass image
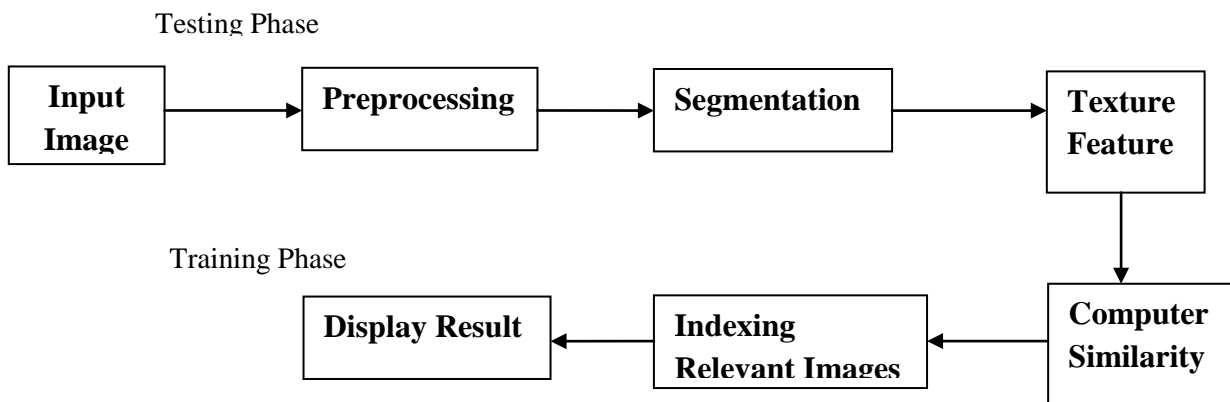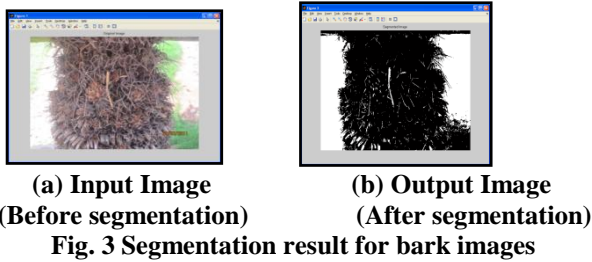
The System Architecture

Fig 1: Block diagram for Content Based Image Retrieval using Index

3.2 Segmentation Method

Image segmentation is one of the most important preliminary steps in CBIR Systems. "Performance of both the shape and layout features depends on good" segmentation. Segmentation refers to the process of dividing the digital image into multiple segments or a set of pixel it is also sometimes known as super pixels. The need of segmentation - is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries, lines and edges in images. Precisely image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The ultimate goal of the segmentation is to identify the semantically meaningful components in the image. Such

components can be used for higher-order representation of the image. Widely used image segmentation methods include threshold methods, feature space clustering, fuzzy clustering, edge-based methods, region growing methods, morphological methods, and graph theoretic methods. Unfortunately, there is not a generic method that works well for all type of images.

OTSU method has long been considered the best practice of threshold auto-selection methods. This method is simple in calculation. And it is widely applied in real-time image processing system. When we segment a-channel with threshold method, this study chooses two-dimension OTSU algorithm which automatically selects the optimal threshold for segmentation. Because two-dimension OTSU algorithm not only takes into account the grayscale information of pixels, but also considers the space-related information of

pixels and their neighborhoods. Typically, an object pixel is given a value of "1" while a background pixel is given a value of "0." Finally, a binary image is created by coloring each pixel white or black, depending on a pixel's labels.

(a) Input Image   **(b) Output Image**
(Before segmentation)   **(After segmentation)**
Fig. 3 Segmentation result for bark images

3.3 Texture Feature Extraction using Gabor Filter and MCCM method

Feature extraction is transforming the input data into the set of features still describing the data with sufficient accuracy. In pattern recognition and image processing, feature extraction is a special form of dimensionality reduction. Feature extraction is used when the input data to an algorithm is too large to be processed and it is suspected to be redundant (much data, but not much information). To extract relevant features from the images resulting in quantitative information for each object. The approach employed in this study was to use statistical texture analysis to determine a set of measures to provide object-background as well as object-object discrimination[2].

3.3.1 Global Filter

Gabor filter is one of the most popular signal processing based approach for texture extraction. Basically, Gabor filters are a group of wavelets, with each wavelet capturing energy at a specific frequency and specific direction. It acts as a local band pass filter with certain optimal joint localization properties in both the spatial domain and the frequency domain. An image is filtered with a set of Gabor filters with different preferred orientations and spatial frequencies and the features, which are obtained from a feature vector, is used further.

Texture features are found by calculating the mean and X Variation of the Gabor filtered image. For a given image I(x, y) with size PQ, its discrete Gabor wavelet transform is given by a convolution:

Where s and t are filter mask size variables and is the complex conjugate of which is a class of self-similar functions generated from dilation and rotation of the following mother wavelet:

Where W is the modulation frequency. The self similar Gabor wavelets are obtained through the generating function:

 Where m and n specify scale and orientation of the wavelet respectively with m = 0, 1…M-1 and n =0, 1…N-1 and

The variables in the above equation are defined as follows:
 In our implementation, we used the following constants
 Ui = 0.05, Uh = 0.4, s and t range from 0 to 60.
After applying Gabor filters on the image with different orientation and at different scale, we obtain an array of magnitudes:

These magnitudes represent the energy content at different scale and orientation of the image. Texture feature are found

by calculating the mean ìmn and standard deviation ómn of the energy magnitude

$$\mu_{mn} = \frac{E(m,n)}{P \times Q}$$

$$\sigma_{mn} = \frac{\sqrt{\sum_x \sum_y \left(G_{mn}(x,y) | - \mu_{mn} \right)^2}}{P \times Q}$$

3.3.2 Modified Color Co-occurrence Matrix (MCCM)

Another approach is to extract texture features from the image is MCCM was proposed as an image feature. First CCM was simplified to account the number of certain colored pairs between all possible adjacent pixels in the image. For adjacency, four – connectedness was chosen. In this case, the diagonal matrix of CCM explains the color histogram of pixels that belong to the homogeneous regions.

We are calculated the following texture feature for image extraction using MCCM technique:
1. Mean: It is a measure of brightness,, where is pixel at location(r, s).

$$mean = \mu_p = \frac{1}{n^2} \sum_{r=0}^{n-1} \sum_{s=0}^{n-1} p_{r,s}$$

2. Standard deviation: It is a measure of contrast

$$s\tan dard$$

$$deviation = \sigma_p = \left[\frac{1}{n^2} \sum_{r=0}^{n-1} \sum_{s=0}^{n-1} [p_{r,s} - \mu]_p^2 \right]^{\frac{1}{2}}$$

3. Entropy: It is a measure of randomness and complexity of the image. Complex textures tend to have higher entropy

$$energy = e = -\sum_i \sum_j p(i,j) \log p(i,j)$$

4. Energy: Energy is a measure of textural uniformity of an image. Energy reaches its highest value when gray level distribution has either a constant or a periodic form.

$$energy = E = \sum_i \sum_j (p(i,j))^2$$

5. Contrast: Contrast is a difference moment of the *P* and it measures the amount of local variations in an image.

$$contrast = c = \sum_i \sum_j (i-j)^2 p(i,j)$$

This measure provides evidence of how sharp the structural variations in the image are.

6. Angular second moment: The angular second moment gives a strong measure of uniformity. It is a measure of the homogeneity of an image and can be defined as

$$angular \quad second$$

$$moment = \sum_i \sum_j \{p(i,j)\}^2$$

7. Correlation: The correlation feature is a measure of gray-level linear dependency of the image. Correlation feature is defines as

$$correlation = \frac{\sum_i \sum_j (ij)\, p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

Where, and are the means and standard deviations of

8. Maximum Probability:

$$maximum \qquad probability = \max_{i,j}(i,j)$$

9. Inverse Difference moment: Inverse difference moment is the measure of local homogeneity and is defined as

$$Inverse \qquad difference$$

$$moment = \sum_i \sum_j \frac{1}{1+\left(i-j\right)^2}\, p(i,j)$$

3.3.3 Color Space

Each color space has its own appear background and application region. When segmenting a color image, the selection of color space plays a decisive role on the segmentation results. The commonly color spaces used in color image processing include RGB color space, HIS color space, Lab color space, and so on. At present, the general color digital images are RGB format. RGB color space is based on the theory of three-basic color to build. RGB format is the most basic color space. Other color space models can be obtained through the RGB format conversion. But the RGB color space is not a homogeneous visual perception space, it is not conducive to image segmentation based on color feature. HSI color space uses color characteristics of a direct sense of the three quantities: the brightness or lightness (I), hue (H), saturation (S) to describe the color. This method is more in line with the human eye habits to the description of the color, but the expressed colors are incomplete visual perceived color. The Lab color space is a homogeneous space for visual perception, the difference between two points in the Lab color space is same with the human visual system[1]. It applies to represent and calculate of all light color or object color, so the Lab color space is very effective in color image analysis.

Color: One of the most important features that make possible the recognition of images by humans is color. Color is a property that depends on the reflection of light to the eye and the processing of that information in the brain.

Color Descriptors

There are many different ways of extracting the color information from an image, some of the basic ones are color histograms and color moments.

Color Spaces:

1. RGB: this color space comes from an additive model in which the three primitive colors red/green/blue are added together to reproduce the all range of colors. Similar in a way to the HVS it is widely used and present in all CRT monitors.

2. HSV: stands for Hue, Saturation, Value, it is a cylindrical-coordinate representation and is known for its inductivity i.e. close to how a person would describe a color. HSV color space it is a popular choice for manipulating color. The HSV color space, representing hue, saturation and color value (brightness) has the shape of a hexagonal cone. The angle is given by the hue, the distance from the centre of the cone by the saturation and the vertical position by the value.

3.3.4 Realization of the Color Channel Separation

The principle of color sub-channel method is to decompose color space components of color image into multiple single-channel images in color space. The color space can choose color space RGB, HSI, Lab or other color space. The single-channel image information after decomposition can be described as the gray-scale image.

4. SIMILARITY/DISTANCE MEASURE

Similarity measure is one of the key items in the process of image retrieval that decides the effectiveness and the efficiency of the retrieval technique. Content-based image retrieval calculates visual similarities between a query image and images in a database. Accordingly, the Retrieval result is not a single image but a list of images ranked by their similarities. Different similarity/distance measures will affect retrieval performances of an image retrieval system significantly. The visual similarities between a query image and images in an image database are determined as an alternative to exact image matching in case of content based image retrieval. Consequently a list of images ranked in order of their resemblance with the query image is enlisted as a result of retrieval. Lately, numerous similarity measures have been developed for image retrieval that works on basis of approximations of the distribution of features. The retrieval performance of an image retrieval system is greatly influenced by different similarities or distance measures.

Similarity measure for texture: Euclidean Distance Algorithm:

1. Decompose query image.
2. Get the energies of the first dominant k low frequency components.
3. For image i in the database obtain the k energies.
4. Calculate the Euclidean distance between the two sets of energies, using

Increment i. Repeat from step 3**:** Using the above algorithm, the query image is searched for in the image database. The Euclidean distance is calculated between the query image and every image in the database.

5. INDEXING AND RETREVIAL USING R* TREE STRUCTURE

R*-trees are a variant of R-tree used for indexing spatial information. R*-trees support point and spatial data at the same time with a slightly higher cost than other R-trees[5].

An R*-Tree satisfies the following properties

- ⚔ The R*-tree is a balanced disk-based tree structure.
- ⚔ Root has at least two children unless it is a leaf.
- ⚔ Spatial objects are clustered based on the proximity of their locations.
- ⚔ Each sub-tree is bounded by the Minimum Bounding Rectangle (MBR) of all objects in it.
- ⚔ Each node is a disk page and has >= m (min # of entries) entries.
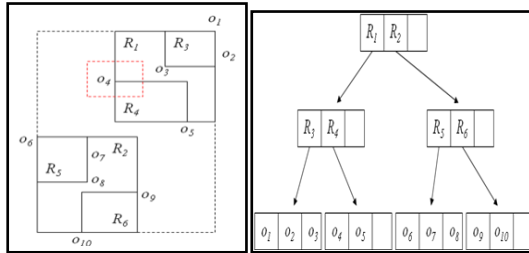- ⚔ Efficiency metric = number of disk-pages accessed


Fig. 4 Structure of R*-Tree

First of all, the R*-tree clearly outperforms the R-tree variants in all experiments Moreover the most popular variant, the linear R-tree, performs essentially worse than all other R-trees the following remarks emphasize the superiority of the

R*-tree in comparison to the R-trees

- The R*-tree is the most robust method which is underlined by the fact that for every query file and every data file less disk accesses are required than by any other variants To say It m other words, there is no experiment where the R*-tree is not the winner[8].
- The gain in efficiency of the R*-tree for smaller query rectangles is higher than for larger query rectangles, because storage utilization gets more important for larger query rectangles This emphasizes the goodness of the order preservation of the R*-tree (ie rectangles close to each other are more likely stored together in one page).
- Surprisingly in spite of using the concept of Forced Reinsert, the average insertion cost is not increased, but essentially decreased regarding the R-tree variants.
- Minimization of area, margin, and overlap is crucial to the performance of R-tree / R*-tree.
- The R*-tree attempts to reduce the tree, using a combination of a revised node split algorithm and the concept of forced reinsertion at node overflow. This is based on the observation that R-tree structures are highly susceptible to the order in which their entries are inserted, so an insertion-built (rather than bulk-loaded) structure is likely to be sub-optimal. Deletion and reinsertion of entries allows them to "find" a place in the tree that may be more appropriate than their original location. à Improve retrieval performance.

6. EXPERIMENTAL RESULT

Experimental results show segmentation of plant images, then classify images into bark and leafy mass plants, generate R*-tree Structure for given query image and finally result shows retrieval results of query image.
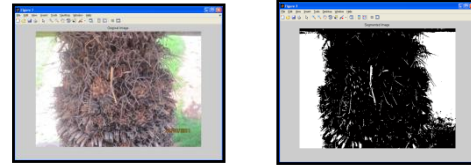
(a) Original Image   **(b) Segmented Image**

Fig.5 Clustering and indexing of the Cycas sps bark Original and Segmented plant image
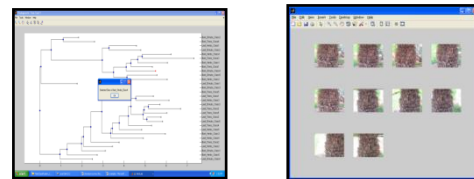
(a) R* Tree Generation   **(b) Retrieval Results**
Fig.6 Retrieval results of R* tree generation and retrieval results

6.1 Evaluation of Performance

Given a collection of documents, a set of queries and human expert's responses to the above queries, the ideal system will retrieve exactly what the human dictated. Performance in terms of precision and recall is calculated. The most popular way to evaluate the performance of a retrieval system is to calculate the percentage of relevant documents retrieved and also their relative order. Ideally, a system should retrieve all the relevant documents first, keeping the number of non-relevant documents that are retrieved before the relevant ones as minimum as possible.

Precision and Recall are the most widely used retrieval performance measurement in literature. It is based on categorical matching. The precision P and recall R are then defined as Recall is the percentage of the total relevant documents retrieved and is defined as:

Precision refers to the capability of the system to retrieve only the relevant documents. Precision can be expressed as:

For each query, the precision of the retrieval at each level of the recall is obtained. The experimental results show that the proposed system can improve the retrieval accuracy as well as reduce the time for retrieval.

A set of recall and precision curves are joined together and this curve is called the precision-recall curve. Recall is inversely proportional to precision. For the sake of comparison, we prefer a non-increasing precision-recall curve and hence we use an interpolated precision-recall curve as our retrieval measure.
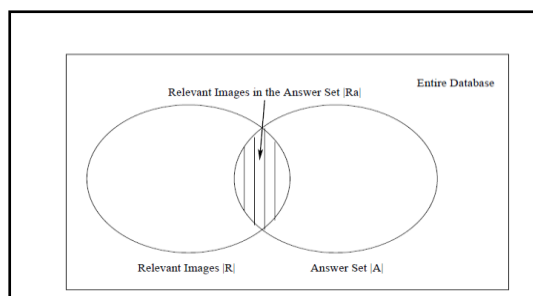
Fig.7 The diagrammatic representation of precision and recall

Table 1: Retrieval efficiency of Herbs

Sl.no	Planet Name	Leafy mass	Bark	Leafy mass	Bark
1	Catharanthes	16	18	80%	90%
2	Colacasia esculenta	14	18	70%	90%
3	Costus Speciosa	16	18	80%	90%
4	Seasamum indicum	14	16	70%	80%
5	Ocimum sanctum	16	14	80%	70%

Table 2: Retrieval rate of feature extraction for herbs leafy mass

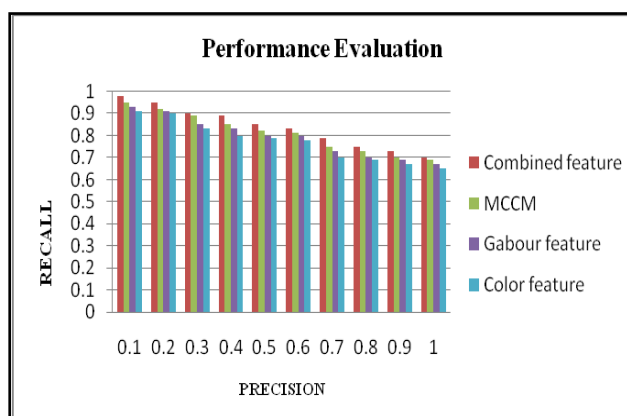| Precision	Combined feature	MCCM	Gabour feature	Color feature
0.1	0.98	0.95	0.93	0.91
0.2	0.95	0.92	0.91	0.9
0.3	0.9	0.89	0.85	0.83
0.4	0.89	0.85	0.83	0.8
0.5	0.85	0.82	0.8	0.79
0.6	0.83	0.81	0.8	0.78
0.7	0.79	0.75	0.73	0.7
0.8	0.75	0.73	0.7	0.69
0.9	0.73	0.7	0.69	0.67
1	0.7	0.69	0.67	0.65

Fig. 8 Performance Evaluation graph for Herbs leafy mass

7. CONCLUSION

The goal of Content-Based Image Retrieval (CBIR) systems is to operate on collections of images and, in response to visual queries, extract relevant image A clustering and indexing techniques has been proposed which is suitable for plant image database. Content Based Image Retrieval by Indexing method to improve retrieval performance with good accuracy, more result can be produced. Results shows that OTSU segmentation is highly effective for extracting plant image from background. Combination of gabor filter and MCCM are very effective for feature extraction from plant image. We have also presented a novel approach for image segmentation to extract the region features effectively. R*- Tree data structure is used in indexing the region features. The experimental results show that the proposed system can improve the retrieval accuracy as well as reduce the time for retrieval. There are no universal system / method for indexing or retrieval. We can try to develop something that robust. Indexing base on regional features give better result than global features.

However, we need to perform further experiments with a larger dataset and with different datasets to get more confidence in experimental results with better feature extraction. The future of CBIR depends a lot on the collective focus and overall progress in each aspect of image retrieval, and how much the average individual stands to benefit from it.

8. REFERENCES

[1] A.Drimbarean, P.F.Whelan, "Experiments in colour texture analysis", (Feb 2001), PP 1161-1167.

[2] Christoph Palm, Thomas M. Lehmann, "Classification of color textures by gabor filtering", (2002), vol. 11, PP 195-219.

[3] Jiatao Sonl, Zheru Chi, "Bark Classification by Combining Grayscale and Binary Texture Features", Proceedings of 2004 Inrernational Symposium on Intelligent Multimedia, VJdeo and Speech Processing, (October 2004), PP 450-453.

[4] Jongan Park, "Image Indexing using Spatial Multi-Resolution Color Histogram", PP 1-4.

[5] Kian-Lee Tan, "An Evaluation of Color-Spatial R.etrieva1 Techniques", International Conference, (September 1997), PP 1078-1082.

[6] M.V.Sudhamani, Dr.C.R.Venugopal, "Image Retrieval From Databases: An Approach Using Region Color and Indexing Technique", International Journal Of Computer Science and Network Security, Vol.8,(Jan 2008), PP 54-63.

[7] Mohamad Obeid, Bruno Jedynak, et.al. "Image Indexing and Retrieval Using Intermediate Features", (Sep 2001), PP 531-533.

[8] Sang Hyun Bae, Dae Young Lee, et al, "Feature-Based Similarity Retrieval based on SOM-Based R*-Tree", PP 1-13.

[9] Suresh Pabboju, Dr. A.Venu Gopal Reddy, "A Novel Approach for Content-Based Image Indexing and Retrieval System using Global and Region Features", International Journal of Computer Science and Network Security, VOL.9, (February 2009), PP 119-130.

[10] Shyi-Chyi Cheng, Tian-Luu Wu, "Fast indexing method for image retrieval using k-nearest neighbors searches by principal axis analysis", Journal of Visual Communication and Image Representation,(2006), PP 42-56.

[11] Suihua Wang, Ailing Zhao, "An Image Indexing and Retrieval Model Using Reasoning Services", International Conference, (2009), PP 193-196.

[12] Seong-o shim,Tae-sun choi, "Image Indexing By Modified color co-occurrence Matrix", IEEE, (2003), PP 493-496.

[13] Xiaosong Wang, Xinyuan Huang, et.al., "A Color-texture Segmentation Method to Extract Tree Image in Complex Scene", International Conference on Machine Vision and Human-machine Interface, (2010), PP 621-625.