**Australian Journal of Basic and Applied Sciences**, 9(27) August 2015, Pages: 598-602

ISSN:1991-8178

**Australian Journal of Basic and Applied Sciences**

Journal home page: www.ajbasweb.com

# Performance Based Image Indexing and Accessing Through A Technique in Data Cube

[1]R. Karthikeyan and [2]Dr. Celine Kavida A

[1] *Dept. of CSE ,Vel Tech Multitech Dr RR & Dr SR Engineering College, India*
[2] *Dept.Of Physics ,Vel Tech Multitech Dr RR & Dr SR Engineering College, India*

**ARTICLE INFO**

***Article history:***
*Received 23 June 2015*
*Accepted 25 August 2015*
*Available online 2 September 2015*

***Keywords:***
*Image Indexing, Feature Extraction, Knowledge Discovery, Multidimensional Indexing, Spatial Data*

**ABSTRACT**

Background: Image mining is an important retrieval efficient mechanism which accesses the data from the database storage and arrange in the form of queues and priority. Primary and Secondary indexes used currently the retrieval system for similarity hidden phase. In this case indexing has been done in similarity space. Indexing techniques ranging from standard methods such as inverted access methods and multidimensional methods. A specific indexing scheme is necessary for web image mining which presents and efficient color indexing scheme for similarity based retrieval system with the advancement in new image retrieval mechanism, gives better performance in color special retrieval techniques. Good indexing results in better classification and clustering of large image databases. Hence, the indexing mechanism is very important step in web image mining.



**To Cite This Article:** R. Karthikeyan and Dr. Celine Kavida A., Performance Based Image Indexing and Accessing Through A Technique in Data Cube. ***Aust. J. Basic & Appl. Sci.,*** *9(27): 598-602, 2015*

## INTRODUCTION

In the World Wide Web, the availability of a variety of sophisticated data and images are in massive collection of database storage, which retrieves the data based on the certain principals and also technique. Here, the efficient extraction is important through visual features and contents are needed to provide a meaningful indexing visual data (Zheu Liu, 2014). Most existing approaches to image indexing and retrieval use the textual keyword. Search and retrieval are performed on the keyword records and the associated images are retrieved after the textual search is complete (Xiaoqian Jiang, 2007). The main drawback of such an approach is that different users may have different interpretations for one image content and therefore, they may use different words to describe the image.

To enable such queries it is necessary to combine methods from pattern recognition (to detect and represent the content based features) and database techniques (to efficiently index and retrieve the relevant images based on those features) as well as learning capabilities to include new data entries.

In general image can be indexed by their image features. Image retrieval strategies based on either text-based or content-based retrieval. Text-based retrieval (manual) has the problem with Labor intensive, more prone to inter indexer consistency problems then indexing of text and Of-ness, thing-ness about-ness ambiguities. But Content-based retrieval (automated) is based on its contents (any information that can derived from image itself) such as Color, Texture, Shape, Color Layout (both color feature and spatial relations) (http://en.Wikipedia.org/wiki/).

The block diagram of content-based image retrieval system is shown in figure 1.It is highly based on user's query image and the feedback of the user. Since the early 90's, content-based image retrieval has become a very active research area. Many image retrieval systems, both commercial and researches have been built. Most image retrieval support one or more of the following option (Chang, S.F., 1998).

- Random browsing.
- Search by example.
- Search by sketch.
- Search by text.
- Navigation with customized image categories.

QUBIC (Query By Image Content), Virage, Retrieval Ware, Photobook VisualSEEk and WebSEEk, Netra, MARS (Multimedia Analysis and Retrieval System) and ART MUSEUM these are all the example content based image retrieval system (Yong Rui).

**Corresponding Author:** M.S. Padmavathi, Research Scholor, PG and Research department of Computer Science and Applications, DG Vaishnav College, Arumbakkam, Chennai – 106.
E-mail: padmanivas_2002@yahoo.co.in.

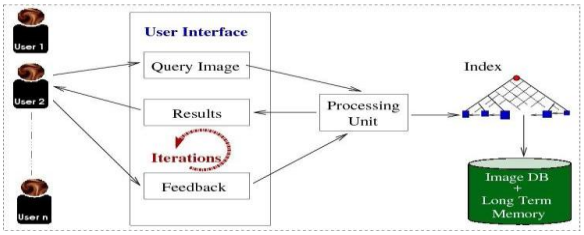


**Fig. 1:** Existing Content-Based Image Retrieval System.

In this paper we propose a new approach to handle multi-dimensional image feature by using data cube. Consider images are stored in data base .These images are arranged in random manner. When the user gives the query image to data base, it will create the clusters by using clustering algorithm (i.e. K-means clustering algorithm). Before images are searched in the data base we have to check whether the given query image is valid or not. This can be achieved by data preprocessing technique (i.e. cleaning, noise reduction technique). After correcting the query image, it will search the similar images in the data base. After that clusters are created based on the similarity mean value of the query image. That is two clusters are created first one is relevant images based on the given query image. Second one is outliers. We can avoid the result of outlier only take consideration on the relevant query image results. After extracting the relevant images from date base, we will apply association rules (i.e. apriori algorithm) used for identifying the most frequently used image features. In-order handle multi-dimensional feature or data we introduce new concept data cube which is used handle many features at the same time. By using this data cube we can perform flowing operation.

### 1.1 *Overview of Image Database:*

Our image database mainly consists of 5 functional units

• the image database itself with the index structure,
• the image processing and interpretation unit with the image query construction unit,
• the textual user interface,
• the automatic image query construction unit,
• the similarity determination and retrieval unit, and
• the learning unit for index structure updating.

Two query types are possible: textual keywords and image content based queries. For inputting the textual queries the user is provided by the system with a vocabulary for image description, see The same terms are used by the signal-to-symbol transformation unit which automatically transforms the numeric features extracted from the image into a symbol.

The content based query is automatically processed from the image by the image processing algorithm and feature extraction unit. This results in another problem with image databases caused by pattern recognition. An accurate automatic detection and recognition algorithm never works for all kind of images. Such an algorithm is rather domain dependent than generic. They suggest providing an image database with standard procedures for image enhancement, image analysis and feature extraction. The application of these procedures to the image is left to the user. He can process the query in an interactive fashion with the help of the procedures provided by the system. However, this requires knowledge of the user about image processing technology. He needs to be trained on image processing and pattern recognition technology in order to understand what effect an image processing and pattern recognition procedures produces on the image.

## 2. *Indexing and retrieval:*

### *2.1. Vantage Point Tree:*

Yianilos introduced the Vantage Point Tree, or VP tree for short in. A VP tree relies on pseudo-metrics. First a point **v** (vantage point) is selected. To construct the index, the database points are sorted according to their distance from **v** (i.e., in ascending order of $d(^{2}; \mathbf{v})$). The median distance is computed and all points having a distance smaller than the median are assigned to the left subspace of **v**, while the remaining ones are assigned to the right subspace. This procedure is recursively applied to the left and right subspace. The simplest VP trees rely on simple operations to select the appropriate vantage point among a random subset of the points associated with each node of the tree. It is clear that the choice of vantage points at each level of the VP tree plays an important role in the performance of the indexing algorithm. In our implementation, we select the point which is furthest from the center of gravity as vantage point. Our experiments prove that this is a better choice than random selection. Because the vantage point tree is a binary tree structure, binary search algorithm can be used to search the tree. The searching complexity is then ($O$ log2 $N$).

Indexing based on VP trees is faster than using K-d trees, since one does not need to project points onto each dimension during the querying stage. Moreover, similarity of regions is expressed as distances between feature vectors, which make VP trees the more natural choice (also using distances) and especially well suited for our task.

### *2.2. Off line database construction:*

When constructing the database, we first extract all invariant regions from each image in the database and compute the feature vector of moment invariants for each region. Then we construct the VP tree. As distance measure in feature space, we use the Mahalanobis distance, to correctly take into account the different variability and interdependence of the different components of the feature vectors. To this end, the covariance matrix was estimated based on all the regions extracted from the database images. At each leaf of the VP tree we store not only the feature vector, but also the invariant region and the database image from which it was extracted.

### *2.3. Retrieval implementation:*

When a new query image has to be processed, the first step is again to extract all invariant regions from the query image and to compute the feature vectors of moment invariants for them. Each region extracted from the query image is used as a query region to find similar regions in the database based on the binary tree searching algorithm. Here, two regions are considered similar if the Mahalanobis distance between them is below a predefined threshold. We say the query region is *matched* to a region in one of the database images if the latter is retrieved from the VP tree. Once all the regions in the query image have been processed, the images in the database can be ranked based on the number of matches found, i.e. based on the number of regions they contain that are similar to regions in the query image.

## *3. Multi-Dimensional Indexing:*

There are many multi-dimensional indexing techniques are available. The existing popular multi-dimensional indexing techniques are Bucketing algorithm, k-d tree, priority kd-tree, quad-tree, K-D-B tree, hB-tree, R-tree and its variants R+-tree and R*-tree. In addition to the above approaches, clustering and Neural Nets, widely used in Pattern Recognition, are also promising indexing techniques. Among those techniques BA-KD-tree gave the best performance.

The following three characteristics of the dimension-reduced data can be used to select good existing indexing algorithm.

- The new dimension components are ranked by decreasing variance.
- The dynamic range of the dimension is known.
- The dimensionality is still fairly high.

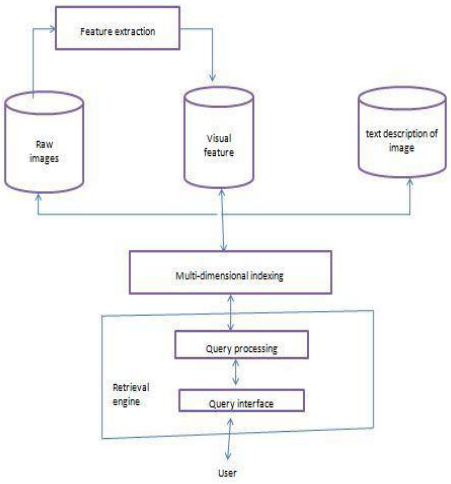


**Fig. 2**: An image retrieval system architecture (Yong Rui ).

In figure 2 there are three databases are used, for storing raw images, storing the extracted visual feature of the images and finally storing the textual description of the images. By using multi-dimensional indexing scheme we can handle three databases and index the images. It can index the large collection image-set in a fast manner. Query interface should be user friendly and it is graphical based representation.

### *3.1. Query Image Preprocessing:*

It is one of data mining preprocessing technique used for Detecting and correcting corrupt or inaccurate annotated text from the query image. It is used to identify incomplete, incorrect, irrelevant query image. By using of this technique we can determine whether the given query image is valid or not. Furthermore, the dirty data can cause confusion for the mining procedure, resulting in unreliable output.

### *3.2. Image Search in Database:*

It is the most important phase for image retrieval. Here we apply k-means algorithm to partition the images based the given query image. K-means algorithm is based on iteratively reassigning

objects to cluster improve the partitioning is known as iterative relocation. Here cluster centers are represented by mean value. It is highly based on distance based. The purpose of the creating this cluster is the object within a cluster are "similar" to one another and "dissimilar" to objects in other cluster in term of data set attributes. Cluster variation can be calculated by sum of the squared error between all objects in Ci and the centroid ci, defined as k E= Σ Σ dist(p,ci) i=1 p € c Where E is the sum of the squared error for all objects in the data set. p is the point space representing a given object and ci is the centroid of the cluster (Jiawei Han,)

The main objective of the clustering technique is we can extract the required similar images from data base by creating the clusters based on the mean value. We can also identify the outliers (i.e. dissimilar to query image). By extracting the outlier from database we can get relevant image which are more require by user. After this stage we would obtain the needed images from database.

### *3.3. Image Pattern Analysis:*

In this stage we have to find most frequently used image features by using pattern analyzing technique. Pattern analysis is process identifying pattern (i.e. most frequently occurred image feature) in the database. In-order to get the image pattern we apply Apriori algorithm. It is based on an iterative approach known as a level-wise search. It can be achieved by Apriori property (All nonempty subset of a frequent item-set must also be frequent.) is used to reduce the search space. This algorithm follows two-step process is known as join and prune step. This algorithm is also has the property is known as anti-monotonicity (i.e. if a set cannot pass a test all of its supersets will fail the same test as well) (Jiawei Han,)

It follows the following steps to predict frequent image features.

- Scan the all image features in the cluster and find count for each feature occurrences.
- It compares the minimal support count values to the image feature occurrences value.
- It chooses feature only high occurrence value compare than minimal support count and it forms candidate image feature set.
- After the forming candidate feature set, it checks whether sub set of its feature is frequent or not. If it is not frequent it will prune that image feature set.
- This procedure will follow until it finds the most frequent feature set.

### *3.4. Deployment of Data Cube and Image Retrieval:*

After predicting frequent image features, we have to deploy those features in to the data cube for performing or handling multidimensional image features simultaneously. Bitmap indexing algorithm is used for index the image features. It has advantage of quick access on data cube. This algorithm based cardinality (distinct values) of the columns [9]. Bitmap indexing can be achieved by bitwise logical operation on bitmap in-order to answer the given query. It can do AND, OR or XOR operation and it is only useful for read-only system. In data warehouse Bitmap index can be achieved by using star schema. This schema use two tables such as fact table and dimension table. By using a single fact table we can index many dimension tables simultaneously. By using this bitmap index we can effectively handle and index the image features.

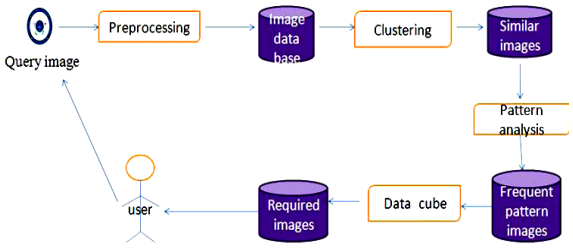


**Fig. 3:** Proposed architecture for image retrieval.

Figure 3 shows over all architecture of the image retrieval system from the user's query image to find the similar query images. Finally similar query images are display to user by using image retrieval system as we discuss in the early literature.

### *Conclusion & Future Work:*

In this paper we propose a new image indexing in multidimensional data cube by using bitmap index. Bitmap can perform only on low-cardinality columns but not well in high-cardinality and It does not have definitive design to index so further step of this project is increase the performance in high-cardinality and create a standard index format to it.

## REFERENCES

Zheu Liu,Houqiang Li,Wengang Zhou,Ruizhen Zhao and Qi Tian, 2014."Contextual Hashing for arge-Scale Image Search", *IEEE Trans. On Image Processing*,. 23: 1606-1614.

Xiaoqian Jiang,Wanhong Xu,Yiheng Li,Latanya Sweency, 2007. Ralph Gross,Daniel Yurovsky,"2D Image Database Indexing" *IEEE conference paper,ICME*.

"Wikipedia search CBIR system,http://en.Wikipedia.org/wiki/Content-based_image_retrieval

Chang, S.F., A. Eleftheriadis and Robert McClintock. 1998. "Next-generation Content Representation,Creation and Searching for new Media Application" in Education.IEEE proceedings, 86(5): 602-615.

Yong Rui and Thomas S. Hung, Shih-Fu Chang."Image Retrieval: Current Techniques, Promising Directions and Open Issues"**.**

Google image search, "http://cvit.iiit.ac.in/ projects/cbir"**.**

Jiawei Han,Micheline Kamber and Jian pie. "Data mining concepts and techniques**",** 3rd ed.ch. 4: 146-148

Wikipedia search on Data cube, "http://en.wikipedia.org/wiki/Data_cube".

Wikipedia search on Bitmap Index, "http://en.wikipedia.org/wiki/Bitmap_index".