ISSN 2320 – 2602
Volume 3, No.10, October 2014
International Journal of Advances in Computer Science and Technology
Available Online at http://warse.org/pdfs/2014/ijacst033102014.pdf

CREATING AN EFFECTIVE IMAGE INDEXING STRUCTURE BY USING MULTI-DIMENSIONAL INDEXING TECHNIQUE IN DATA CUBE

M Yesudoss Winster[1], R Krthikeyan[2], K Vijay Kumar [3]

[1]Dept. Of CSE,Vel Tech Multitech Dr RR & Dr SR EngineeringCollege,India,yesudosswister@gmail.com

[2]Dept.Of CSE ,Vel Tech Multitech Dr RR & Dr SR EngineeringCollege,India,rkarthikeyan@veltechmultitech.org

[3]Dept.Of CSE,Vel Tech Multitech Dr RR & Dr SR EngineeringCollege,India,vijaykumaranmail@gmail.com

ABSTRACT

Now a day the usage of image is rapidly increased, so indexing the exact image in the data base is very difficult because an image will have more than 4000 features [1]. It is highly impossible to store and index these much amount of features for a single image representation. In-order to make an efficient indexing structure we have to reduce the image features by using association rules (i.e. frequent image features are to be identified) but reduction image feature does not affect reliability of image results. So in this paper we introduce new image indexing technique by using data cube (multidimensional indexing) for handling multidimensional image feature.

Key words: Image feature, Association rule, data cube, Multidimensional indexing.

1. INTRODUCTION

With the explosive proliferation of the internet and the World Wide Web (WWW), the amount of digital image data accessible to users has grown highly. So sizes of the Image databases are becoming larger and more widespread, and there is a growing need for effective and efficient image retrieval and indexing systems [2]. 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) [3]. 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 [4].

- 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 [5].

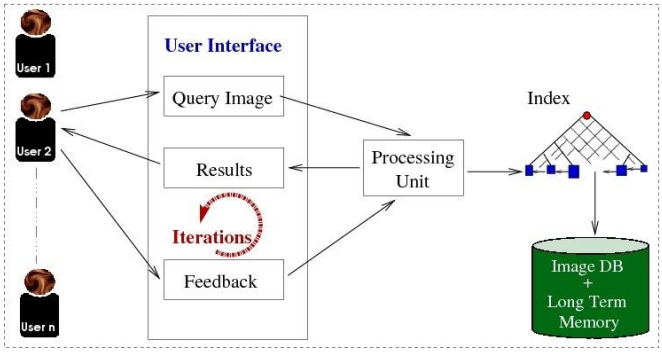

Figure 1: Existing Content-Based Image Retrieval System [6].

2. OUR APPROACH

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.

- Roll-up (Summarizing the data along the dimension or dimension reduction).
- Drill-down (Transform most summarized to most detail form).
- Slice (It performs operation on single dimension of the given cube)
- Pivot (Rotation of the cube).
- Dice (Can do operation on a sub cube based on the selection of the sub cube) [7].

Image features can be indexed by bitmap indexing algorithm for fast retrieval of images.

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 [5].

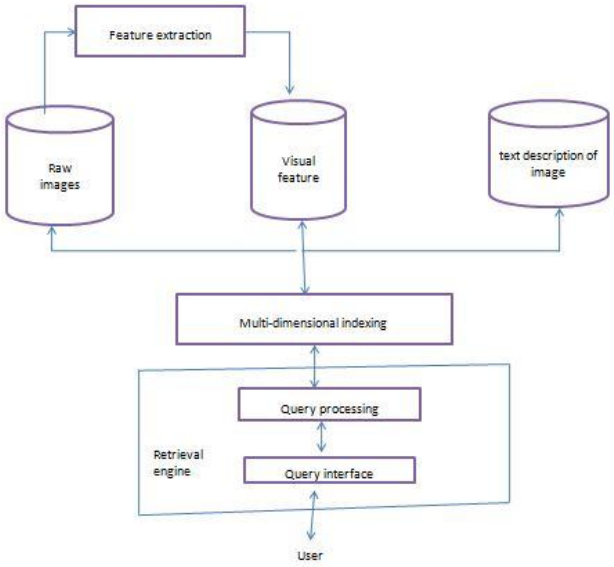

Figure 2: An image retrieval system architecture [5].

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 [7].

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 C_i and the centroid c_i, defined as

$$E= \sum_{i=1}^{k} \sum_{p \in c_i} dist(p,c_i)$$

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 c_i is the centriod of the cluster [7].

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) [7].

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.

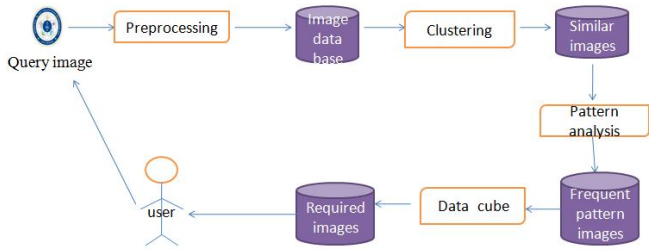

Figure 3: Proposed architecture for image retrieval.

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.

4. CONCLUSIONS AND 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

1. Zheu Liu,Houqiang Li,Wengang Zhou,Ruizhen Zhao, and Qi Tian."Contextual Hashing for Large-Scale Image Search", *IEEE Trans. On Image Processing* , Vol. 23, pp. 1606-1614, April 2014.
2. Xiaoqian Jiang,Wanhong Xu,Yiheng Li,Latanya Sweency,Ralph Gross,Daniel Yurovsky,"2D Image Database Indexing" *IEEE conference paper,ICME 2007*.
3. "Wikipedia search CBIR system,"http:// en.Wikipedia.org/wiki/Content-based_image_retrieval".
4. S.F. Chang, A. Eleftheriadis, and Robert McClintock."Next-generation Content Representation,Creation and Searching for new Media Application" in Education.IEEE proceedings,Vol. 86(5),pp. 602-615, September 1998.
5. Yong Rui and Thomas S. Hung, Shih-Fu Chang."Image Retrieval: Current Techniques, Promising Directions and Open Issues".
6. Google image search,"http://cvit.iiit.ac.in/projects/cbir".
7. Jiawei Han,Micheline Kamber and Jian pie."Data mining concepts and techniques", 3rd ed.ch. 4, pp.146-148.
8. Wikipedia search on Data cube, "http://en.wikipedia.org/wiki/Data_cube".
9. Wikipedia search on Bitmap Index, "http://en.wikipedia.org/wiki/Bitmap_index".

.