ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 5 , May 2017

Content Based Image Retrieval Using Big Data

SurakshaJadhav, Bhagyashree Shinde, VaibhavPawar, Virendra Singh Bhatediya.

B.E. Student, Department of Computer Engineering, STES's NBN Sinhgad School of Engineering, Ambegaon (bk), Sinhgad Road, PUNE

ABSRACTION: In the digital age of graphics and image processing, image retrieval using visualcontents of an image such as colour, shape, texture and spatial layout, from the perspective of human vision among the widely stored database of thousands of digital images such as Corel Image Database is a comprehensive image retrieval technique rather than cumbersome text based image retrieval. MongoDB with GridFS, an unstructured database system stores an image into the different chunks along with its metadata. This enables the efficient retrieval of images from the clusters formed of big data sets of digital images. Content Based Image Retrieval (CBIR) techniques such as Semantic retrieval, Interactive Machine Learning and Relevance Feedback outputs the required images along with the nearest matching visual contents of that input image. User Interface can accept the pre-existing image, rough image layouts, colors or shapes to be matched from the clustered database. The main motive behind using the Big Data approach in CBIR is to use distributed and unstructured digital image database for faster image retrieval.

KEYWORDS: Mongo DB,CBIR,Feature extraction Techniques, Coral Database, GridFS.

I. INTRODUCTION

Day by day digital area captures devicesproducing more images. Almost in every field images are use and those images placed in the database and as per the user need should to be retrieve from database for that purpose content base image retrieval system use it's the one of the quick solutions[3]. Content base image retrieval focused on the content of the that images that content in the numerical form to make it more quick accessing we are using the Mongo DB because its free and open source and provide cross platform document oriented database programs these program known as NoSQL database program also using the JSON Document file Mongo DB supportive for the long quires and regular expression quires could include the JavaScript function.
Features of Mongo DB as follows:-

1. Supportive to Ad-hoc Quires.
2. Provide indexing.
3. Replication of the data.
4. Mongo DB often uses as file store system.
5. Provide the Load balance between Data of Number of server.
6. Map reduce can use for aggregation.
7. It has the JavaScript execution at server side.
8. Also supportive for the fixed size data.

II. LITERATURE SURVEY

From the study of various sources, it was concluded that in today's era the amount of digital images are growing in a very explosive manner. The storage requirement for storing these images is also increasing from gigabyte to petabyte. Searching and retrieval of particular images from the massive database is not possible when the images in the database are wrongly annotated and described. For getting the correct image, during the search, content based image retrieval can be used to search and retrieve the images from the massive collection of images. The query image is compared with database images on the basis of their feature descriptor; this can improve efficiency of searching and retrieval.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 4, Issue 5 , May 2017

III. CONTENT BASED IMAGE RETRIEVAL

The term "Content-Based Image Retrieval" is used for retrieving the corresponding images from the database based on their feature of images which derived the image itself like texture, color and shape and domain specific like human faces and fingerprints. The retrieval on the based on the content of an image is to be more effective than the text based which is called content based image retrieval that are used for a various applications like vision techniques of computer [3]. Traditionally, search of the images are using text, tags or keywords or annotation assigned to the image while storing into the databases. Whereas if the image which is stored in the database are not uniquely or specifically tagged or wrongly described then it's insufficient, laborious and extremely time consuming job for search the particular image in the large set of databases [1].

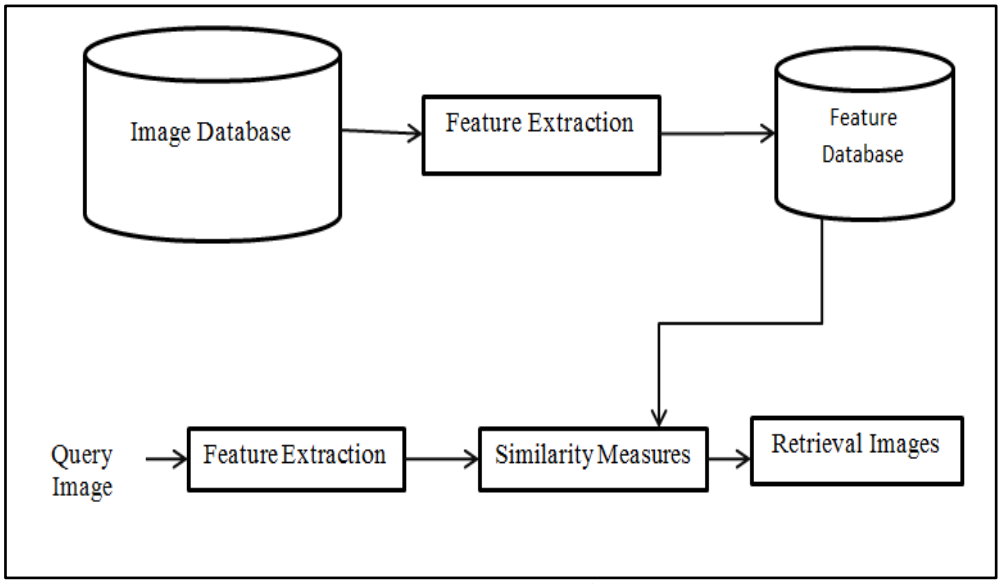


Figure 1 System architecture for CBIR

For these purpose obtaining the most accurate result CBIR system are used which searches and retrieve the query images from the large databases based on their image content like color, texture and shape which derived from the image itself[1].

QUERY BY IMAGE CONTENT (QBIC): Query by Image Content is the first commercial content based retrieval system. It is also called as Content Based Visual Information Retrieval System (CBVIR). This system was developed at the IBM Almaden Research Centre.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 4, Issue 5 , May 2017

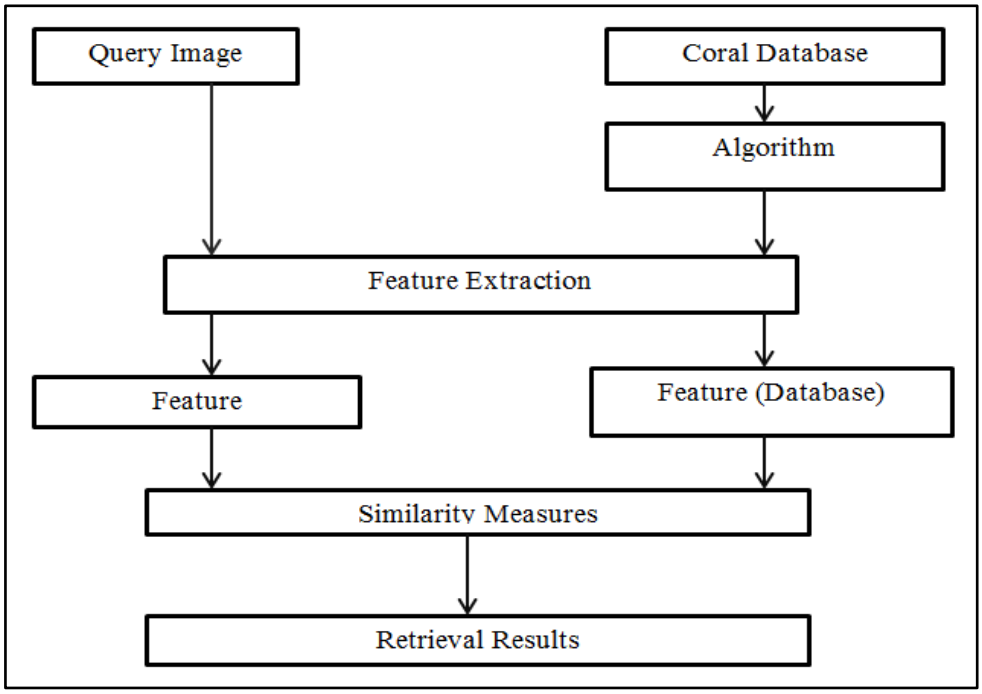


Figure 2 Flowchart of query by image retrieval system.

It works on principle of extracting several features from each image, namely color, texture, shape, size, position features. This gives ability to graphically pose and refine queries based on multiple visual properties such as color, texture and shape. It supports queries based on input images, user-constructed sketches, photos, etc. It is being used in IBM's Multimedia Manager in OS/2, and in DB2 Extenders [1]. QBIR consider two main factors as-.

1. It uses computable properties of image and videos like color, shape, texture, motion of object and other graphical information in the query.
2. It is a Graphical Query Languages which implies the query by drawing, selecting and retrieving the graphical features of digital image.

CORAL DATABASE: We manually divided 10,800 images from the Corel Photo Gallery [6] into 80 concept groups, e.g., autumn, aviation, bonsai, castle, cloud, dog, elephant, iceberg, primates, ship, stalactite, steam-engine, tiger, train, and waterfall. Figure 6 shows some example images. We reorganised the Corel Photo Gallery,

1. .Because many images with similar concepts were not in the same group and
2. Some images with different semantic contents were in the same group in the original database. In the reorganised database, each group includes more than 100 images and the images in the group are category-homogeneous. These concept groups were used in the evaluation of the results of our algorithms [1-5].

MAP REDUCE:Map-reduce is a data processing paradigm for condensing large volumes of data into useful aggregated results. MongoDB uses mapReduce command for map-reduce operations. MapReduce is generally used for processing large data sets.
Map process- Large data divide into small data block.
Spilling-spilling done by the use of background thread.
Partition- Thread partitioned the data.
Sort process- Memory sort method use for sorting.
Merging- merge into one block.
Reducing- this process performed by the copy,sort,reduce phase.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 5 , May 2017

GRIDFS: GridFS is a specification for storing and retrieving files that exceed the BSON-document size limit of 16MB. Instead of storing a file in a single document, GridFS divides a file into parts, or chunks [1], and stores each chunk as a separate document [5].

IV. EXTRACTION OF IMAGE

COLOR BASED FEATURE EXTRACTION: Color is the most commonly used attribute in image retrieval systems. Retrieval in these systems is done on the basis of similarity in color. For each image in database, color histogram is computed which shows pixel position of each color in the image [4]. Most commonly used color based image retrieval methods are RGB and HSV [4].
Steps for including Color features extraction as follow:-

1. Read the input quires.
2. Conversion process between the color spaces of RGB to HSV.
3. HSV's every pixel quantization.
4. Store the 256_vector _value in database.

SHAPE BASED FEATURES EXTRACTION: Well know concept in this extraction is MORPHOLOGY, which provide the operation on the images for the extraction of shape of the images. Morphology perform the two operations Dilation and Erosion, Dilation operation understand the shape of objects of image by including the dilated pixel around the boundaries of the objects of the image, And erosion Remove the pixel for the boundaries of the object of the image[4].
Steps include in the shape based feature extraction as follows:-

1. Read the input query image.
2. Conversion of the RGB input image into image of grey scale.
3. Calculation of 4 morphologies for the Edge map.
4. Calculate the Edge Maps.
5. Store those Edge map into vector form then store into database.

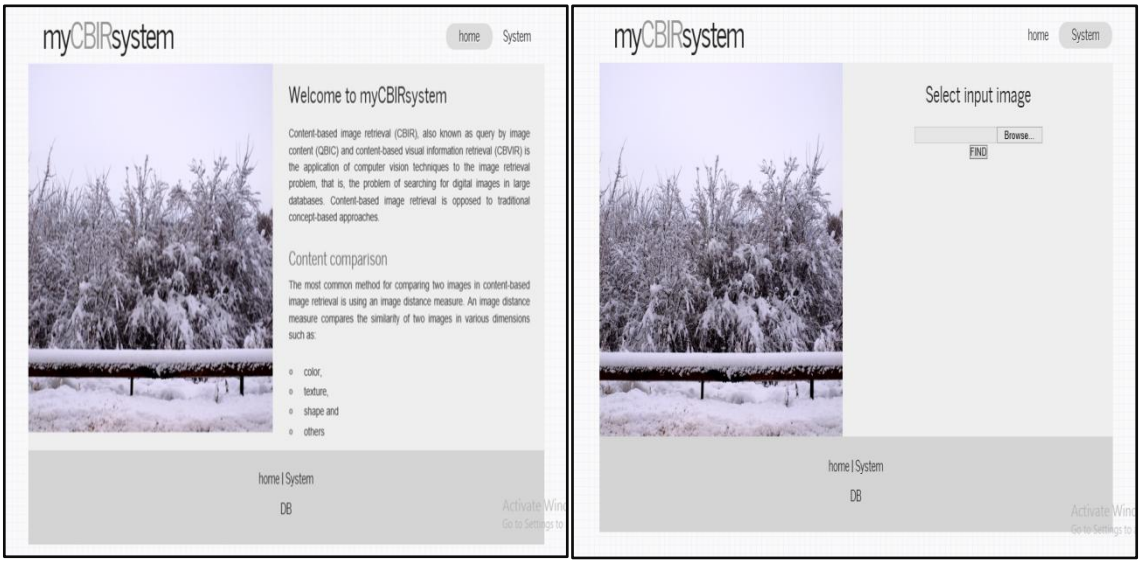
TEXTURE BASED FEATURES EXTRACTION: Texture similarity may not seem to be useful for image retrieval. Texel's are made of intensity of pixel. They help in differentiating between textured and non-textured images in database. The computation of Texel's may be defined based on the direction of texture, its coarseness and regularity. Texture has certain classification like fine, gross with respect to pixel of an image. This retrieves textured regions in images on the basis of similarity to automatically-derived towards representing important classes of texture within the collection of images [4].
Steps include in the texture based feature extraction as follows:-

1. Read the input query image
2. Conversion of the RGB input image into image of grey scale.
3. There is Calculation of GLMC matrix for all four directions 0', 45', 90', 130'.
4. There is Statisticalfeature computation for the entire featurematrix.

SIMILARITES MEASUREMENT: Similarities between images can be compute by using the Canberra Distance measure x and y is the feature vector ofthe input image query.

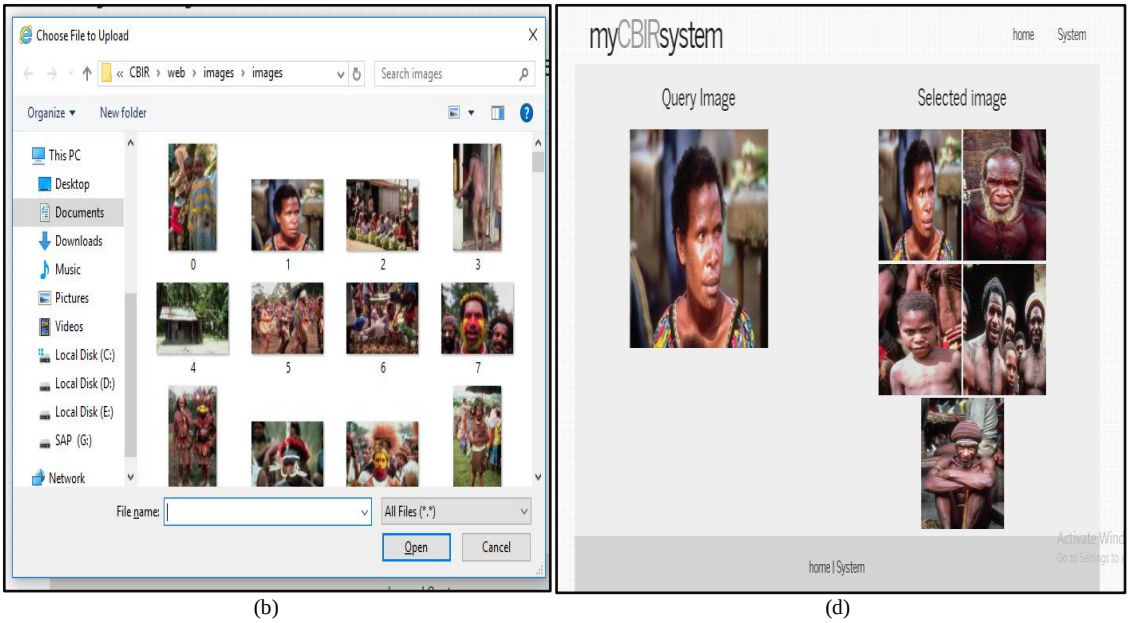
ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 4, Issue 5 , May 2017

V. SYSTEM OUTLOOK



(a) (c)

Figure 3 (a) System Welcome look. (c) Set an input image for processing.



(b) (d)

Figure 4 (b) choosing an images to upload similar result from database. (d) Showing similar Final results to given query images.

VI. CONCLUSION

Mongo DB provide the best image retrieval and also provide the trustworthy background framework using Map Reduce. Map Reduce also helpful for the real time user requirement .when the huge data amount of the data is present in the database ,then the traditional approaches is not that much helpful as compared to the CBIR. We are also concentrating on the increasing the accessing speed and reducing the transmission speed.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 5 , May 2017

REFERENCES

[1] Mrunalini D Pawar,ShradhhaShinde,Ramdas B bagawade, "Content BasedImgaeRetreival : A study", IJARCCE,Vol 5,issue 3,march 2016.

[2] Priyanka Shinde 1, AnushkaSinkar 2, Mugdha Toro 3, Prof.ShrinivasHalhalli, "Image Retrieval Based on its Contents Using Features Extraction", International Research Journal of Engineering and Technology (IRJET), Volume 03, 01 Jan-2016.

[3] SwapnilSaurav, PrajaktaBelsare, Siddhartha Sarkar, "Holistic Correlation of Color Models, Color Features and Distance Metrics on Content-Based Image Retrieval", IRJET, Volume: 02, 07 Oct-2015.

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

[5] http://javabeat.net/mongodb-gridfs-tutorial/.