Private Content Based Multimedia Information Retrieval Using Map-Reduce

Swapnil P. Dravyakar[1], Sunil B. Mane[2], Dr. Pradeep K. Sinha[3]

[1]*PG Student, Department of Computer Engineering and Information Technology, COE, Pune 411005, MH., India*

[2]*Assistant Professor, Department of Computer Engineering and Information Technology, COE, Pune 411005, MH., India*

[3]*Senior Director (Corporate Strategy and R&D), C-DAC, Pune, MH., India*

*Abstract-*Today's large amount of variety multimedia information like audio, video and images are massively produced through mic, digital camera, mobile phones and photo editing software etc. In these paper we proposed a solution for a large database of images which provides secure, efficient and effective search and retrieve the similar images of Query image from the database with the help of a novel technique of Local Tetra Pattern (LTrPs) for Content Based Image Retrieval (CBIR) which carries the interrelationship in between the center pixels and its surrounded neighbours of center pixel by computing difference of gray level. The images which are stored on the database are private or contain private information of particular user and these digital images should not be accessed by others except the user. The user wish for a privacy and security for their images while storing and retrieving the images from the database. For that the private content based image retrieval (PCBIR) used for accessing the similar images from the database of query image even without learned to database admin.

This system incorporates Map-Reduce technique for privately search and retrieve of images over large datasets of user images. Another part of the system is large storage i.e. Hadoop distributed file system (HDFS), Hadoop specify a scheme that processes on large datasets in a distributed cluster of computer by using a simple programming models. Proposed solution will be a system to upload, search and retrieve the query image among the database of images.

Keywords - Map-Reduce, CBIR, Image Retrievals, Local Tetra Patterns (LTrPs), Hadoop, Cloud Computing.

I. INTRODUCTION

Now a day's Technology become a progressively advanced gives a multimedia devices are also become advanced and cheaper in cost results multimedia data in an explosive amount. In which produces multimedia data like image are massive and need to store on a large datasets for various applications such as prevention of crime, medical, security become a necessity to store ,search and retrieve the images from the large datasets in a parallel processing technique for efficiently and effectively.

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 [2].

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 [9]. 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.

In these proposed system we describe the novel tetra patterns for the efficient and effective retrieval of the image which incorporated with Map-Reduce framework such as Hadoop for fast calculation and return the result in shorter time. The Map-reduce framework works in parallel manner which processes on very large image collection of petabyte of storage.

II. PRELIMINARIES

In this Section we give some procedure and Definitions that are used to decide the flow and working of System for effective and efficient similar image retrieval from the large databases of images. Section 2.1, 2.2, 2.3, 2.4 described the image features, section 2.5 present the novel technique of local tetra pattern which are used in the proposed system for search and retrieve the image from the HDFS database. Section 2.6 gives the Map-Reduce framework of Hadoop which describe for the large database of images.

A. Content Based Image Retrieval

Content based image retrieval means retrieving the similar image from the database based on the three major features of Image such as color, texture and shape which the image derived itself. In that 'Content based' refers to search the images from the collection of database based on the content rather than annotation based [1].

B. Color

Color is the one of the most important feature of the image retrieving process [1]. It Is the most basic form which is used to Retrieving the images from the database. The more common approach to comparing on several color spaces which are Red, Green and Blue (RGB), HSV, CMYK, CIE L*a*b* and CIE L*u*v*, and color histogram which identifies the relative proportion of pixels within certain values [1],[2].

C. shape

Shape is another important low level feature of images. Which are used to measuring the shape of different objects for differentiates between various objects, similarity measurements of shapes are difficult. For that two main steps are requiring for retrieving images are extraction of feature and measurement of similarity the edge detection and contour based are the two main categories for that. [1], [2].

D. Texture

Texture is an important feature of an image. Textures are defined by Texel's which is the intensities of pixel in a specific region. Based on the color property the textures are used to classify the image from textured image to non-textured image that are to be used for efficient and effective retrieval of image based on them. Texture gives the spatial relationship of colours in the image and also from the gray tones of themselves for the working of both classification and segmentation. [7].

E. Local Tetra Patterns(LTrPs)

The novel Local Tetra Pattern is Modified from the combination of local patterns (LBP,LTP and LDP).this technique encodes with relationship between the center pixel to its neighbours, based on that its calculated using first order derivative in horizontal and vertical direction of the center gray level pixel g_c.

Let, g_c denotes the center pixel and g_h , g_v denotes the horizontal and vertical pixel of the neighbors of center gray level pixel g_c respectively. In this method the center pixel can be calculated as below.

$$I_{0^\circ}^1(g_c) = I(g_h) - I(g_c)$$

$$I_{90^\circ}^1(g_c) = I(g_h) - I(g_c)$$

And the directions of the center gray level pixel can be calculated as,

$$I_{Dir}^1 = \begin{cases} 1, & I_{0^\circ}^1(g_c) \geq 0 \text{ and } I_{90^\circ}^1(g_c) \geq 0 \\ 2, & I_{0^\circ}^1(g_c) < 0 \text{ and } I_{90^\circ}^1(g_c) \geq 0 \\ 3, & I_{0^\circ}^1(g_c) < 0 \text{ and } I_{90^\circ}^1(g_c) < 0 \\ 4, & I_{0^\circ}^1(g_c) \geq 0 \text{ and } I_{90^\circ}^1(g_c) < 0 \end{cases}$$

From the above equation, image is converted into four values i.e.1, 2, 3, 4 and that four values are defined as a four

directions. After that the second order derivative is applied and forms it an 8-bit tetra pattern is obtained from center pixel then based on their direction of center pixel tetra pattern is divided into four parts [4],[5]. At last the tetra patterns for each direction converted into three binary-patterns. For that the direction of center pixel is 1 then (4*3=12) twelve binary patterns are obtained for four directions. Thus this method used to propose the 13th binary pattern (LP) by using the magnitude of horizontal and vertical first-order derivative using [5].

$$M_{I^1}(g_p) = \sqrt{(I_{0^\circ}^1(g_p))^2 + (I_{90^\circ}^1(g_p))^2}$$

F. Hadoop

Hadoop is a framework that allows for the distributed processing of large datasets, it is also capable of to process small datasets. However it also works on terabyte of data where RDBMS takes hours and fails whereas Hadoop does the same in couple of minutes. The Apache Hadoop is a open-source software project for scalable, reliable, flexible, distributed computing, failure handling [10].

The project includes these modules [10].

Hadoop Common: The common utilities which supports the other Hadoop modules.

Hadoop Distributed File System (HDFS™): It provides high-throughput access to application data in a distributed file system.

Hadoop YARN: A framework which scheduls the job and cluster resource management.

Hadoop Map-Reduce: system for parallel processing of large data sets based on YARN.
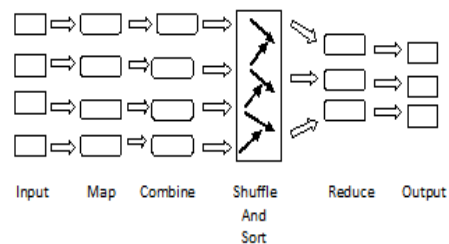
a) Map-reduce

Fig. 1 Map-reduce Architecture

Map-reduce is the application that works on the data stored in HDFS and act as resources scheduler. Map-Reduce is an Framework that works on distributed computing for support parallel computation over a large datasets in multiple petabyte of storage available on cluster of computer. The map-reduce operation can be run on Big Data of large clusters of commodity hardware in reliable and fault tolerant manner. Map-Reduce framework which works on a list of pairs <key, value>into a list of values. The output list can then be saved into Distributed file system then the reducer run to merge the result in parallel [11].
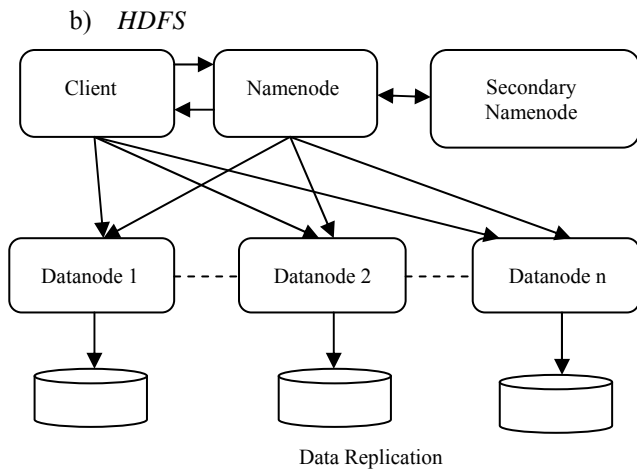
b) *HDFS*

Data Replication

Fig.2 HDFS Architecture

In HDFS Data is divided into chunks. Namenode is the Master of the File System and Datanode is the slave Component of the file system, only one namenode and multiple namenode are running on the Hadoop cluster. Data to be stored on node that is datanode [12].datanode should be replicated one each datanode, if one data node goes down then the data is present on another datanode also the Name node knows where the data is to be stored in which rack.Namenode contain all the data storage information which is stored in datanode. There is another namenode that also contain all the information like namenode called secondary namenode .if namenode fails then it will recover the information from secondary namenode [12].

c) *H-Base*

HBase on the otherhand is built on top of HDFS and provides fast record lookups for large table. This can sometimes be a point of conceptual confation.HBase internally puts your data in indexed "store files "that exists on HDFS for high speed lookups. Use of H-Base when you need random real-time read/write access to your big data [13].

III. LITERATURE SURVEY

R.P.Maheshwari et al. [4]. Proposed the novel image retrieval algorithm using content based image retrieval (CBIR). By local tetra patterns (LTrPs) carries the interrelationship in between the center pixels and its surrounded neighbours of center pixel by computing difference of gray level.

Sitalakshmi Venkatraman et al. [9]. Presents a novel Map Reduce framework for neural network for CBIR from collection for large data in a cloud environment. Classify the color images on the basis of their content and by using Map and Reduce functions accurate parallel results are arrived in real-time and shorter time that can operate within the cloud clusters.

Ashish Oberoi et al. [5]. Represent a framework based on Local Tetra Pattern and Fourier Descriptor for content based image retrieval from medical databases is proposed. The proposed approach formulates the interrelationship between the reference or centre pixel and from its

neighbours, considering the directions i.e. vertical and horizontal calculated using the first-order derivatives.

Dr.Ayyaz Hussain et al. [14]. Have been present the comparison of three different approaches of CBIR based on image feature, distance measure and precision of result. Result of these approaches show that local feature extraction is more important than global level feature extraction.

IV. SUMMARY OF LITERATURE SURVEY

We have researched various papers, In that we are going to concluded that in today's era the amount of digital images are growing everyday in a very explosive manner have to take gigabyte and petabyte of storage. For these, search and retrieve the particular images from that massive database are not possible when the images which are search on the database are wrongly annotated and described. For that the content based image retrieval are used to search and retrieve the images from the massive collection of images. For that images are compared with database images on the basis of their feature descriptor, that image contain itself. the search and retrieve from the massively collection of images from database parallel processing need to be done to give better result in a shorter time .the Map-Reduce Framework (Hadoop) result in real time efficiency even in large image collection occupying the gigabyte and petabyte storage of database.

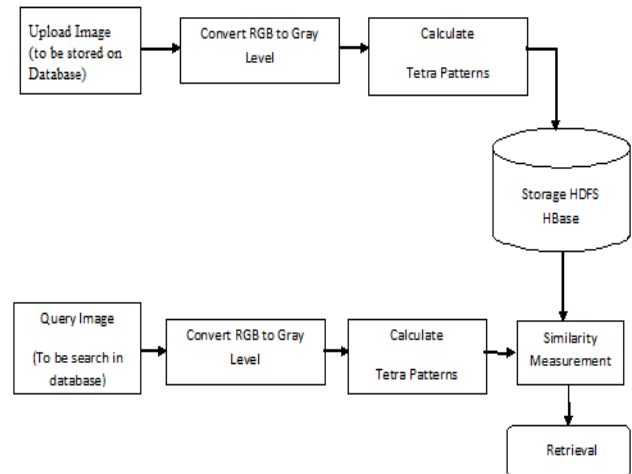## V. PROPOSED SYSTEM DESIGN AND WORKING

Fig. 3 Architecture of Proposed System

In these Proposed System the user upload the images on database of Hadoop such as HDFS then it will convert into gray scale. Which give functionality that the images could be any size, so it be will resize. After resizing, the first-order derivatives of vertical and horizontal axis are applied and the every direction of pixel are calculated based on the center pixel that patterns are divided into four parts. Then tetra patterns are calculated and separated into three binary patterns and also calculate magnitude of center pixel. The feature extracted from the binary patterns and obtaining binary patterns from the magnitude patterns are combined

to form a feature vector and stored this on database. The query image and images in the database are compared by using Euclidean distance for obtaining the similar measurement and the best matched images are retrieved from the database of images in response to query image [4],[5].

VI. Security Analysis

Image Database Privacy: in this system the image stored on the HDFS database of Hadoop is in the .text format which will not give any information the about the images on database even to the database admin.

VII. Conclusions

Thousands of images are growing through the various digital devices and these images are added to the image databases and internet for various applications which needs to store and retrieve the images in effective and efficient manner.

For these purpose we described the novel method LTrPs for PCBIR of large datasets incorporate with map-reduce framework of Hadoop technology for effective image retrieval on cloud environments.

References

[1] Hiremath P.S. and Pujari J., "Content Based Image Retrieval Using Color, Texture and Shape Features*," International Conference on Advanced Computing and Communications, ADCOM ,* pp.780 – 784,2007,

[2] Shankar M. Patil "Content Based Image Retrieval Using Color, Texture and Shape," *International Journal of Computer Science & Engineering Technology (IJCSET)*, Vol. 3, Sept. 2012.

[3] Ryszard S. Choras "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems" *International Journal of Biology And Biomedical Engineering*, Vol. 1, 2007.

[4] Murala S., Maheshwari R.P., and Balasubramanian R., "Local Tetra Patterns: A New Feature Descriptor for Content-Based Image Retrieval," *IEEE Transactions on Image Processing*, Vol. 21,pp.2874 – 2886, May 2012.

[5] Oberoi Ashish, Bakshi Varun, Sharma Rohini and Singh Manpreet "A Framework for Medical Image Retrieval Using Local Tetra Pattern," *International Journal of Engineering Science & Technology*, Vol. 5, pp.27, Feb2013.

[6] Liangliang Shi , Bin Wu ,Bai Wang and Xuguang Yan "Map/reduce in CBIR application," *International Conference on Computer Science and Network Technology (ICCSNT),* Vol.4 , pp. 2465 – 2468, Dec. 2011.

[7] Chapter 7.Textures [Online]. Available http://csweb.cs.wfu.edu/

[8] Muneto Yamamoto and Kunihiko Kaneko, "Parallel Image Database Processing With MapReduce And Performance Evaluation In Pseudo Distributed Mode*," International Journal of Electronic Commerce Studies,*Vol.3, No.2, pp.211-228, 2012.

[9] Venkatraman S.,and Kulkarni S. "MapReduce neural network framework for efficient content based image retrieval from large datasets in the cloud," *12th International Conference on Hybrid Intelligent Systems (HIS),* pp.63 – 68, 4-7 Dec. 2012.

[10] Apache hadoop. [Online]. Available: http://hadoop.apache.org/

[11] Mapreduce - hadoop wiki. [Online]. Available: http://wiki.apache.org/ hadoop/MapReduce.

[12] Hdfs users guide. [Online]. Available: http://hadoop.apache.org/docs/ hdfs/current/hdfs user guide.html.

[13] Apache hbase. [Online]. Available: http://hbase.apache.org/

[14] Khan, S.M.H. , Hussain, A. , and Alshaikhli, I.F.T."Comparative Study on Content-Based Image Retrieval (CBIR)" *International Conference on Advanced Computer Science Applications and Technologies* (ACSAT), 2012 .