

## Local Tetra Pattern for Image Retrieval System Using Hadoop

Mrs. Urvashi Trivedi Student of Master  
of engineering Department of Computer  
Engineering, Sigma Institute of  
Technology, Vadodara  
urvashi\_311@yahoo.co.in

Mrs. Kishori Shekoker  
Assistant Professor, Department of  
Computer Engineering, Sigma Institute of  
Technology, Vadodara  
kishori.cs.engg@sigma.ac.in

**Abstract - In today's world, huge quantity of data, in the form of images, is produced through digital cameras, mobile phones and photo editing software. It is important to develop new CBIR techniques which gives effective and scalable result for real time processing of large image collections. Local tetra pattern (LTrP) is used for managing the large database. The local ternary pattern and local binary pattern encode the relationship. LBP and LTP encodes the relationship between referenced pixel and its surrounding neighbors by calculating gray-level difference. Local Tetra Pattern (LTrP) which carries the interrelationship in between the center pixels and its surrounded neighbors of center pixel. The main objective of study is distribution of image data over a large number of nodes over Hadoop using Map Reduce Technique. Hadoop defines a framework which allows processing on distributed large sets across clusters of computer.**

**Keywords Map-Reduce, HDFS, Content-Based Image Retrieval (CBIR), Local Tetra Patterns (LTrPs), Hadoop.**

### I. INTRODUCTION

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.

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.

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]. 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.

In these proposed system we describe local 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. 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].

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 colors in the image and also from the gray tones of themselves for the working of both classification and segmentation. [7].

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].

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. [1], [2].

### Local Tetra Patterns (LTrPs)

A new technique 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 neighbors, based on that its calculated using first order derivative in horizontal and vertical direction of the center gray level pixel denotes the center pixel and denotes the horizontal and vertical pixel of the neighbors of center gray level pixel Let, respectively. The first order derivative at center pixel  $g_c$  can be defined as

$$I_0^1(g_c) = I(g_h) - (g_c)$$

$$I_{90}^1(g_c) = I(g_v) - (g_c)$$

Where  $g_h$  and  $g_v$  are horizontal and vertical neighbors of  $g_c$  respectively

$$I_{Dir}^1(g_c) = \begin{cases} 1, & I_0^1(g_c) \geq 0 \text{ and } I_{90}^1(g_c) \geq 0 \\ 2, & I_0^1(g_c) < 0 \text{ and } I_{90}^1(g_c) \geq 0 \\ 3, & I_0^1(g_c) < 0 \text{ and } I_{90}^1(g_c) < 0 \\ 4, & I_0^1(g_c) \geq 0 \text{ and } I_{90}^1(g_c) < 0 \end{cases} \quad (3)$$

Where,  $I_0^1(g_c)$  - Value of center pixel in horizontal direction,  $I_{90}^1(g_c)$  - value of center pixel in horizontal direction

After identifying the 13 bit binary pattern, their histogram is calculated and query image is compared with the images of given database. During comparison process, the  $N$  best images similar to query image are selected. The experimental results confirmed that LTrP outperforms LBP, LDT and LTP in terms of retrieval and is able to extract more detailed information from the image.

### III. HADOOP

Hadoop provides open source software framework which is used for storage and large scale processing of datasets on clusters. It has two subparts, Map-reduce and HDFS, Map-reduce for computational capabilities and HDFS for storage. Map-reduce is distributed framework for data processing, especially big data. The Map-reduce process of hadoop complete with two phases Map and Reduce. In Map phase stored split data inputted to map function which will generate intermediate key pair .Wherever reduce phase accept these key value pair as its inputs which will merge

all intermediate values associated with same intermediate key. Figure 1 shows architecture of HDFS.

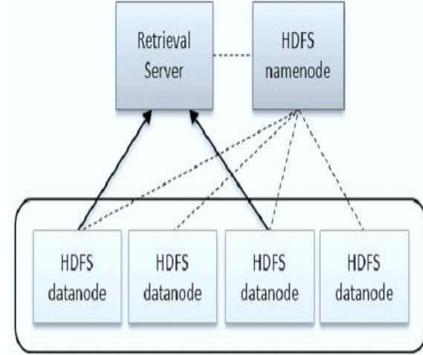


Figure 1: The architecture of HDFS

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 an open-source software project for scalable, reliable, flexible, distributed computing, failure handling [10].

HDFS - 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].

Map Reduce - The parallel framework offered by Map-reduce is highly suitable for proposed CBIR structure with large amount of data. Figure 2 shows the Map reduce technique. We use the open source distributed cloud computing framework hadoop and their implementation of Map-reduce module. Map-Reduce decomposes work submitted by a client into a small parallelized map and reduce jobs, as shown in figure 1. A Hadoop cluster includes multiple worker nodes and a single master. Master node consists of a TaskTracker, Namenode, JobTracker, and Datanode. A worker node acts as both a TaskTracker and Datanode, though it is possible to have compute-only worker nodes and data-only worker nodes. The list of output can then be saved into Distributed file system then the reducer run to merge the result in parallel [11]. Figure 2

shows a multi-node Hadoop cluster. Hadoop provides location awareness compatible file system.

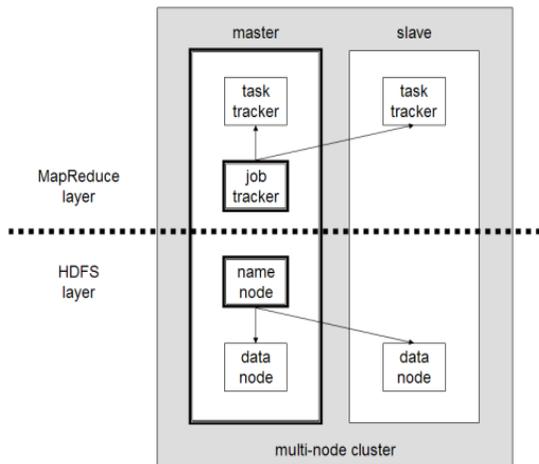


Figure 2 Multi node Hadoop Structure

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].

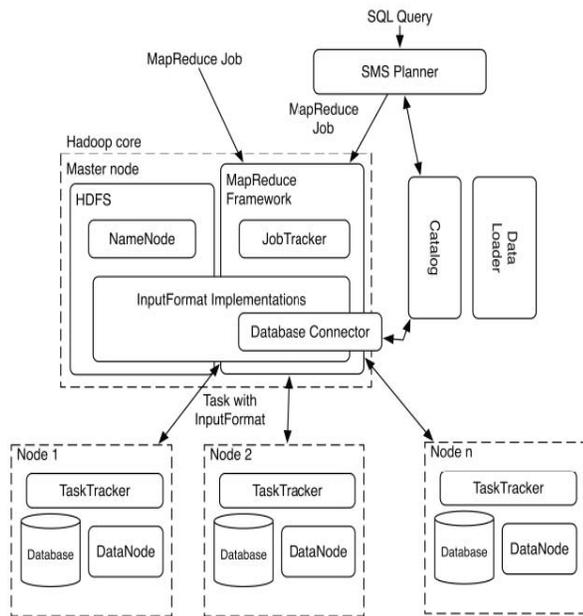


Figure 3 Working Principle of Hadoop Map-Reduce.

Figure 3 shows working principle of hadoop map-reduce. 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].

#### IV. LITERATURE REVIEW ANALYSIS Sitalakshmi

Venkatraman 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 cloud clusters [9].

Ashish Oberoi represent hadoop open source framework based on Local Tetra Pattern for content based image retrieval from medical databases is proposed. It approaches to formulate the interrelationship between the reference or center pixel and from its neighbors, considering the directions i.e. vertical and horizontal calculated using the first-order derivatives [5].

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

R.P.Maheshwari 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 neighbors of center pixel by computing difference of gray level [4].

#### V. PROPOSED SYSTEM DESIGN

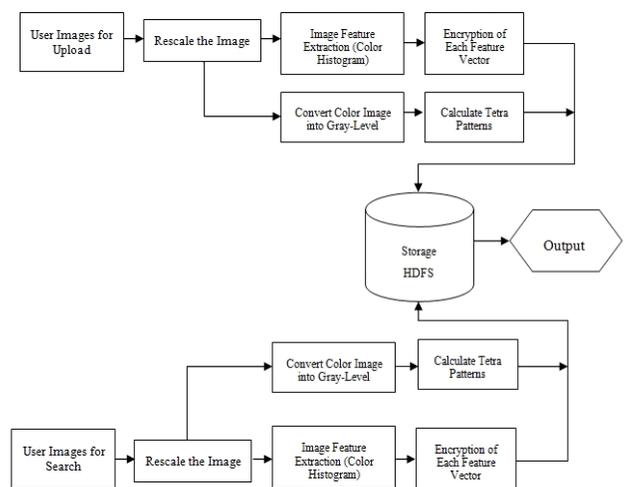


Figure 4 System Architecture

The goal of the proposed system is to detect the most relevant images from the databases. In this paper, the LTrP includes LDP, LBP, LTP and Magnitude Pattern which are used to retrieve feature from the images. 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. CONCLUSIONS

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. 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. Hadoop distributed File system (HDFS) is used to store and retrieve images. Application developed using the proposed approach is fast and efficient in retrieving images. The content based image retrieval algorithm used in the developed application produces accurate results within short span of time and is very reliable. The Hadoop-CBIR developed has immense potential to be used in various fields. This paper presents a content based image retrieval system in Hadoop framework Hadoop has been used in this work to set up a grid in a large scale environment which supports large amount of data processing. It also facilitates accurate retrieval of images matching the queried image. As the proposed image retrieval system is implemented in Hadoop, it is very easy to adapt in cloud environment with minimal overhead.

## VII. REFERENCES

- [1] Hiremath P.S. and Pujari J., “Content Based Image Retrieval Using Shape, Texture, Color, and 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., “LocalTetra Patterns: A New Feature Descriptor for Content-Based Image Retrieval system” 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, 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] Available <http://csweb.cs.wfu.edu/> Chapter 7.Textures [Online].
- [8] Muneto Yamamoto and Kunihiko Kaneko, “Parallel Image Database Processing With MapReduce And Performance Evaluation,” 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] Available: <http://hadoop.apache.org/> Apache hadoop. [Online].
- [11] Available: <http://wiki.apache.org/hadoop/MapReduce>. Mapreduce -hadoop wiki. [Online].
- [12] Hdfs users guide. [Online]. Available: [http://hadoop.apache.org/docs/hdfs/current/hdfs\\_user\\_guide.html](http://hadoop.apache.org/docs/hdfs/current/hdfs_user_guide.html).
- [13] Available: <http://hbase.apache.org/> Apache HBase. [Online].
- [14] Khan, S.M.H. Hussain, A., and Alshaikhli, I.F.T. “Comparative Study on Content-Based Image Retrieval system” International Conference on Advanced Computer Science Applications and Technologies (ACSAT), 2012 .
- [15] Subrahmanyam Murala, R. P. Maheshwari, Member, IEEE, and R. Balasubramanian, Member, IEEE,” Local Tetra Patterns: A New Feature Descriptor for Content-Based Image Retrieval,” IEEE Trans. on Image Processing, vol. 21, no. 5, May 2012.
- [16] Ashish Gupta, “Content Based Medical Image Retrieval Using Texture Descriptor, IJREAS” Volume 2, Issue 2 (February 2012), ISSN: 2249- 3905.