ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A Survey on Content Based Image Retrieval System Using HADOOP

Mrs. Urvashi Trivedi*, Mrs. Kishori Shekoker
*Student, Master of Computer Engineering, Sigma Institute of Engineering, Vadodara, India
Professor, Department of Computer Engineering, Sigma Institute of Engineering, Vadodara, India

Abstract

Content-based image retrieval (CBIR) - an application of computer vision technique, addresses the problem in searching for digital images in large databases. This emerging approach includes the Local Binary Pattern (LBP), Local Derivative Pattern (LDP), Local Ternary Pattern (LTP) and Magnitude Pattern. The ability to handle very large amounts of image data is important for image analysis and retrieval applications. With digital explosion of image databases over internet pose a challenge to retrieve images that are relevant to user query efficiently and accurately. It becomes increasingly important to develop new CBIR techniques that are effective and scalable for real time processing of very large image collections. Content based image retrieval system based on Hadoop, 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 Local Tetra Pattern (LTrPs) for Content Based Image Retrieval (CBIR).

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

Materials and methods

Over the last decade there has been a fast increase in the volume of image and video collections. The information is available in huge amount and it is very difficult to access it. To overcome these kinds of difficulties in the early 1990s, Content-Based Image Retrieval (CBIR) emerged as a promising means for describing and retrieving images. In accordance with CBIR, images are indexed by their visual content, such as color, texture, shape, and spatial layout instead of being manually annotated by text-based keywords.

One of the most important visual attributes in an image is its Shape. Its concept is invariant to translations, rotations and scaling and it is a binary image representing the extent of the object. That's why shape presentation is one of the most challenging aspects of computer vision. Shape representations can be roughly classified in two major categories: boundary-based and region-based. In boundary-based, shape is represented by its outline, while in the region-based, shape being formed of a set of two-dimensional regions. The feature vectors extracted from boundary-based representations provide a richer description of a shape allowing the development of multi-resolution shape description. Another important visual attribute is its Texture, since it is presented almost everywhere in nature. Textures may be described according to their spatial, frequency or perceptual properties.

a. Image Retrieval

Image retrieval is the task of retrieving or getting digital images from a database. Many hospitals use picture archiving and communication system, which is basically a computer network that is used for storage, retrieval, and distribution of image data.
Image retrieval systems differ in the way in which querying and retrieval is done. The possible kinds of queries were already introduced in the introduction:

- Meta information, like the patient's name
- Like a textual description "An X-ray image showing a fracture in the lower left arm".
- Visual information Querying by visual information is called content-based image retrieval (CBIR).

b. CBIR

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. CBIR operates on a totally different principle from keyword indexing. Primitive features characterizing image content, such as color, texture and shape, are computed for both stored and query images, and used to identify the 20 stored images most closely matching the query. Semantic features such as the type of object present in the image are harder to extract, though this remains an active research topic. Video retrieval is a topic of increasing importance here, CBIR techniques are also used to break up long videos into individual shots, extract still key frames summarizing the content of each shot, and search for video clips containing specified types of movement.

Patterns used for image retrieval

1. LBP **-** The Local Binary Pattern operator was introduced for texture classification. Given a center pixel in the image, the LBP value is computed by comparing its gray value with its neighbors. The LBP operator on facial expression analysis and recognition is successfully reported in. Furthermore, the LBP is incorporated into multi-scale heat-kernel face representation for the purpose of capturing texture information of the face appearance. Face image is decomposed into different scale and orientation responses by convolving with multi-scale and multi-orientation Gabor filters. In the second phase, BP analysis is used to describe the neighboring relationship not only in image space but also in different scale and orientation responses. Due to the discriminative power and computational simplicity, LBP texture operator has become a popular approach in various applications. It can be seen as a unifying approach to the traditionally divergent statistical and structural models of texture analysis. Local Derivative pattern explains the feasibility and effectiveness of using high-order local patterns for face representation.

2. LDP - It considered the LBP as the non-directional first-order local pattern operator and extended it to higher order (nth-order) called the Local Derivative Pattern. The LDP contains more detailed discriminative features as compared with the LBP. An LDP operator is proposed, in which based on a binary coding function (n-1) th order derivative is calculated.

3. LTP - It extended the LBP to a three-valued code called the Local Ternary Pattern, in which gray values in the zone of width ±t around g_e are quantized to zero, those above (g_e + t) are quantized to +1, and those below are quantized to 1, i.e., indicator is replaced with three-valued function and the binary LBP code is replaced by a ternary LTP code.

4. LTrP - The idea of local patterns (LBP, LDP and LTP) has been adopted to define LTrPs. The LTrP describes the spatial structure of the local texture using the direction of the center gray pixel.

The LBP, the LDP, and the LTP extract the information based on the distribution of edges, which are coded using only two directions (positive direction or negative direction). Thus, it is evident that the performance of these methods can be improved by differentiating the edges in more than two directions. This observation has motivated us to propose the four direction code, referred to as local tetra patterns (LTrPs) for CBIR.

Advantages of the LTrP over other patterns

The LBP and LDP are able to encode images with only two distinct values (either “0” or “1”) and the LTP is able to encode images with three (“0”, “-1” or “1”) distinct values. However, the LTrP is able to encode images with four distinct values as it is able to extract more detailed information i.e why it is known as Local Tetra Pattern.

The LBP and the LTP encodes the relationship between the gray value of the center pixel and its neighbours, whereas the LTrP encodes the relationship between the center pixel and its neighbours based on directions that are calculated with the help of (n-1) th-order derivatives.

HADOOP

Hadoop is open source software framework for storage and large scale processing of datasets on clusters. Hadoop has two subparts first one is Map-reduce for computational capabilities and second is 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.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

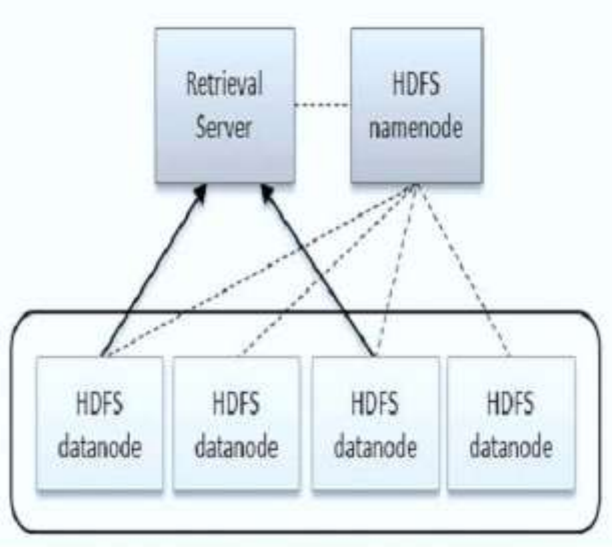

Figure 1: The architecture of HDFS

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 - Hadoop Distributed File System (HDFS) is a subproject of apache Hadoop Project which is designed to provide fault tolerant file system designed to run on commodity hardware. HDFS can create, move, delete or renames the files like traditional file system but the difference is the method of storage because it includes two actors which are NameNode and DataNode. A DataNode store data in HADOOP and NameNode is centerpiece of HDFS .HDFS store data reliably even in presence of failure including NameNode failure, DataNode failure and network partitions. Generally HDFS uses a master/slave architecture in which one device control one or more other devices. HBASE is column-oriented database management system that runs on top of HDFS. Unlike relational database system HBASE does not support a sql. HBASE system comprise a set of tables. Each table contains rows and columns much like a traditional database. The programmer can access a table easily by APIs provided by HBASE.

MapReduce - 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 small Hadoop cluster includes a single master and multiple worker nodes. The master node consists of a JobTracker, TaskTracker, NameNode and DataNode. A slave or worker node acts as both a DataNode and TaskTracker, though it is possible to have data-only worker nodes and compute-only worker nodes. In a larger cluster, the HDFS is managed through a dedicated NameNode server to host the file system index, and a secondary NameNode that can generate snapshots of the NameNode's memory structures, thus preventing file-system corruption and reducing loss of data. Map-reduce framework works in parallel manner which processes on very large image collection of petabyte of storage. Map-reduce is the application that works on the data stored in HDFS and act as resources scheduler. Map-Reduce is a 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]. Figure 2 shows a multi-node Hadoop cluster. Hadoop provides location awareness compatible file system.

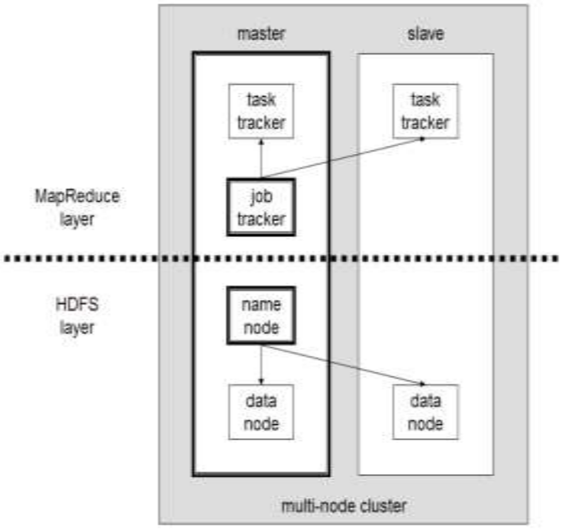

Figure 2 Multinode Hadoop Structure

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 shows working principle of hadoop map-reduce.

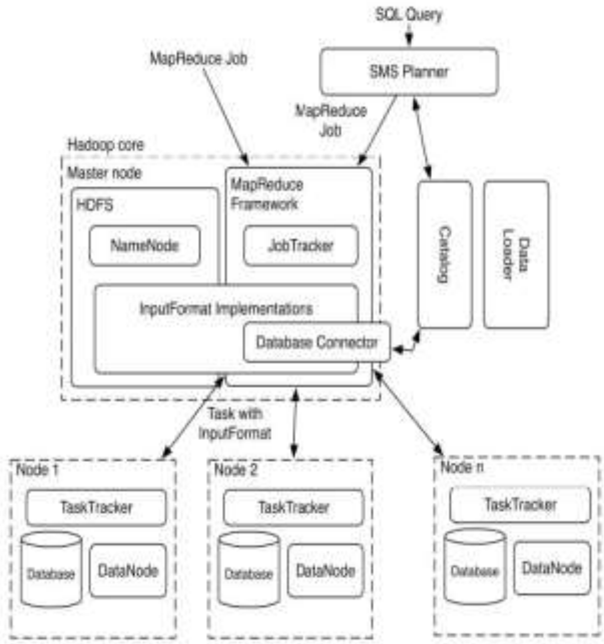

Figure 3 Working Principle of Hadoop Map-Reduce.

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].

Literature review analysis

The Paper presents a 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 [9].

The Paper 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 center pixel and from its neighbors, considering the directions i.e. vertical and horizontal calculated using the first-order derivatives [5].

The Paper has 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 [14].

The Paper the analysis of image indexing and retrieval algorithm using local tetra patterns (LTrPs) for content-based image retrieval (CBIR) has been proposed. The standard local binary pattern (LBP) and local ternary pattern (LTP) encodes the relationship between the referenced pixel and its surrounding neighbors by computing gray-level difference. The proposed method encodes the relationship between the referenced pixel and its neighbors based on the directions that are calculated using the first-order derivatives in vertical and horizontal directions [15].

The Paper proposed a high-order local pattern descriptor, local derivative pattern (LDP), for face recognition. LDP is a general framework to encode directional pattern features based on local derivative variations. The nth order LDP is proposed to encode the (n–1) the order local derivative direction variations, which can capture more detailed information than the first-order local pattern used in local binary pattern (LBP). Different from LBP encoding the relationship between the central point and its neighbors, the LDP templates extract high-order local information by encoding various distinctive spatial relationships contained in a given local region [16].

The Paper proposed the 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].

The Paper 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 [14]

I have researched various papers, In that I surveyed that in today's era the amount of digital images are growing every day in a very explosive manner have to take gigabyte and petabyte of storage. For these, search and retrieve the particular images from that 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. The Map-Reduce Framework (Hadoop) result in real time efficiency even in large image collection occupying the gigabyte and petabyte storage of database.

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 peta byte. 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.

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.

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., "LocalTetra 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/h dfs 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 .
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.