# Content Based Image Retrieval System for Medical Diagnosis using Had loop

**Chandrashekhar Mahakulkar[1] Mangesh Jagtap[2] Nitin Harale[3] Sandeep Babbur[4] Prof. Avinash Palave[5]**

[1,2,3,4,5]Trinity College of Engineering and Research, Pune

*Abstract*— Heavy usage of internet causes the explosive growth of images. It's impossible to maintain and handle amount of images, pictures, gifs using the conventional technique. New methodologies and techniques are required to handle huge amount of database and which must provide the more promising results to the user. Hence, we are introducing "Content Based Image Retrieval for medical purposes and diagnosis using Hadoop". The main objective of this is to handle a humongous and large amount of data with the principle of fragmentation, Map Reduce, shape, colour etc.

***Key words:*** Map Reduce, CBIR, fragmentation, HDFS, medical Database

## I. INTRODUCTION

Before the introduction of content based image retrieval system, there is a method called Text Based different images Retrieval system is used previously. In this method, images are obtained on the basis of ids or tags allotted with the image. The main issue with this technique is, we have to assign the id to every image in the database manually and that is almost impossible to task to assign id to huge amount of images and it is not impossible to memorize all those ids. That's why this technique called content based diffrent images retrieval system has been introduced. In this technique the attributes of the images are used as search criteria. These attributes are automatically evaluating different features.

Content Based Image Retrieval is the technique for recollecting a specific image from the image database based on the content of the image. Here the content is nothing but the attributes associated with that image. There are so many kinds of attributes like colours, textures, shapes etc. On the basis of these attributes, the images are obtained from the image database according to the user made query. And that's why the content based image retrieval technique for image retrieval as in comparison to Text Based different kinds of Images Retrieval.

In this system which we are trying put a simplified approach for effective retrieval of images based on MapReduce technique which is generally used for parallel processing and gives the result in minimum time. In the proposed system we are evaluating colours, shapes, textures attribute of the images.

## II. PRELIMINARIES

### *A. Content Based Image Retrieval*

The Content Based Image Retrieval is the issue of obtaining the images from the database by analysing the attributes of the images rather than its metadata. The attributes might be primal features such as colours, textures, shapes etc.

### *B. Colour*

Colour is the one of the fundamental attribute which is used in image obtaining process. Colours are generally defined in three dimensional views. These are listed RGB, HSV, HSB (Hue, Saturation, and Brightness). Image formats such as EXIF, TIFF, GIF, WEBB Poses the RGB colour technique to store information. Colour information of the image is stored and kept in the form of colour different histograms.

### *C. Texture*

Texture is nothing but the visual attribute associated with the image. Texture is used to differentiate out of the areas of image with same colour. Different texture features like degree of contrast level, spatiality and meticulously, coarseness, directionality and randomness etc. This methodology that we are using for analysis of the texture and it is related to the colour co-occurrence.

### *D. Shape*

Shape is one of the lowest level attribute of the image which is used to measuring the shape of specific object associated with that image or database. The natural objects are generally recognized by their shapes, patterns or other attributes. There are different types of shape features such as circularity(circle), convexity (concave and convex), Lake

### *E. Attribute Extraction*

Attribute extraction is the process of conversion in which the input data into set of attributes. In attribute extraction technique, colours, textures and shapes attributes of the image are extracted. MATLAB, Octave, SciPy, Sage these are the software available for attribute extraction.

### *F. Similarity Evaluation*

Similarity evolution is the process of finding our inexactly the solution based on computation of resemblance function between a pair of images or multiple images. Euclidean distance is the technique for similarity evolution in which the distance and space between two points are calculated. In reference to the distance the images are obtained. The images with the less difference and more exactness are provided as an output.

### *G. Hadoop*

Hadoop database is open source software for storage, maintained and large scale processing of datasets on classification of clusters. Hadoop has two subparts first one is MapReduce does computational potential and second is HDFS for storage. MapReduce signally deals with the distributed framework for data processing of images especially called big data. This MapReduce process of Hadoop completes with two sections Map and Reduce.

In Map section stored split data is given as input to map function which generates intermediate key or ids pair. Wherever reduce section accept these key or ids value pair

 

as its input which makes to merge all intermediate in between values associated with same intermediate key or id.

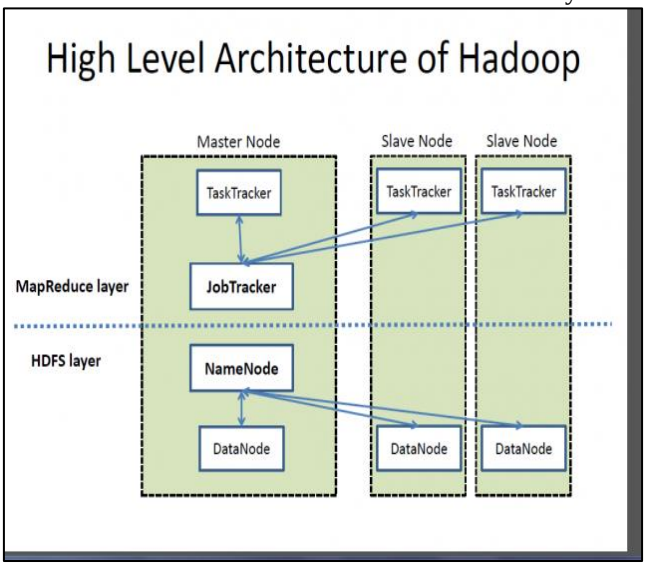


Fig. 1: High Level Architecture of Hadoop

### *H. HDFS*

HDFS generally referred as Hadoop Distributed File System. In HDFS, data is divided into small units or chunks. HDFS is mainly sub exists of Name node and data node. HDFS uses the master/slave type of architecture where Name node act as master of file system where as data node act as slave of file system. Data node stores actual dataset and name node stores the metadata or related key or id for the data node. Name node and data node both are duplicated to handle the if not getting any success and to provide the reliability.

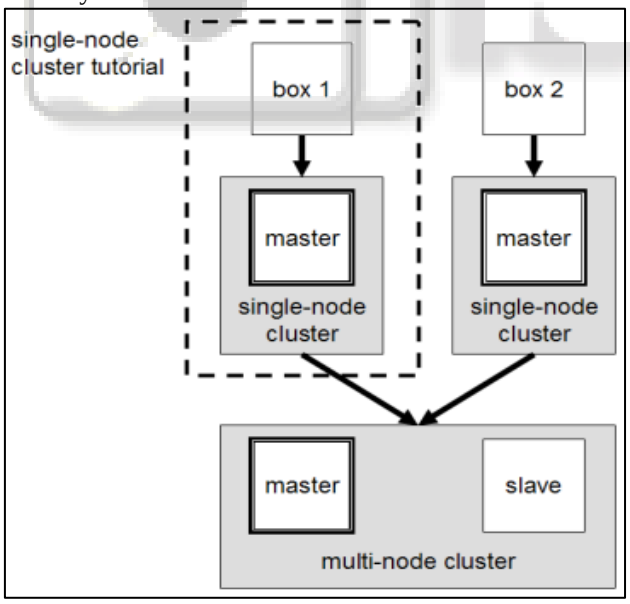


Fig. 2: Single-node Cluster Tutorial

### *I. Map Reduce*

Map Reduce is a programming technology used processing and evaluating large datasets with parallel and distributed algorithms on different cluster. Map reduce program is existing of two basic methods which are Map () method which performs filtering, sorting while a Reduce () method performs a summary information.

Map Step: The master node takes the input values, divide it into smaller sub-problems or units, and distribute them to working nodes. A working node may do this method again, which might lead to multilevel tree structure. The working node processes the smaller problems that is they resolve smaller chunks first and passes the answer back to the master node.

Reduce Step: The master node then gathers all the relevant data which obtained master node and all sub-problems and combines them to give us the final output.

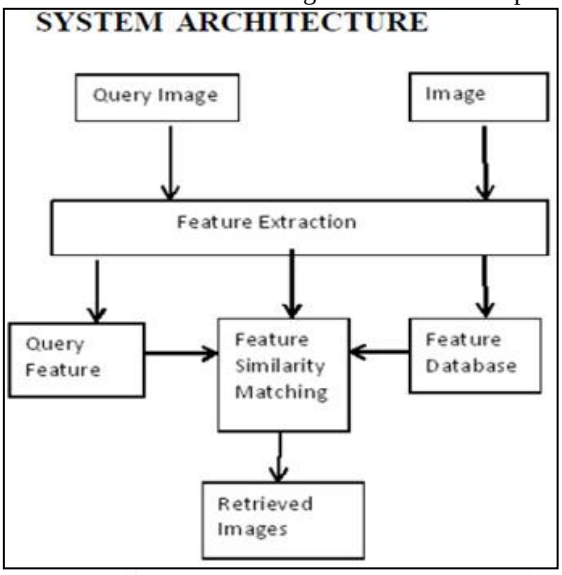


Fig. 3: System architecture

This above diagram shows the of idea system architecture. In this, user will upload the image. Attributes of that image are evaluated such as colours, attributes, texture attributes and shape attributes. These attributes are evaluated collaterally by using the Hadoop Map Reduce technique and are stored in the Hadoop database which might be in the form of attribute vector. Whenever the user provides any query image, the attributes of the query image is evaluated and these attributes are estimated with the attribute vector in the database. These images with the minimum space or less distance are provided as a final result to the user.

## III. CONCLUSION

With the advent of various search engines, image searching has become an easier task. But all the search engines use text based retrieval techniques. Though CBIR is a happening topic, we cannot expect the entire upheaval of existing techniques with CBIR. But certainly, CBIR can be used to complement the existing machinery to provide better results. The CBIR methods presented herein use low- level features to generate results. The purpose of this paper was to improve the exactness and to give clarity regarding CBIR application by allowing the system to obtain more images similar to the source image. The new algorithms under research and also the recently published ones seem to be extremely invasive on the image. Also each new algorithm is always seen to have certain regions where it works best and poor. The proposed technique might increase the precision from an average of 44% to an average of 72%.

## ACKNOWLEDGMENT

We express my sincere gratitude towards my guide Prof: Avinash Palve his constant help, encouragement and inspiration throughout. Without his parental guidance, this work would never have been a successful one. I also like to convey my sincere gratitude to Prof: S S. Choudhary (HOD), all faculty members and staff of Department of Computer Engineering, TRCOE, Pune for all necessary cooperation in the accomplishment of this project. Finally, I would like conclude this by thanking my family and friends.

## REFERENCES

[1] Ludovic Paulhac, Pascal Makriss, and Jean-Yvees Rumel. diffrent solid textures of the database for segmentation and classification experiments. In in 4th International Conference on Computational Vision Theories and Applications, pages 135–141, February 2009.
[2] Jayashree Kalpathy-Cramer, Henning M¨uller, Stevens Bedrickk, Ivan Eggell, Alba Garc´ıa Seco de Herrera, and Theodora Tsikrika. The CLEF 2011 medical image obtaining and classification of different tasks. In Working Points of CLEEF 2012 (Cross Language Evaluation Forum), September 2011.
[3] Katherine P. Andriole, Jeremy M. Wolfe, and Ramin Khorasani. Optimizing analysis, imagination and navigation system related to large image data sets: One 6000–section CT scan can ruin your whole day. Radiology, 259(2):346–362, May 2011.
[4] J. Zhang, X. Liu, J. Luo, and B. Lang, "Dirs.: Distributed image retrieval system of images based on MapReduce and Hadoop technique," in Pervasive Computational complexities and Applications, 2010 6th International Conference on, 2010.
[5] G. Pass and R. Zabih, "Histogram Refinement for Image Retrieval system based on the contents," 4th IEEE Workshop on Applications and queries of Computer Vision and imagination, pp. 96-102, 1996.Amarnath Gupta and Ramesh . Visual data retrieval. Communications of the ACM and related applications, 40(5):70–79, May 1997.
[6] Chapter 7. Textures [Online]. Available http://csweb.cs.wfu.edu/
[7] J. Shashank, "Content Based Image Retrieval Using Colour, shape and colour features," International Conference on Advanced studies computations and Communications, pp.780 – 784,2007.
[8] Chapter 8. Textures [Online]. Available http://cswebs.cs.wfu.edu/
[9] Apache hbase. [Online]. Available: http://hbase.apache.org/
[10] Rodrigos . Calheiroos, Ranjan, Anton Belogllazov, C´esar A. F. De Rose, and Rajkumar Buyya. Cloudsim: diffrent toolkits for modeling and simulation of cloud computing environments and evaluation of souce providing algorithms. Software: Practice and Experience, 41(1):23–50, January 2011.
[11] Zhang W, Dickinson S, Sclaroff S, Feldman J, Dunn S. Shapebasedindexing in a medical image database. Procs. IEEEWorkshop on Biomedical Image Analysis, Los Alamitos, CAs:IEEE Computer Engineering Society 1999; pp. 221-230.
[12] C. Carson, M. Thomas, S. Belongiee, P.M. Hellersteins, P.Malik, Blobworld: a system for regionbased image indexingand retrieval, in: D.P. Huijsmans, A.M.M. Smeulders(Edss.), Proceedings of the Third International ConferenceOn Visual Information Systems (VISUAL'99), no. 1614 inLecture Notes in Computer Science, Springer-Verlag, Amsterdam, The Netherlands, 1999, pp. 509—516.
[13] Chang, C.-L., Cheng, B.-W., & Su, J.-L. (2004). Using case-based reasoningto establish a continuing care information system of discharge. Expert Systems with diffrent Applications, 26(4), 601-613.

 