Volume 4, Issue 12, December 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Retrieval of Images Using Map Reduce

Hinge Smita, Gaikwad Monika, Chincholkar Shraddha
Department of Computer Science and Engineering,
Savitribai Phule Pune University

Abstract- Since arrival of CBIR system, there is an issue of retrieving images that are relevant to users query. Today many websites like google.com, facebook.com etc have lots of images on their servers. Many users access those websites regularly. So there is a need of fast retrieval of these images on users demand. Therefore this paper introduces an effective approach of retrieving images by using Hadoop mapreduce. The main objective of this is processing a large scale of images by the use of parallel processing.

Keywords- CBIR,HDFS

I. INTRODUCTION

Interest in digital images has increased enormously over last few years due to the rapid growth of images on World-Wide-Web. The problem with traditional method of image retrieval is manual annotation. This led to the rise of interest in techniques for retrieving images on the basis of automatically derived features such as color, texture and shape.-a technology now referred to as content based image retrieval.

The field of image retrieval has been an active research area for several decades and has been paid more and more attention in recent years as a result of increased growth of digital images. However, the technology still lacks maturity, and is not yet being used on a significant scale .As with the development of internet; the digital images are increasing at rapid speed. Therefore to retrieve and manage those images has become an important issue. Hadoop is ideal for storing large amounts of data over distributed system. Hadoop is a kind of open source software under the Apache Foundation. For its powerful parallelization, Hadoop has gradually become a popular technique in data mining, searching, recommendation, etc. For a long time, high computation tasks caused by computing complexity and big amount of data during storing, indexing, etc. have been bottlenecks for constructing a CBIR system. Therefore, we study, design and implement a system based on Hadoop to explore the solution to the problem. Hadoop uses HDFS as its storage system. You can then access and store the data files as one seamless file system. Access to data files is handled in a streaming manner, meaning that applications or commands are executed directly using the MapReduce processing mode.

II. PROCEDURE

The model proposed in fig1 consists of two phases:
1] Feature extraction phase
2] Similarity matching phase

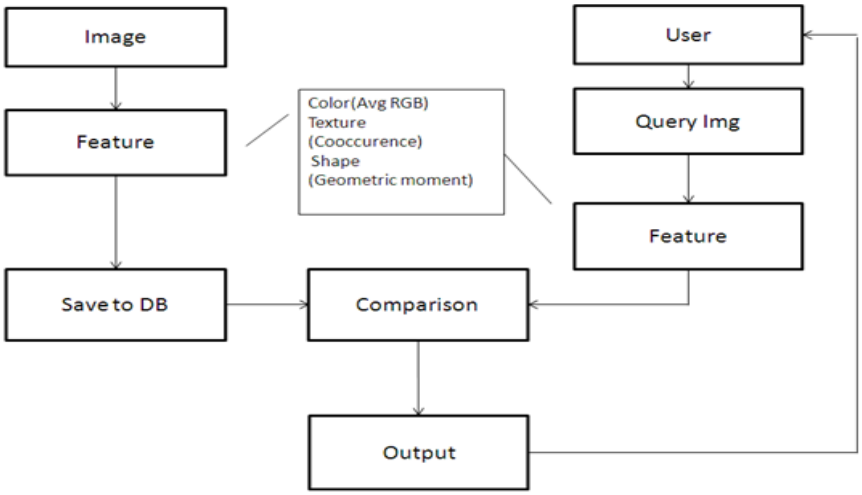


Fig 1: Architectural Diagram for proposed system

Feature is nothing but the content of image that describes the image. The features are extracted from the image in the form of feature vector. The first phase includes extraction of features from all the images in the database and stores it in the form of feature vector in HDFS. The features extracted in the proposed method are color feature, texture feature and

shape feature. These features are extracted parallelly by using hadoop MapReduce. In order to retrieve similar images from the database we must have some similarity measurement technique. The second phase is similarity matching. This phase includes comparison of features of query image with the feature vector in the database. Those images with the less distance are retrieved.

III. DESIGN AND ANALYSIS

We have proposed two phases of the system:

1) Feature Extraction phase
2) Similarity matching phase

A. Feature Extraction Stage:

The various features obtained from the images in the proposed CBIR System are based on Color, texture and shape. All these three features combined, are used to retrieve better images

1) Color Feature:

Each and every color image is having three components: Red(R), Green (G), Blue (B).The features are extracted based on these three components.

Average RGB is the method for extracting the color features from image. We use average RGB to calculate color similarity. Average RGB is to compute the average value in R, G, and B channel of each pixel in an image, and use this as a descriptor of an image for comparison purpose.

Averaging color values is almost identical to averaging numbers, except with the added initial step of finding the red, green and blue components of the color. To do this we can use **bitwise operators.**

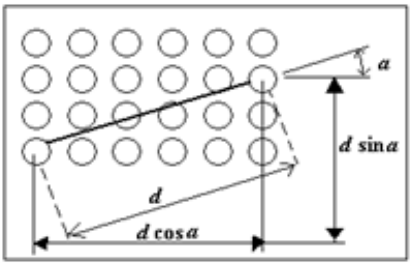
R: Number=pixel>>16 & 0xFF
G: Number=pixel>>8 & 0xFF
B: Number=pixel & 0xFF

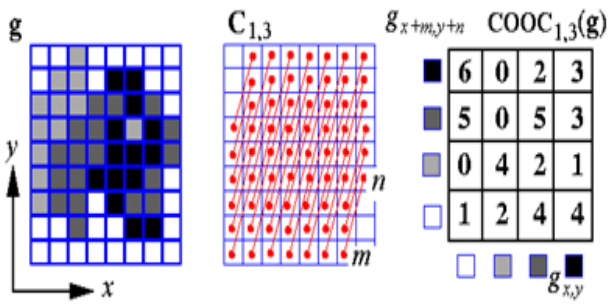
2) Texture Feature:

Image textures are defined as images of natural textured surfaces and artificially created visual patterns, which approach, within certain limits, these natural objects. Image sensors yield additional geometric and optical transformations of the perceived surfaces, and these transformations should not affect a particular class of textures the surface belongs.

Many statistical texture features are based on co-occurrence matrices representing second-order statistics of grey levels in pairs of pixels in an image. The matrices are sufficient statistics of a Markov/Gibbs random field with multiple pair wise pixel interactions.

A co-occurrence matrix shows how frequent is every particular pair of grey levels in the pixel pairs, separated by a certain distance *d* along a certain direction *a*.





3) Shape feature:

Shape descriptions are an important task in content-based image retrieval. It is a mapping that converts the shape space into a vector space and satisfies the requirement that two similar shapes will also have close-to-identical shape descriptors. Fourier descriptors (GD) prove to be more advantageous than other techniques in terms of computation complexity, robustness, easy normalization and retrieval performance. The two-dimensional moment (for short 2-D moment) of a 2-D object R is defined as:

$$m_{pq} = \iint_R x^p y^q f(x,y) dxdy$$

Where f (x,y) is the characteristic function describing the intensity of R, and p+q is the order of the moment. In the discrete case, the double integral is often replaced by a double sum giving as a result:

$$m_{pq} = \sum\sum_R x^p y^q f(x,y)$$

B. Feature comparison stage

The output of the first stage is the features which are extracted from the images. The feature vector of the query image and images int the database are calculated and stored separately.

The Euclidean distance between point's p and q is the length of the line segment connecting them ($\overline{\mathbf{pq}}$).

In Cartesian coordinates, if p = (p_1, p_2,..., p_n) and q = (q_1, q_2,..., q_n) are two points in Euclidean *n*-space, then the distance from p to q, or from qto p is given by:

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2+\cdots+(q_n-p_n)^2} = \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

The position of a point in a Euclidean *n*-space is a Euclidean vector. So, p and q are Euclidean vectors, starting from the origin of the space, and their tips indicate two points. The Euclidean norm, or Euclidean length, or magnitude of a vector measures the length of the vector:

$$\|\mathbf{p}\| = \sqrt{p_1^2+p_2^2+\cdots+p_n^2} = \sqrt{\mathbf{p}\cdot\mathbf{p}}$$

where the last equation involves the dot product.

A vector can be described as a directed line segment from the origin of the Euclidean space (vector tail), to a point in that space (vector tip). If we consider that its length is actually the distance from its tail to its tip, it becomes clear that the Euclidean norm of a vector is just a special case of Euclidean distance: the Euclidean distance between its tail and its tip.

The distance between points p and q may have a direction (e.g. from p to q), so it may be represented by another vector, given by

$$\mathbf{q}-\mathbf{p} = (q_1-p_1, q_2-p_2, \cdots, q_n-p_n)$$

In a three-dimensional space (*n*=3), this is an arrow from p to q, which can be also regarded as the position of q relative to p. It may be also called a displacement vector if p and q represent two positions of the same point at two successive instants of time.

The Euclidean distance between p and q is just the Euclidean length of this distance (or displacement) vector:

$$\|\mathbf{q}-\mathbf{p}\| = \sqrt{(\mathbf{q}-\mathbf{p})\cdot(\mathbf{q}-\mathbf{p})}.$$

which is equivalent to equation 1, and also to:

$$\|\mathbf{q}-\mathbf{p}\| = \sqrt{\|\mathbf{p}\|^2+\|\mathbf{q}\|^2-2\mathbf{p}\cdot\mathbf{q}}.$$

1) One dimension

In one dimension, the distance between two points on the real line is the absolute value of their numerical difference. Thus if *x* and *y*are two points on the real line, then the distance between them is given by:

$$\sqrt{(x-y)^2} = |x-y|.$$

In one dimension, there is a single homogeneous, translation-invariant metric (in other words, a distance that is induced by a norm), up to a scale factor of length, which is the Euclidean distance. In higher dimensions there are other possible norms.

2) Two dimensions

In the Euclidean plane, if p = (p_1, p_2) and q = (q_1, q_2) then the distance is given by

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \sqrt{(p_1-q_1)^2+(p_2-q_2)^2}.$$

This is equivalent to the Pythagorean Theorem.

Alternatively, it follows from (*2*) that if the polar coordinates of the point p are (r_1, θ_1) and those of q are (r_2, θ_2), then the distance between the points is

$$\sqrt{r_1^2+r_2^2-2r_1r_2\cos(\theta_1-\theta_2)}.$$

3) Three dimensions

In three-dimensional Euclidean space, the distance is

$$d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}.$$

4) N dimensions

In general, for an n-dimensional space, the distance is

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_i - q_i)^2 + \ldots + (p_n - q_n)^2}.$$

IV. CONCLUSIONS

In this paper, we have represented an effective approach for content based image retrieval by combining color, texture and shape features and using Hadoop MapReduce for parallel processing. The proposed CBIR method over HADOOP framework environment used MapReduce processing to cover thousands of images distributed over internet. Huge of image are added to image database for application like pattern recognition ,medical ,arts, security etc. Due to huge capability of HADOOP storage we can store big image database in HDFS. This is very important to retrieve similar images effectively stored at various nodes using CBIR based on color feature. We can also extend proposed work of CBIR using different features like shape and texture etc.

ACKNOWLEDGMENT

We wish to avail this opportunity to acknowledge our profound indebtedness and extend our deep sense of gratitude to our guide Prof. Ajay K. Gupta, Assistant Professor , IOKCOE for his valuable guidance profound advice and encouragement that has feel to the successful completion of this project.

REFERENCES

[1] Prof.Deepti Chikmurge ”Implementation of CBIR Using MapReduce Over HADOOP“,“International Journal of Computer, Information Technology & Bioinformatics (IJCITB)” ISSN: 2278-7593, Volume-2, Issue-2.

[2] Swapnil P. Dravyakar et al ”Private Content Based Multimedia Information Retrieval Using Map-Reduce”, “International Journal of Computer Science Engineering and Technology(IJCSET) “| April 2014 | Vol 4, Issue 4,125-128,ISSN-2231-0711.

[3] S. Mangijao Singh , K. Hemachandran, “Content Content-Based Image Retrieval using Color Moment and Gabor Texture Feature”,“IJCSI International Journal of Computer Science Issues”, Vol. 9, Issue 5, No 1, September 2012 ISSN (Online): 1694-0814.

[4] Priyabrat Pattnaik ,”A Survey on Text Based Indexing Techniques in Hadoop“,” International Journal of Advanced Research in Computer Science and Software Engineering “, Volume 3, Issue 11, November 2013 ,ISSN 2277-128X.

[5] Yong Rui Thomas S Huang and Sharad Mehrotra,”Content Based Image Retrieval With Relevance Feedback In Mars”.

[6] Swati V. Sakhare & Vrushali G. Nasre,” Design of Feature Extraction in Content Based Image Retrieval (CBIR) using Color and Texture.

[7] 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,

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

[9] 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

