International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Feature Extraction and Content Based Image Retrieval for High Resolution Remote Sensing Images

T Naga Raju, Chittineni Suneetha

Abstract: ***These are the days where we are very rich in information and poor in data. This is very true in case of image data. Whether it is the case of normal images or satellite images, the image collection is very huge but utilizing those images is of least concern. Extracting features from big images is a very challenging and compute intensive task but if we realize it, it will be very fruitful. CBIR (Content Based Image Retrieval) when used with HRRS (High Resolution Remote Sensing) images will yield with effective data.***

Keywords : Content Based Image Retrieval, Feature Extraction, HDFS, Map-Reduce, Remote Sensing Images

I. INTRODUCTION

Images that are retrieved from remote satellites are referred to as Remote Sensed Images. Such remote sensing images are used in numerous disciplines such as analyzing geography of earth including hydrology, geology, glaciology, oceanography, meteorology, and ecology. It has also a great usage in intelligence reporting, commercial analysis, planning methodologies and humanitarian applications. Remote sensing images refer from the usage of aircraft, satellite based data to identify and classify object on the surface of the earth. These Images are divided into two categories such as active images and passive images. Active images are the one which are captured when a signal is emitted by the respective satellite and it reflection is detected by the sensor. Passive images are those which are captured by the sensor with help of only sunlight. Remote Sensing can extract even data from dangerous, very remote and inaccessible areas of Earth. The images of such high importance are not given much focus. The Remote Sensing images consist of several resolutions such as spatial, radiometric, temporal, and spectral. All the resolutions in one image gives it more worth and information. Extracting correct and useful information from these images is aa very challenging task and very technical intensive task. Pictures are very powerful devices for communication i.e., it is a visual portrayal of unlimited information. If we are successful in extracting that useful information from the picture and then transforming it into data it will be very useful in every aspect. There are different applications of data extracted from Remote Sensing images which are helpful for developing our economical well-being as standard of living every picture has its own semantical values depending on its nature as every picture raises from a distinct area of subject. These images should be properly organised and the system should be very well trained for every sort of picture. Every digital picture is a collection of discrete pixels which are to be carefully analysed to extracted brilliance. CBIR (Content Based Image Retrieval) is one of the most prominent and successful technique to extract information from pictures. Approach can be further improved to be applicable with remote satellite images which are with high definition. This depends on feature extraction procedure of extracting features from an image this procedure should go through machine but not the human intervention in order to save time. Huge amount of data is already available and amount of data is produced every day [1]. The procedure or methods followed to store and transmit images is very old and naive it is also a very well known problem [2, 3]. Even a lot of approaches and methods are established for feature extraction still there is a huge semantic gap between human brain perception of images and computerized perception of images. Human brain is more capable of visually understanding and image accurately in a lesser time when compared to the computer [4]. Huge amount of research should be still dedicated in this area to design procedures to perform accurately in a reasonable time complexity [5]. Our research also contributes in the same way to design a new algorithm which CBIR for Remote Sensing images.The currently available systems for CBIR can only extract limited feature sets which confine the output retrieval efficiency rate [6]. The first step is to extract maximum possible features from images and next step is to utilize those features i.e., after features are extracted from the images that features are to be classified and analysed to extract Useful information from images. This involves huge computational power and challenging step towards data extraction [7]. The feature similarity comparison is achieved through machine learning algorithms in which the machine is trained with huge variety of images. After the machine is fully trained and equipped with all sort of images the machine behaves well and classify features and extract valuable information from Remote Sensing images.

Manuscript published on 30 September 2019
***** Correspondence Author
T Naga Raju*, Research Scholar, Acharya Nagarjuna University, Guntur, India.
Dr. Chittineni Suneetha, Department of MCA, RVR&JC College of Engineering, Guntur, India.

II. RELATED WORK

Feature extraction and CBIR is an active field for multimedia community research since Inception of 1990. This process was very ambiguous because of huge image datasets and diversity of image contents in its early stages. But there on it received huge attention [8].

Das et al. [9] proposed an approach where extracting the features of an image is achieved through image binarization. This enhances retrieval of images and its identification using content based image recognition. Latest update system using two public data sets and reduce the sizes of features to 12:00 regardless of the dimensions of the image. This is not suitable for high resolution images as binarization of those images is hectic task then training the system.

Seetharaman and SelvaRaj [10] came up with a method for image retrieval buy statistical tests they are applicable for structured and textured images as input. Hate the entire image was considered to as a textured image where as for structured images the shape is separated into different regions depending upon its nature. If this approach is applied to big images the images cannot be divided into structured and textured images.

Madhavi et al. [11] propose an approach which is called as image retrieval using interactive genetic algorithm which is used to calculate selective features of an image and then comparing them to related images against those features. The approach was tested to prove the efficiency of the proposed approach. The test set was about 10,000 images. To implement this methodology for extracting features from high resolution images is nearly impossible as the test set it selves forms a huge burden to handle.

Zeng [12] et al. came up with a procedure where a local structure descriptor is designed for retrieval of images. Local structure descriptor is designed based upon the local structures which are beneath colors; it has a combination colour texture and shape as a single unit of image retrieval. They also proposed algorithm for extracting features which can extract histogram for local structure using its local structure descriptor. As we use only high resolution images and Remote Sensing images in this scenario colour and texture will vary depending upon the time, sunlight, mist, fog and current climatic conditions when the image was clicked. So the local structure descriptors will not be valid every time for the same image.

Ali et al. [4] proposed a robust CBIR approach where the used two algorithms namely Speeded Up Robust Features (SURF) and Scale Invariant Feature Transform (SIFT) extraction of features from images. It is robust to rotation and scale change prayer algorithm is very robust for illumination changes. Both the algorithms are used to enhance effectiveness of CBIR approach.They evaluated the method on Corel-1500 etc. But was not at all implemented on high resolution or remote sensing Data images thus it is a dilemma whether the approach works well with big images or not.

Nagaraju et al. [13] described a distributed framework methodology to handle high resolution Remote Sensing images through Hadoop and HDFS. This is an approach where a single high resolution image is handled in no time with distributed framework. The images with split into dimension based and memory based and then uploaded to HDFS via PUT command. This has two phases such as pre-processing and map-reduce. This is a robust way of handling large files by converting and merging of files.

We hereby come up with an approach which extends this idea of Distributed Framework for High Resolution Remote Sensed Images with CBIR for feature extraction and comparison.

III. METHODOLOGY

The following figure describes CBIR methodology for normal images.

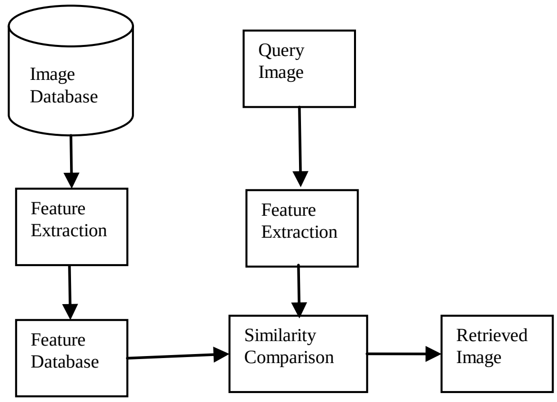


Figure 1: Block diagram of Content Based Image Retrieval System

The above figure describes the methodology followed for content-based image retrieval for normal images. This is only applicable for normal and small size images and it performs very well when used with medium resolution images. Initially all the images for which the machine is to be trained it is saved in a database called as image database then and in the further step maximum possible features from all those images are extracted from the database by the machine by analyzing each and every image from the database. All these features are saved and Stored for further purpose in in another database called as feature database. This feature database is a repository of all the features extracted from the image database. When an image is given as a query to the system all the features from those images are extracted and compared against the feature database which is already built initial steps. After this comparison is successful images with similar features are retrieved by the system. In this process the disadvantages that the machine does not currently learn from the images and neither the image database not the feature database is updated with time. If we apply machine learning algorithms to this approach we can come up with a most powerful and best-in-class methodology for feature extraction from images. This is also not suitable to work with high resolution remote Sensing images because no two images shot in a small period of time will be similar. Thus up gradation of image database as well as the features of those is mandatory in this concern. So we modify the above approach to be applicable for high resolution Remote Sensing images with distributed framework working with content based image retrieval approach.

Algorithm:

Step 1: Accept the Input image and parallelly perform Steps 2 and 7
Step 2: Divide the image into required amount of sub-images
Step 3: Using an Image Processing Tool Pre-process the images
Step 4: Upload every sub-image to HDFS
Step 5: Apply Map-Reduce to the uploaded sub-images
Step 6: Store these images into a huge database
Step 7: Extract maximum possible features from the given sub-images via supervised Learning
Step 8: Store all the extracted features into a Feature database
Step 9: Accept a query image and extract maximum possible features from it
Step 10: Compare these features using K-NN algorithm to the Feature DB which is built
Step 11: Retrieve the required image which is a result of comparison of features
Step 12: Continuously train the system to learn the features from the Image Database

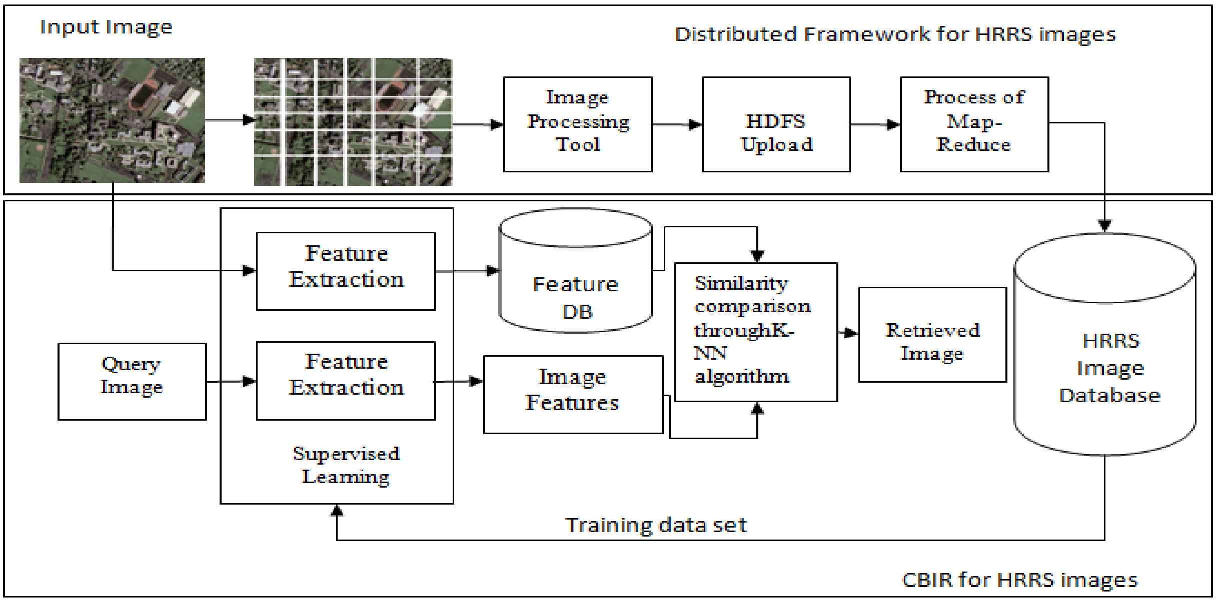


Figure 2: Block Diagram of CBIR approach for distributed framework for HRRS images

Detailed description:

The high definition and high resolution remote Sensing images are taken as input and divided into even sized chunks without any loss of dimension, detail and data. High resolution images cannot be analysed properly in a single go accurately so we divide those images so that it will be easy to handle. The size of each piece of the image can be chosen depending upon the computational power available.

All the image parts are passed through an image processing tool so that any lost dimensions and details are recovered which are lost during division of images. Dimensionality reduction does not apply on satellite images at every detail is important. Image processing tool ensures that whole image can be recovered from summing up image parts so that no valuable information from the image is lost by dividing it.

Every image part is uploaded to Hadoop Distributed File System (HDFS) using HDFS command called as put command. Every image part is referred as a block in HDFS. Every single block is uploaded at a time can be retrieved at a time from the file system. Any operations such as access, upload, modify are done on a block at a time. These are called as block operations in HDFS.

Procedure of Map-Reduce is applied on Hadoop Distributed File System which performs Map-Reduce algorithm on each and every block of HDFS. After map-reduce is applied to every block in the database the result is saved in image database for easy retrieval and access. This image database consists of high resolution remote Sensing Data images which become a huge repository of images retrieved by satellites from Earth containing very useful information.

Accept a query image for which comparison has to be performed and extract maximum possible features from it. The feature extraction procedure for HRRS images as well as for query images is done through supervised learning which a pre-dominant branch of machine learning. Through supervised learning feature extraction procedure performs faster and foolproof also extract as many features by redundantly learning from the previous results. The features which are extracted from HRRS images are stored into a feature database which serves for future.

The features which are extracted from the query image are compared against features from feature database recurrently through K- Nearest Neighbour algorithm which trains the system every time a decision is made. We here apply a machine learning algorithm instead of normal comparison for feature comparison such that the system performs better and can handle any sort of image with any detail. The machine becomes more robust in making decisions for comparisons as it frequently learns from every result. Incessant learning is a predominant feature of this methodology as the system continuously learns whenever it extracts features from input images as well as query images. The repository of HRRS images which is called as image database will act as training data for learning to extract features from the images.

Extracting features from query images from the test data set. By applying this machine learning algorithm such as supervised learning and k nearest neighbour algorithm we make the procedure more powerful and robust in making decisions for feature comparison.

For example if one is interested to know in how many hectares paddy crop is irrigated, we should give a sample of image with paddy field as a query image. The feature and texture from this image is extracted and compared with each and every image which is shot by the satellite from the Earth and give the results in which locations paddy is irrigated after that if we calculate the total are of the images, we can get the appropriate result. The same can be applied to rivers, lakes, forests, populated areas, etc.

Algorithm for K-Nearest Neighbour approach:

1. Every data set must be loaded
2. 'k' value must be initialized
3. Repeatedly iterate initiating from starting point to end point training to derive the output

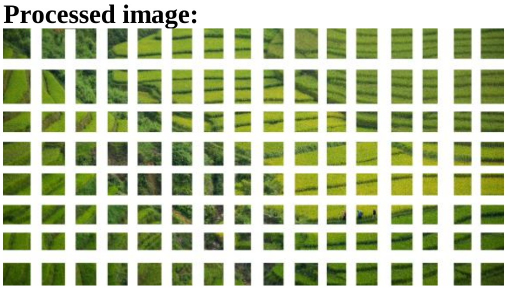
a) Calculate the distance for every training data row with the whole test data
b) Use a distance metric. Mostly used distance metric is Euclidean distance
c) Chebyshev, cosine metrics also can be used as distance metrics
d) According to the derived value of distances, sort out all of them in an ascending order
e) The top-most k-rows are extracted from above sorted array
f) The class of above rows are extracted
g) Return the predicted class (output) [14]

IV. Results

A satellite image of farm lands is taken and the amount of land where paddy is irrigated is extracted.

Input image:

Processed image:



Output:

The satellite image is given as the input which has to be processed and the features are to be extracted. The image is partitioned into equal sized pieces to make it processed by the Map-Reduce programme. Then every piece of the image is retrieved by the CBIR system and all the features of it are extracted then stored into the feature database. After the image is input we posed a query that it should return how many hectares of paddy is irrigated in that image. Every image piece is compared against the paddy feature and the amount of land where paddy is irrigated is calculated and returned by the system. The output is given as 110 hectares of paddy is irrigated and the result is derived in 50ms.

V. Conclusion

Efficient extraction of features from images forms a very crucial role rather than collecting images. Such important task must be performed efficiently and effectively such that all features are properly known. We have come up with a solution which also effectively extracts features but still more faster and parallelly executing algorithms can be developed.

REFERENCES

1. Wan, Shouhong, et al. "Incorporating Spatial Distribution Feature with Local Patterns for Content-Based Image Retrieval." *Chinese Journal of Electronics* 25.5 (2016): 873- 879.
2. Kekre, H., S.D. Thepade, and A. Maloo, Extended Performance Appraise of Image Retrieval Using the Feature Vector as Row Mean of Transformed Column Image.
3. Kekre, H., et al., Improved Shape Content Based Image Retrieval Using Multilevel Block Truncation Coding.
4. A. Ali and S. Sharma, "Content based image retrieval using feature extraction with machine learning," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, 2017,pp.1048-1053. doi : 10.1109/ICCONS.2017.8250625
5. Chun.Y, Kim.N, Jang.I,"Content-Based Image Retrieval Using Multi resolution Color and Texture Features," IEEE Transactions On Multimedia, Vol. 10, No. 6, October 2008, pp.
6. Mutasem K. Alsmadi, An efficient similarity measure for content based image retrieval using memetic algorithm, Egyptian Journal of Basic and Applied Sciences, Volume 4, Issue 2, 2017, Pages 112-122, ISSN 2314-808X, https://doi.org/10.1016/j.ejbas.2017.02.004.
7. Jain, Mihir. Towards efficient and scalable visual processing in images and videos. Diss. International Institute of Information Technology Hyderabad, 2010.
8. Jaime-Castillo S, Medina J M, Sánchez D. A system to perform cbir on x-ray images using soft computing techniques. In Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on, pp. 1314–1319.
9. Das R, Thepade S, Bhattacharya S, Ghosh S. Retrieval architecture with classified query for content based image recognition. ApplComputIntell Soft Comput 2016;2016:2.
10. Seetharaman K, Selvaraj S. Statistical tests of hypothesis based color image retrieval. J Data Anal Info Process 2016;4(02):90.
11. Madhavi KV, Tamilkodi R, Sudha KJ. An innovative method for retrieving relevant images by getting the top-ranked images first using interactive genetic algorithm. ProcComputSci 2016;79:254–61.
12. Zeng Z. A novel local structure descriptor for color image retrieval. Information2016;7(1):9.
13. T. Nagaraju, Dr. Ch. Suneetha, Distributed Framework for Processing High-Resolution Remote Sensing Images, Volume-8 Issue-10, August 2019
14. Sindhu. B, Dr. B. Sujatha,Voice Recognition System through Machine Learning, IJITEE Volume 8, Issue 10, 2019