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

Content Based Image Retrieval using Statistical Parameters of a Medical Image

M. Thilagam, K. Arunesh

Abstract: ***We present a framework that permits in classifying medical images so as to recognize conceivable diseases that affected. This is done by Image retrieval from the collection of dataset by inputting the query image. Content based Image retrieval (CBIR) is the way toward seeking comparable pictures from a picture database dependent on the visual substance of the given query image. Even though some studies present general method in image extraction, there are no efficient methods in medical image retrieval with accuracy. To overcome and to eliminate these flaws our proposed CBIR method examined with the accurate and efficient way for feature extraction from medical images. The images used are grey scale image. The dataset holds the n number of images related to medical particularly brain tumor images. To retrieve the related images from the dataset and get the corresponding details, image is given as an input i.e., query image. Initially, the query image is analyzed by shape, texture and histogram and the result obtained from this is compared with the similar images in dataset. The similarities between the images are found by implementing the Matching Score algorithm. This algorithm provides accuracy in matching the image that helps greatly at the time of classification. The results of computation is said to be the features for the given image. Also the cost for processing the image is comparatively low. The technique has been examined on standard image dataset and satisfactory results have been achieved.***

Index Terms: ***CBIR, Histogram, Matching score, Brain tumor, Image retrieval.***

I. INTRODUCTION

"Content based" in CBIR alludes to analyzing the image by considering it as query. So CBIR is usually recognized as Query by giving image as input. Here the "Content" refers texture, shape, color or other components which can be inferred from the picture itself. Numerous advancements had been made in digital photography that facilitates the user to store high quality images and furthermore helped in expanding the systems speed. The concept of image retrieving is a PC framework where the images are searched in bulk database and needed image is recovered. There are 2 methods in recovering framework, one is text based and other is content based. In most of the cases text based is used where retrieving will rely on image along with keyword. But the drawback of this is irrelevant images are also shown which fails in providing accuracy and also if the accurate keyword is not given, search result would be irrelevant. This minimizes the performance of retrieving and also remains as time consuming process. To overcome this content based image retrieval is used that allows the user to search and retrieve the needed image just by giving the query image.

CBIR [1] undergoes the process of image retrieval by considering the contents that are present visually in image. CBIR does not consider the direct image-image searching mechanism. Such methodology isn't feasible for implementation when it comes in real time applications, the reason is that image data are nearly immense in size. To overcome this in CBIR, reasonable component extraction methods from the image is required with the goal that important and significant images can be recovered based on features that were extricated from the trained images. Various CBIR strategies [2, 3] were created dependent on considering the noteworthy features like texture, histogram and shape [4]. Histogram is generally utilized in CBIR systems because it is the most conspicuous low level feature among others and it additionally remains constant when scaling, rotated or if there is any spatial changes in the pictures. It has been analyzed that the matching histogram based on CBIR strategies is generally straightforward and quicker.

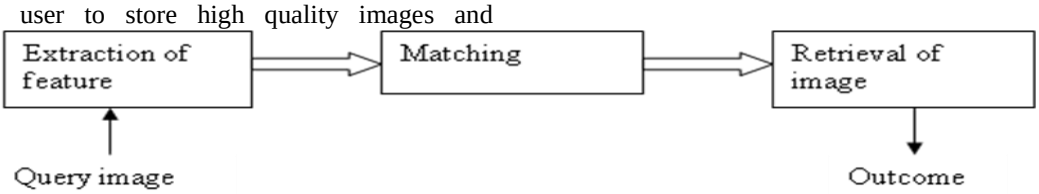

Figure 1. Block diagram of CBIR

Here in our proposed CBIR technique, we have done image acquisition initially, it is gene-rally defined as process retrieving from particular source, here the images are acquired from dataset that holds the collected images related to brain tumor, these images are said to be trained images. After acquiring, the image must be preprocessed to improve the data in image that eliminates unwanted distortions or image features are enhanced to avoid redundancy in upcoming processing.

Manuscript published on 30 September 2019
* Correspondence Author
Ms M. Thilagam*, Department of Computer Science, Sri S Ramasamy Naidu Memorial College, Sattur, Virudhunagar, Tamilnadu, India.
Dr. K. Arunesh, Department of Computer Science, Sri S Ramasamy Naidu Memorial College, Sattur, Virudhunagar, Tamilnadu, India.

The image is converted from RGB to gray scale image, usually image is the combination of Red, Green and Blue and it is converted to grayscale image. In technical terms, 24 bit (i.e., RGB values) is converted to 8 bit (i.e., Grayscale values). After this wavelet transform is done with the factory for 2D, this converts the image to two dimensional formats. The converted image is reshaped to three dimensional matrixes because the image given may in 2D or 3D, to make every image uniform for further processing every image is converted to 3D matrix. The other term Gaussian Projection is used for compressing the retrieval images to be presented in the same page. The image pixels are converted into histogram, a graphical depiction of pixels in image. From this the mean value, standard deviation and variance is calculated for every image. Finally, the matching images based on this value are retrieved for classification.The upcoming sections of this paper deals with: related work in section 2, working method of proposed CBIR in section 3, statistical parameters is presented in section 4, proposed method and obtained results is demonstrated in section 5 and 6. And at last section 7 brings the conclusion to paper.

II. RELATED WORK

Swain and Ballard [5] considered color histogram and comparability measure was finished utilizing histogram convergence distance measurement between the histograms of images. The other histogram based image retrieval was proposed by Malik and Baharum [6] where RGB image is transformed to grayscale and the acquired grayscale was again processed by Laplacian channel. The histogram of processed image was utilized in CBIR. Wang, Yond and Yang [7] presented a strategy for retrieving of images by combining most predominant colors, shape and texture features that depend on pseudo Zernike significance. Liu and Yang [8] utilized color difference histogram in image retrieval. Kekre and Sonawane [9] additionally conceived CBIR strategies utilizing colored histograms and some factual parameters. Murala, Gonde and Maheswari [10] have come up with CBIR that depends on the color histogram and Gabor wavelet. Imran et.al [11] have deteriorated picture into 16 color blocks that are non overlapping and from each block, they have registered different measurable qualities as a part of an feature vector. Singh and Hemachandran [12] have recommended a CBIR dependent on color and Gabor texture component.

III. CONTENT BASED IMAGE RETRIEVAL

CBIR is process of retrieving images considering visual features like histogram, shape and texture. Reason behind its advancement is that in numerous expansive databases of images, conventional strategies for image indexing have demonstrated to be inadequate, relentless, and very tedious. These old strategies for indexing image that ranges from storing and retrieving the images from databases by giving keyword, but this does not provide efficiency in retrieving the images. This is not CBIR process. In CBIR, feature of every image is extracted and stored into the database and these features are compared with the feature of query image and provide the right image that matches the features. There are primary two steps involved,

- **Preprocessing**
- **Feature extraction & Matching**

The initial phase in the process is preprocessing the image by removing the unwanted distortions from the image and in second phase features are extracted from images to a recognizable extent and includes matching those extracted features to yield an outcome that are similar.

A. Preprocessing

Process where the image is enhanced, segmented and converted into the required form are said to be image preprocessing. This helps greatly to extract the features of images accurately. In image enhancement, the contrast and brightness is made clearer. In segmentation, images are divided into segments based on color and segmented to blocks. Finally the image conversion is done where the image is converted to required color space from the actual color space of image.

B. Feature extraction & matching

Here to retrieve the images for the given query image feature extraction of every image in database is done, extraction is based on CBIR. Statistical parameters, histogram, color moment and wavelet transform. Finally similarity matching is done for the given query image.

IV. STATISTICAL PARAMETERS

Execution of any CBIR strategies relies upon the extraction of features from the images. CBIR will give proficient and viable result when all the possible features are considered, yet cost for computation increases subsequently. So from a functional perspective, the element of the feature vector must be chosen appropriately so that the execution will not get affected. Accordingly, in CBIR, determination of feature vector with appropriate dimension is tedious task. Here we have measured some of the statistical feature of the images. This statistical feature creates noteworthy outcomes in the characterization of the datasets. Some normally utilized measurable parameters are mean, standard deviation and gradient. These parameters are figured legitimately from an image histogram.

A. Color Moment

This technique is utilized for color extracting from image. This helps in differentiating images by considering image color. It greatly helps in matching the color of various images. Prior thing in color moment is to calculate color distribution from an image which is said to be probability distribution. First (mean), second (variance) and third (gradient/skewness) order moments are calculated using this. At first, original histogram is changed over into a type of a standardized histogram where x-axis indicates intensity level, y-axis and r_i signifies the evaluated probability $p(r_i)$ of level r_i. The mean (μ) is computed for the intensity value range [LB, UB] as

$$\mu = \sum_{i=LB}^{UB} r_i p(r_i)$$

The Standard Deviation is denoted for the intensity range [LB, UB] as

$$\sigma = \sqrt{\sum_{i=LB}^{UB} (r_i - \mu)^3 p(r_i)}$$

The standard deviation demonstrates the differentiation of the image in the specific noteworthy bin of the image histogram. It is utilized to evaluate the values of intensity distribution about mean in every bin of histogram. When the standard deviation value in a specific image block is less than the other image block values, then it is said that there is maximum contrast in the specific block when compared to other blocks of images.

The gradient or skewness can be computed as

$$\gamma = \frac{1}{\sigma^3}\sum_{i=LB}^{UB}(r_i - \mu)^3 p(r_i)$$

Skewness estimates the level of asymmetry that is exhibited by the data. In case if the skewness becomes zero, then the histogram is symmetric by taking mean into account.

B. Wavelet transform

The method used to recover image by considering texture is known as wavelet transform. Wavelet is the representation of small wave. Wavelet change is the process of transforming a signal into arrangement of wavelets. It changes the image into multi scale portrayal with frequency and spatial attributes. The principle objective is to register the pixel intensity of the images. In this system the image is categorized into 4-sub groups every one of different frequencies that is high-high, low-low, low-high and high-low. Once the sub-band on which the frequencies are thought most is received it is utilized for further image processing. The sub-band energies are obtained. Within this obtained energy range, the distance of the given query image and every other image is determined. The images that are minimum in distance are put at the top places.

When comparing the two images, distance among them is measured using Euclidean distance formula,

$$D(FV_q, FV_{DB_{im}}) = \{\sum_{i=0}^{L-1}(FV_q(i) - FV_{DB_{im}}(i))^2\}^{1/2}$$

Where, feature vector of query image $FV_q(i) = (FV_q(0), FV_q(1), \ldots\ldots FV_q(L-1))$ and feature vector of database image

$$FV_{DB_{im}}(i) = (FV_{DB_{im}}(0), FV_{DB_{im}}(1), \ldots\ldots FV_{DB_{im}}(L-1)),$$

L is the dimension of images feature vector.

Difference among two feature vectors of image is Euclidian distance. Here two feature vectors are created, one for texture and other is for color. The final vector is obtained for this image by combining the two vectors. Similarity difference among the feature vector in the database image and query image is estimated. Database images have Euclidian distance that is equivalent to zero will be a similar as query image. Here we need to locate the common images so we need to set a limit of Euclidian distance.

C. Flow chart of Proposed CBIR

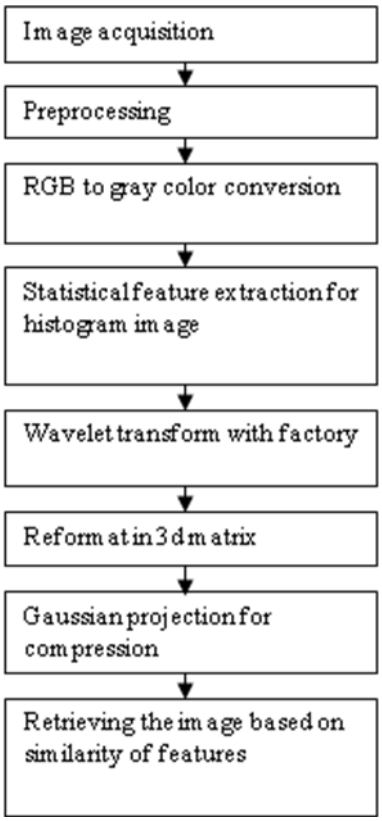

Figure 2. Flowchart of Proposed CBIR

D. CBIR Algorithm

Input: Query image for undergoing comparison is given as input.

Output: Images are undergone with some process and the most alike images to query images are retrieved.

Consider an image IM_q do

1. Image preprocessing.
2. RGB values in the images is converted to Gray scale value.
3. Mean, μ is computed for the intensity value [LB, UB], $\mu = \sum_{i=LB}^{UB} r_i p(r_i)$, then
4. Standard Deviation, σ is represented for the same intensity range [LB, UB],
$$\sigma = \sqrt{\sum_{i=LB}^{UB}(r_i - \mu)^3 p(r_i)},$$
5. Skewness is been formed by the eqn,
$$\gamma = \frac{1}{\sigma^3}\sum_{i=LB}^{UB}(r_i - \mu)^3 p(r_i)$$
6. Creature feature vector FV_q for given query image by color extraction.
7. The same process from step 2 to 5 is repeated for every image in the database DM_{im}, where $DM_{im} = \{IM_{DB1}, IM_{DB2}, \ldots\ldots IM_{DBn}\}$, here n represents the total images in database.
8. Similarity between the images is found by matching score algorithm i.e., the wavelet transform, finding Euclidian distance,

$$D\,(FV_q, FV_{DB_{im}},) = \{\sum_{i=0}^{L-1}(FV_q(i) - FV_{DB_{im}}(i))^2\}^{1/2}$$

9. The images that match the given query image is displayed with Gaussian projection.

V. EXPERIMANETAL RESULTS

The experimental results shows that the images obtained for the given query image is more accurate and the time taken for recovering is less. The matching score algorithm and the steps involved in CBIR shows that the process is undergone flawlessly. Accuracy in recovering images can be very helpful in the classification process that helps in prediction of brain tumor.

The figure 3 shows the query image. Query image can be any image that is chosen from database. The steps mentioned in algorithm is done for the query images also because only the statistical values and wavelet transform is used to compare and recover images that are related to query image. Using the Gaussian compression, the recovered images are displayed accordingly.

Figure 3. Image given as Input

The given query image is transformed into gray scale image from RGB image. The converted gray scale image is depicted in the figure 4.

Figure 4. Gray scale image that is converted from RGB

The images that are recovered from the database after further processing are depicted in figure 5. These are the images that matches the given query image. And these images can be used for prediction of brain tumor.

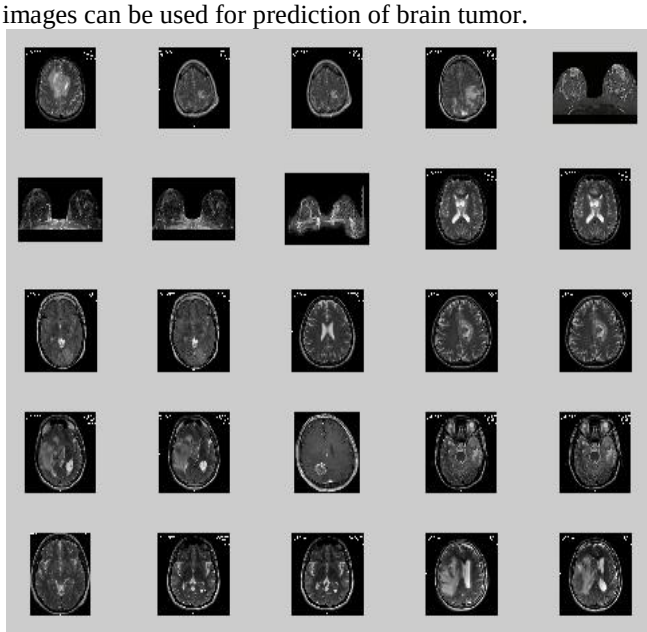

Figure 5. Matching images with Gaussian projection

VI. CONCLUSION

We have proposed an approach to execute content based medical image retrieval because it has gained significant area in retrieving the images. It is an integrated methodology utilized to extract texture and color from images. Here efficient methods and techniques are implemented for analyzing medical images. With the help of single feature, proper results can never be obtained. So there occurs a need to consider multiple features in extracting, this provides more beneficial results in retrieving images. For extraction of color feature, high order color moments are utilized which is said to be color descriptor.

The results obtained are found to be similar to the given query image. The future work will be the enhancement of proposed work that will be done by utilizing different techniques. We will attempt to propose another procedure which will give progressively adequate results. Clustering method will be implemented to decrease the time of searching images. Classification will be performed to enhance the framework execution with the goal that results are accurate.

REFERENCES

1. Deole PA, Longadge R. Content based image retrieval using color feature extraction with knn classification. IJCSMC. 2014 May; 3(5):1274–80.
2. Shete DS, Chavan MS. Content based image retrieval: Review. IJETAE. 2015 Oct; 5(10):1–6.
3. Smeulders, Arnold WM, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. "Content-based image retrieval at the end of the early years." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 12 (2000): 1349-1380.
4. Chaudhari R, AM Patil. Content based image retrieval using color and shape features. IJAREEIE. 2012 Nov; 1(5):1–7.
5. Swain, M. J., & Ballard, D. H. (1991). Color indexing. *International journal of computer vision*, *7*(1), 11-32.
6. Malik, Fazal, and Baharum Bin Baharudin. "Feature analysis of quantized histogram color features for Content-Based image retrieval based on laplacian filter." In *International Conference on System Engineering and Modeling. IACSIT Press, Singapore*, vol. 34. 2012.
7. X. Y. Wang, Y. 1. Yong, and H. Y. Yang, "An effective image retrieval scheme using color, texture and shape features," Computer Standards & Interfaces, vol. 33, no. 1, pp. 59-68,2011.
8. Liu, Guang-Hai, and Jing-Yu Yang "Content-based image retrieval using color difference histogram." *Pattern recognition*46, no. 1 (2013): 188-198.
9. Kekre, H. B., and Kavita Sonawane. "Use of equalized histogram CG on statistical parameters in bins approach for CBIR." In *2013 International Conference on Advances in Technology and Engineering (ICATE)*, pp. 1-6. IEEE, 2013.
10. Murala, Subrahmanyam, Anil Balaji Gonde, and Rudra Prakash Maheshwari. "Color and texture features for image indexing and retrieval." In *2009 IEEE International Advance Computing Conference*, pp. 1411-1416. IEEE, 2009.
11. Imran, Muhammad, Rathiah Hashim, and Noor Eliza Abd Khalid. "Color histogram and first order statistics for content based image retrieval." In *Recent Advances on Soft Computing and Data Mining*, pp. 153-162. Springer, Cham, 2014.
12. K.Vengatesan,M.Kalaivanan Assistant Professor, Department of CSE Published a paper on, "Recommendation System Based on Statistical Analysis of Ranking From User" Information Communication and Embedded Systems (ICICES) 29 April 2013, pp. 479-484 .
13. Singh, S. Mangijao, and K. Hemachandran. "Content-based image retrieval using color moment and gabor texture feature." *International Journal of Computer Science Issues (IJCSI)* 9, no. 5 (2012): 299.
14. Singha, Manimala, and K. Hemachandran. "Content based image retrieval using color and texture." *Signal & Image Processing* 3, no. 1 (2012): 39.
15. Shriram, K. V., P. L. K. Priyadarsini, and V. Subashri. "An Efficient and Generalized approach for Content Based Image Retrieval in MatLab." *International Journal of Image, Graphics and Signal Processing* 4, no. 4 (2012): 42.
16. Pattanaik S, Bhalke DG. Beginners to content based image retrieval. International Journal of Scientific Research Engineering and Technology. 2012 May; 1(2):040–4. 10.
17. Kumar, V.D.A., Kumar, V.D.A., Malathi, S, Vengatesan.K,Ramakrishnan.M "Facial Recognition System for Suspect Identification Using a Surveillance Camera" Pattern Recognition and Image Analysis ,July 2018, Volume 28, Issue 3, pp 410–420

AUTHORS PROFILE

Ms M.Thilagam, Assistant Professor, Department of Computer Science, Sri S Ramasamy Naidu Memorial College, Sattur, Virudhunagar District, Tamilnadu, India.
Mail-Id: thilagamca2016@gmail.com
Pursing Ph.D in image processing and data mining from Madurai Kamaraj University, Madurai. She has completed M.Phil Computer Science in manonmaniam sundaranar university, Tirunelveli in 2014. She has 9 years of Teaching Experience in Atrs and Science College and Published 10 articles in various journals and international conferences.

Dr. K. Arunesh, Dean & Associate Professor, Department of Computer Science, Sri S Ramasamy Naidu Memorial College, Sattur. Virudhunagar District, Tamilnadu, India. Mail-Id: arunesh_naga@yahoo.com
Dr. K. Arunesh has received Ph.D. in Computer Science from Bharathidasan University, Tiruchirappalli, Tamil Nadu, India. He has a teaching excellence of 29 years. He is currently an Associate Professor of Computer Science and Dean, Academic Affairs at Sri.S.R.N.M College, Sattur, Tamil Nadu, India. His current research interests include Knowledge Discovery in Databases, Big Data Analytics, Data Mining and Recommender Systems. He has acted as a resource person and delivered many guest lecturers in his areas of Interest. He has published his research articles widely in many reputed International journals including Scopus indexed journals and International conferences. He has been a conveyor of International Conferences and state level workshops. He has acted as a conference chairperson and he is leading and serving in Program committees. He is currently holding positions on the editorial boards, Subject Expert and Board of studies and also acting as a reviewer of reputed International journals and conferences. He has guided many M.Phil scholars and at present he is a Ph.D research supervisor in Madurai Kamaraj University. He is rendering his service to society is a spark of inspiration to the students.