International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

Content-Based Image Retrieval Using Discrete Wavelet Transform And Convolutional Neural Network

Mohandass S[1], Madhu Shre R K[2]

[1]*Department of Electronics and Communication Engineering, PSG College of Technology*
[2]*Coimbatore, IndiaDepartment of Electronics and Communication Engineering, PSG College of Technology, Coimbatore, India*

Email:[1]*smd.ece@psgtech.ac.in,* [2]*madhukamal97@gmail.com*

Abstract— Due to the development of internet and digital media techniques the size of digital image collection is increasing rapidly. Various techniques for storing, searching and retrieving images become essential for large image archives. This leads to an existence of Image Retrieval System. Content-Based Image Retrieval (CBIR) is a process of retrieving the expected images from large image databases based on the features of the query image and this is performed by various feature extraction techniques. In this work, two different methods of CBIR have been proposed. The first method is CBIR using Discrete Wavelet Transform (DWT), which extracts the texture feature of an image and histogram extracts the color feature of an image. And also the results will be ranked according to the chi-square distance calculation. The second method is CBIR using Convolutional Neural Network (CNN). In CNN method the features are extracted using VGG16 pre-trained model and canny edge detection is used for detecting the edges of an image. Finally, the performances of DWT and CNN based methods are analyzed and compared.

Keywords— DWT, histogram, chi-square distance calculation, CNN, VGG16 model, canny edge detection

1. INTRODUCTION

Large collection of images occur in many applications like stock photos and footage, military, world-wide web, medical imaging and many other applications. This leads to expanding the area of interest in computer vison to find a particular image and to locate those images to meet some descriptions. The old method like traditional text-based image searching method leads to problem of annotating large volumes of databases accurately not feasible, remarks often ambiguous due to human perception and valid for a particular language only. It necessiates too much responsibilities on the end-user. Later, the term Content-Based Image Retrieval has been originated in 1992, which was coined by Toshikazu Kato [1] to describe experiments related to automatic retrieval of images from a database, based on the color, shape and texture present in the image. It has been used to describe the process of retrieving the expected images from a large database on the basis of image features. CBIR is the application of computer vision algorithm that is also known as Query

by Image Content. That is the process of searching for images in large databases according to the contents of the query image. The term content mainly refers to colors, shapes, textures, or any other information that can be derived from the image itself. In this paper an attempt has been made to search an image from large dataset and rank the retrieved results based on the similarity of the images. Content-Based Image Retrieval is to excerpt visual content of an image like color, shape or texture of an image and they are defined spatially.

2. LITERATURE SURVEY

Dileshwar Patel et al [2] proposed the image retrieval system that uses both local and global shape features for retrieve the most similar images from the database. To obtain both features, some pre-processing steps, like object segmentation using Minimum Error Thresholding and border extraction, are first carried out. To extract the global shape feature the Grid Based method is used. In this work the image is divided into smaller areas and extracts local features by applying DWT and singular value decomposition. Finally, it computes the similarities between the global and local features of the query image and all the images in the database, to give the most possible matches as a result.

Domonkos Varga et al [3] proposed a simple but effective deep learning framework based on CNN and Support Vector Machine. The first and the most important part of CNN is the convolutional phase. It works as an extractor of image features. An image is passed through several stages of filters or convolution kernel, creating new images called convolution maps. Some of the intermediate filters reduce the resolution of the image by using a local maximizing operation. Therefore, the created convolution maps get flattened and concatenated to obtain the feature vector values, called as CNN code. The SVM takes this CNN code from the output of the convolution phase as a new feature vector for training.

In [4] the edge of image is derived from the YCbCr matrix using the canny edge detection method and the RGB histogram is calculated as a global statistical illustration of color distribution representing the image. To further enhance the image representation capabilities, color features are also incorporated with the histogram and applied Haar wavelet transform to effectively reduce computational steps and help improve search speed. The Manhattan distance is applied to retrieve the similar images that are placed in the image dataset.

In [5], Akshara Preethy Byju et al provided the fully comprehensive study of the recent Content Based Image Retrieval which is used to design a mechanism for image retrieval to get better retrieval results with lesser computational complexity. From the survey, it has been summarized that, Color Moment, Color Histogram, Color Coherence Vector, HSV histogram and Color Descriptor etc., techniques are used to extract color features. Haar Wavelet Transform, Gabor Wavelet Transform, Discrete wavelet transforms, and Gray Level Co-Occurrence Matrix etc., are used for extracting texture features. Canny Edge Detection, Edge Detection Histogram, and edge histogram descriptor etc., are used to extract shape features. Similarity measurement is carried on by Euclidean distance, Chi- square Distance, Wavelet Depose, Naïve Bayes, and K-Means Clustering methods.

3. PROPOSED METHODOLOGY

CBIR is also known as Query by Image Content [6] because it involves retrieving of expected images from the large database according to the contents of the query image. Contents mainly refer to the Color, Texture and shape of the query image which can be automatically extracted from the images using feature extraction techniques. Here, two methods are used for feature extraction they are DWT and CNN. In proposed methodology the images will be stored in a database and the features such as Color, Texture and Shape will

be extracted from the stored database and then a query image will be given so that again the features of the query image will be extracted. Comparing both the extracted features of query image and database the result will be obtained according to the similarities. In this proposed work the database consists of 500 images of different classes such as Rose, Tiger, Parrot, Ship and Building. From these the features will be extracted. Fig.1, is the block diagram of proposed methodology.

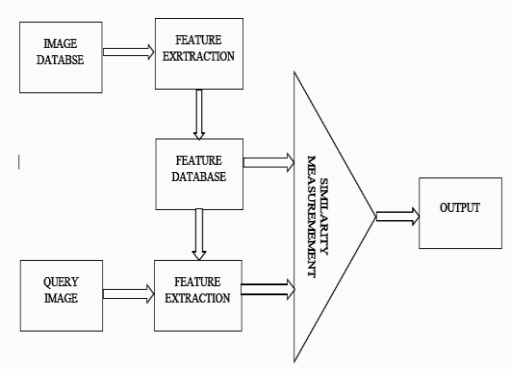

Fig.1 Block diagram of proposed methodology

a. Wavelet Decomposition Method

Discrete Wavelet Transform is used for feature extraction method [7]. In DWT an image of size (M x M) is decomposed by low-pass and high-pass filters on row and column wise operation and down sampled by 2. After the decomposition process four output images of equal size will be produced such as Approximated Coefficient, Horizontal Coefficient, Vertical Coefficient and Diagonal Coefficient of an image at different resolution levels. As resolution of image increases, it requires lot of disk space. So, DWT is used to reduce the size of an image without compromising the quality of an image and hence resolution increases, even in the presence of noise.

b. Texture Feature Extraction

Texture is one of the feature used to partition the images into regions of interest. Texture provides information in the spatial arrangement of intensities and to capture the similarities between images. For detecting texture, DWT is used. First step involves dividing the given image into 8×8 = 64 equal regions and then applying Wavelet transformation on each region. Next step is calculating mean and variance of four sub images corresponding to each region. Progressing all Mean and Variance feature vectors of every region. At the end, the obtained vectors will describe texture information of an image. These vectors are normalized and flattened to get the texture feature.

c. Color Feature Extraction

Color feature is designed to capture the most dominant color in an image. For detecting color, color histogram is used [4]. The image is divided into 5 parts, the first 4 parts belong to 4 corners and the fifth part belongs to the center region of the image. Later, the comparison is made between the same parts of the target and the query image for better results.

d. Chi-Square Distance Calculation

Chi-square distance means comparing two parameters the image stored in database and the query image whether they are related together or not. The search and retrieval is based on measuring the distance between query image and the target image [8]. The rank of each target image is calculated based on their distance from the query image. When there is no difference while comparing the parameters from database and query image, then it is known as null hypothesis. Suppose, there is a difference between the database and query image then that is known as alternate hypothesis.

e. Canny Edge Detection

Canny edge detection technique suppresses the noise while detecting the edges. The canny edge detection algorithm is composed of following steps:

- At first step, the given input RGB image is converted to Grayscale image.
- Next, Gaussian blur will be used to remove the noise [9] to get smooth and flawless image.
- Then the Sobel filtering will be used [4], to calculate the gradient of the image intensity for each pixel. It finds the direction increases from light to dark and the rate of change of direction. So, the approximated gradient will be obtained.
- Due to gradient magnitude process thick edges will be obtained. Hence the process of Non-Maximum Suppression (NMS) used to thin out the edges. It is a prediction of the movement of edges. For example, in a 3x3 image pixel there will be only 4 possible directions 0, 45, 90 and 130 degrees. Thus the edge has to be definitely oriented to one of these 4 directions. This is a kind of approximation, if the orientation angle is observed to be 5 degrees then it is taken as 0 degrees. Similarly, if it is 43 degrees, it taken as 45 degrees.
- After NMS there will be still some noise. So, to remove noise, thresholding will be used. In this process two values high and low thresholds will be set. Any pixel with a high threshold value will be consider as a stronger edge and low threshold value is not an edge. So, all low threshold values are set to 0. The values in between the low and high threshold levels are considered as weak edges.
- Then edge tracking approach is used to connect the weak edges to strong/actual edges. Weak edges which are not connected to the stronger ones are to be removed.

f. Convolutional Neural Network

CNN is most often applied in image processing to find object in an image. CNN is used for fast image retrieval composed of feature extraction and classification. CNN is made up of multiple layers such as convolution layer, Pooling layer, Rectified Linear unit layer and a fully connected layer. The convolutional layer works by placing a filter over an array of image pixels. This creates a convolved feature map. Pooling layer down samples or reduces the sample size of a particular feature map [10]. The output of this layer is pooled feature map. There are two methods of pooling layer they are Maximum Pooling and Average Pooling. Maximum Pooling takes the maximum input of the particular convolved feature and Average Pooling takes the average input of the convolved feature. Then the ReLu layer acts as an activation function ensuring the non-linearity. Without this layer if the data being fed into each layer it would lose its dimensionality. At last, the fully connected layer allows performing classification of images that are present in the dataset.

g. CNN with VGG16 Model

VGG16 is already trained over 138 million images and 1000 classes. It has 16 different layers followed by pooling layers [10]. The convolutional base layer will be at the beginning of the neural network and the fully connected layer will be at the end of the neural network. The features are extracted using VGG16 model. For feature extraction the convolutional base alone going to be used and a new fully connected layer will be introduces and train it on own model. Therefore the total number of features extracted is 512.

IMPLEMENTATION OF PROPOSED WORK

h. Implemention of DWT-Histogram Method

Here the input images are converted to gray scale image. The complexity of gray level images is lower than that of color images. In extracting the texture feature color is not needed. So it is converted into grayscale so that computing DWT will be easier. Then the image is divided into 8x8 regions. Now for each region, 1 level wavelet decomposition is performed [4]. Therefore 4 sub images will be obtained such as approximated, horizontal, vertical and diagonal details of an image. Then for each sub image, mean and variance values

will be obtained. That mean and variance values are the texture vectors of an image. The obtained feature vectors are stored in Excel files.

A	B	C	D	E	F	G	H	I
dataset\1.jpg	0.2680588	-3.85E-05	-9.60E-06	7.98E-07	0.2684088	-1.42E-05	4.31E-05	2.92E-06
dataset\10.jpg	0.0753016	-5.15E-05	-8.02E-05	2.14E-06	0.1046796	-3.34E-05	-0.000105	-1.86E-05
dataset\100.jpg	0.0393053	4.16E-05	-7.25E-05	-5.20E-06	0.0415163	-0.000297	-4.39E-05	2.60E-06
dataset\101.jpg	0.132046	-0.003425	-0.002446	-0.001568	0.0504542	-0.0016	0.0001703	0.0003731
dataset\102.jpg	0.0880499	6.52E-05	-0.000771	-0.00025	0.1293421	0.0003485	0.0001904	-0.000482
dataset\103.jpg	0.0296784	-1.23E-05	-0.000619	-0.000411	0.0472177	-0.000801	0.0003376	-0.000801
dataset\104.jpg	0.0280264	-0.000244	-4.75E-05	-3.76E-05	0.0436	7.57E-05	-0.000318	-0.000135

Fig.2 Texture feature vector

Fig. 2 shows that one image is divided into 64 parts and each part is decomposed into 4 sub parts. Totally there are 256 parts for each image. So that there are 256 mean values and 256 variance values totally 512 vectors are present in this csv file. Then by applying histogram the image is converted to HSV color space. Ellipse and rectangular function are used to separate the regions of an image. The combination of these two functions gives 5 regions of an image. Fig. 3 presents the image divided into 5 regions using color histogram.

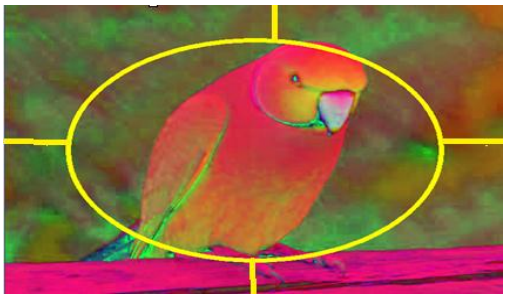

Fig.3 Dividing images into 5 regions

Chi-squared distance is used to measure the similarity between query image and database images. Two chi squared distances are calculated, one with respect to texture features and other with respect to color features. 70% of color and 30% of texture results are considered. Finally they are ranked and images with less distance value will be retrieved [6].

i. Implemention of Canny Edge Detection method

In canny edge detection Gaussian filter is used to remove the noise present in the image. Then the gradient calculation is applied to detect the edges along 4 directions. Due to this gradient calculation thick edges will be produced. So, the NMS is used to thin out the edges by finding the pixels with maximum value. After that, double thresholding will be used to obtain strong, weak and irrelevant pixels for edges and those irrelevant pixels will be removed. Finally, edge tracking by hysteresis is used to convert weak edged pixels to strong edges.

j. Implemention of CNN with VGG16 Model

In VGG16 model the convolution base portion will be considered for feature extraction after each convolutional block max pooling will be used to reduce the dimension of the image size by using 2x2 kernel and then the fully connected layer will be flattened and a new fully connected layer will be added.

4. RESULTS AND DISCUSSION

Canny edge detection method provides better results than sobel and laplacian filters. Canny completely removes the noise present in the image. Fig. 4 is the output of Canny Edge Detection.

Fig. 4 Canny edge detection output

First the output of DWT method used for finding similarities of given query and the database images, are shown as given below. Fig. 5 is the given query image which is a parrot image. Fig.6 is the top 44 retrieved images from the database using DWT method.

Fig. 5 Query image for DWT

Fig. 6 Retrieved Images of DWT method

Similarly, Fig. 7 is the given query image which is a tiger image. Fig. 8 is the top 31 retrieved images from the database using DWT method. It is clear from the output that the DWT method has given some of the images of other animals like Lion, Cheetah etc at the output. Hence the accuracy of the method is low.

Fig.7 Query Image of Tiger

Fig.8 Retrieved images of Tiger

Second the output of CNN method used for finding similarities of given query and the database images, are shown as given below. Fig. 9 is the given query image which is a rose flower image. Fig. 10 is the top 55 retrieved images from the database using CNN method. From the output figure it is clear that, rose flowers are retrieved with 100 % accuracy.

Fig.9 Query image for CNN

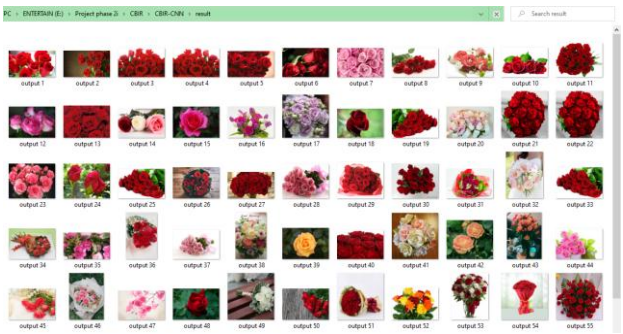

Fig. 10 Retrieved Images of CNN method

Similarly, Fig. 11 is the given query image which is a parrot image. Fig. 12 is the top 30 retrieved images from the database using CNN method. From the output figure it is clear that, parrot images are retrieved with 100 % accuracy.

Fig. 11 Query image for CNN

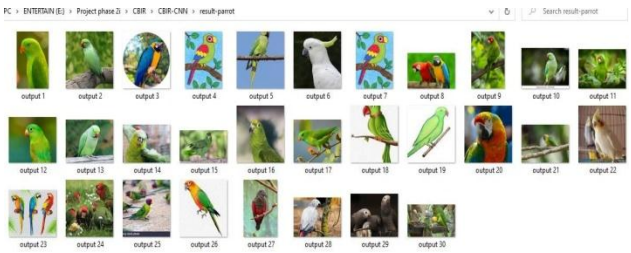

Fig. 12 Retrieved Images of CNN method

Table I and II show the accuracy for DWT and CNN methods respectively. From the table it is clear that CNN gives an accuracy of more than 98% for all the query images, whereas the accuracy of DWT method is from 45% to 83% based on the query image.

TABLE I. ACCURACY FOR DWT METHOD

Requested Images	**10**	**30**	**50**	**100**	**Accuracy**
Rose	100%	83.3%	86%	60%	83.32%
Tiger	90%	66.6%	62%	56%	68.65%
Parrot	100%	76.6%	60%	44%	70.15%
Ship	50%	50%	46%	34%	45%
Building	60%	80%	70%	40%	62.5%

TABLE II. ACCURACY FOR CNN METHOD

Requested Images	**10**	**30**	**50**	**100**	**Accuracy**
Rose	100%	100%	100%	100%	100%
Tiger	100%	100%	100%	97%	99.25%
Parrot	100%	100%	97%	100%	99.25%
Ship	100%	98.89%	100%	95%	98.47%
Building	100%	100%	97%	100%	99.25%

5. CONCLUSION

The overall accuracy for CBIR using DWT method is 65.72% and for CNN method 99.24%. From this, it is observed that CBIR using CNN with VGG16 model attains high accuracy compared to DWT method. DWT method extracts only the handcrafted features like color and texture. Since VGG16 model is a deep learning method with 16 convolutional layers it is more accurate than other methods. Thus each of the images in the dataset is well

trained through VGG16 model and precise feature vectors are obtained. From the results it is inferred that more similar images are retrieved using CNN with VGG16 model than DWT method.

6. REFERENCES

[1] Toshikazu Kato, "Database architecture for content-based image retrieval", Image Storage and Retrieval Systems. International Society for Optics and Photonics, pp. 112–123, April 1992.

[2] Dileshwar Patel, Amit Yerpude, "Content based Image Retrieval: A Review", International Journal of Engineering Research & Technology, vol. 3, no. 20, June 2015.

[3] Domonkos Varga, Tamas Szir, "Fast content-based image retrieval using Convolutional Neural Network and hash function", International Conference on Systems, Man, and Cybernetics, vol. 9, no. 12, Oct. 2016.

[4] Vijayakumar Bhandi, K. A. Sumithra Devi, "Image Retrieval Using Features From Pre-Trained Deep CNN", International Journal of Scientific & Technology Research, vol. 9, no. 6, Jun. 2020.

[5] Akshara Preethy Byju, Begüm Demir, Lorenzo Bruzzone, "A Progressive Content-Based Image Retrieval in JPEG 2000 Compressed Remote Sensing Archives", IEEE Journal on Geoscience and Remote Sensing, vol. 58, no. 8, Aug. 2020.

[6] Rehan Ashraf, Mudassar Ahmed, Sohail Jabbar, Shehzad Khalid, Awais Ahmad, Din Gwangil Jeon, "Content Based Image Retrieval by Using Color Descriptor and Discrete Wavelet Transform", Springer Journal, vol. 7, no. 44, Jan. 2018.

[7] Jigisha M. Patel, Nikunj C. Gamit, "A Review on Feature Extraction Techniques in Content Based Image Retrieval", IEEE Journal, vol.5, no. 7, Aug. 2016.

[8] Amjad Shah, Rashid Naseem, Sadia, Shahid Iqbal, and Muhammad Arif Shah, "Improving CBIR Accuracy using Convolutional Neural Network for Feature Extraction", IEEE Journal, vol. 5, no.8, July 2020.

[9] Guoyong Duana , Jing Yanga , Yilong Yanga, "Content-Based Image Retrieval Research", International Conference on Physics Science and Technology, vol.4 no.3, Nov.2011.

[10] M.Thilagam, K.Arunish, "Content-Based Image Retrieval Techniques: A Review", IEEE Journal, vol. 9, Dec. 2018.

[11] Ahmadi, A.; Sajadian, N.; Jalaliyan, H.; Naghibirokni, N. Study And Analysis of Optimized Site-selection for Urban Green Space by Using Fuzzy logic: Case Study: Seventh Region of Ahvaz Municipality. IARS' International Research Journal, Vic. Australia, v. 2, n. 2, 2012. DOI: 10.51611/iars.irj.v2i2.2012.23.