Performance Evaluation of Content-Based Image Retrieval Using Block Truncation Coding and CNN

A. P. Nilawar¹ and C. G. Deth²

¹Assistant Professor, Department of Electronics and Communication Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India

²Director, UGC HRDC, R.T.M.N.U., Nagpur, Maharashtra, India

ABSTRACT

Retrieving similar images from a large image database is a critical task. The solution to this problem is the use of a Content-Based Image Retrieval System. The images are described through their content, there is three predominant content existing in an image like color, shape, and texture. In this paper, we are evaluating the performance of the CBIR system using two methods. The first method consists of Block Truncation coding (BTC) with Grey level co-occurrence matrix (GLCM). The second method consists of Use of CNN. The feature extraction technique is achieved based on an input query image from the database and features are saved in a feature dataset. A proposed strategy retrieves similar images from a database that fulfills the user's desire. The similarity measurement can be done using the Euclidean distance and hashing technique. The overall performance of the retrieval system has been analyzed through the parameters Precision and Mean Average Precision. The experimental result shows encouraging results using CNN which leads to improving accuracy.

KEY WORDS: CBIR, BTC, GLCM, FOM, CNN.

INTRODUCTION

CBIR incorporates the extraction of features from an image essentially based on its content, estimating likeness between the query image and image in the database, and positioning from generally tantamount to the assorted image. CBIR system has two essential stages in the main stage the pre-handling for image database alluded to as image highlight extraction based on the sort of features like color, shape, and surface. What's more, in the second stage estimation of likeness from database image alluding query image gave using the user.

The above figure explains the basic operation of the CBIR System. In the retrieval system, there are essentially three most important fields text-based image retrieval, retrieval based on visual feature, and semantic features. In a text-based image retrieval system images are provided through labels and for retrieval cause similarity of these labels are used. A predominant disadvantage of these types of structures is that a lot of human work evolved for image labeling and if the database of images in a massive quantity then it is pretty difficult. And labeling of images is no longer user friendly it relies upon database creators labeling strategies (Song, H., Li, X. and Wang, P). Then the visual-based photo retrieval systems used for image retrieval purposes and low-level features are considered for feature extraction (Ganar, A.N et.al.). Presently CBIR is based on visual features along with semantic features in use. There are three most important features are existing in an image specifically color, texture, and shape.

ARTICLE INFORMATION

*Corresponding Author: aa.nilawar@gmail.com

Received 14th Oct 2020 Accepted after revision 28th Dec 2020

Print ISSN: 0974-6455 Online ISSN: 2321-4007 CODEN: BBRCBA

Thomson Reuters ISI Web of Science Clarivate Analytics USA and Crossref Indexed Journal

NAAS Journal Score 2020 (4.31)

A Society of Science and Nature Publication,
Bhopal India 2020. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

Doi: <http://dx.doi.org/10.21786/bbrc/13.14/27>

Figure 1: Basic CBIR System

Color Features: Images are largely categorized into grayscale images and color images. In a grayscale image color pixel having a solely grayscale area while in a color image three color intensity ranges are used. In the color image red, green and blue intensities are used. Color histogram, color coherence, and color moments are important methods used for image retrieval

Texture Features: It measures the homogeneity of a pixel over repeated patterns in the image. We can format a retrieval system the use of two tactics particularly structural and frequency-based approaches

Shape Features: It gives edges or outlines of an object existing in an image. Region and boundary-based techniques are used in the retrieval systems based totally on shape features

Neural Network: A neural network consists of the input layer, hidden layer, and output layer. Convolution Neural Network is used for feature extraction from images.

Liturrature Survey

Block Truncation Coding: As explained in (Gahroudi, M.R. and Sarshar, M.R.), BTC has proven its use in the compression area and additionally, we can efficaciously use it in the CBIR application. BTC is an image ciphering technique, various advanced ciphering methods developed the usage of BTC (Qiu, G.).BTC can be used in image compression, ciphering, and retrieval applications. It consists of two tactics encoding and decoding. In the encoding process, it will divide the entire image into several image blocks having two quantizers particularly high and low. In this technique mean value and standard deviation are calculated for every block. Then the mean value is compared with every pixel in the image if the mean value is greater than the pixel value changed by using 1 in any other case pixel value replaced via 0. First, we will convert the image into a grayscale image, then the grayscale image is divided into several blocks of different sizes like 2 by 2, 4 by 4, 16 by 16, and so on. In the next stage, we will compute the mean value and standard deviation. After computing, the mean value bit map pattern of an image is generated.

Table 1. Gray Scale Image

140	142	144	145
146	141	146	142
145	141	144	142
142	138	141	144

Table 2. Binary bitmap pattern

0	0	1	1
1	0	1	0
1	0	1	0
0	0	0	1

In Table, I grayscale image of block size 4 by 4 considered. The mean value for this block is 142.5 and the standard deviation is 2.199. If the mean value greater than the pixel value, then the pixel value is replaced by 1 otherwise pixel value is replaced by 0. By using this method table II was generated. In decoding two quantization levels as high and low are used to reconstruct value from the encoded block. Figure II shows the encoding and decoding process using BTC.

Figure 2: Illustration of BTC

2.1.1 Similarity Measurement: Euclidean distance is the most frequently used technique to discover the similarity between images (Wang, L., Zhang, Y. and Feng, J.). Euclidean distance measures the similarity between the query image and images existing in the database. Assume X is a set of query images and Y is a set of images in the database. The measuring technique can be carried out on $X=[X_1, X_2, \dots, X_m]$ to every component in $Y=[Y_1, Y_2, \dots, Y_n]$ the usage of the following equation

$$d = \sqrt{\sum_{i=1}^K (x_i - y_i)^2} \quad (1)$$

Where,

X_i and Y_i represents the feature vector of the query image and image in the database at component i respectively K represents the complete number of the component in individual feature vectors

d represents the similarity rating between query and image database

Convolutional Neural Network: The Convolutional Neural Network is a neural network having four steps. The first step is applying convolution to input. Convolution is a mathematical operation in which two functions are used. The values of the functions are changed due to convolution. For the application of CNN in image processing, the feature detector is needed. Generally, the Feature detector is a matrix of size 3×3 . In convolution operation, this feature detector is applied to an image, and after convolution operation, we get a matrix with some values, which are different from the original (Wang, L., Zhang, Y. and Feng, J.). In this step, a feature detector is applied to the input image which results in the generation of the feature map. After feature map generation for removing unwanted information, the ReLU rectifier function is used. In the second stage, the pooling operation is performed. There are three types of pooling are used namely max pooling, mean pooling, and sum pooling (Li, J. and Wang, J.Z.). Stage three consists of flattening which generates a feature map. And in the last stage full connection is done. The softmax and cross-entropy are two functions used for improving CNN (Kuo, C.).

Figure 3: CNN Used for Feature Extraction

The above figure III shows Convolutional Neural Network which is used for feature extraction. In this network, CNN along with ReLU is used along with max pooling. The input for this network is an image. The output provides a feature vector of 4096 elements. This network is also known as the VGG16 network (Simonyan, K. and Zisserman, A.).

Similarity Measurement: The similarity between the query image and images in the database is measured by various methods. Euclidean distance, cosine distance, and Hamming distance is a popular algorithm for measuring similarity. In this paper, we will use the Approximate Nearest Neighbor method (ANN). ANN also is known as Locality Sensitive Hashing (LSH). Let us consider the database of images having various categories like bus, bird, planes, etc. After applying CNN feature vectors are generated for each image. Each vector represents the feature of an image. In ANN each Feature vector is provided to the hashing function. The hashing function is nothing but thresholding. It will convert feature vector value in either 0 or 1. After applying the hashing function the vectors having closer values are put into one common

basis vector. In this way, various basis vectors were created. During the retrieval basis vector of the query, the image is compared with the basis vector of the image database and provides results. In this process since the basis vector consist of various vectors hence retrieval takes less time as compared to other functions.

Proposed Methods

Method I: Block Truncation coding with GLCM: In this method first, we apply BTC to all images in the database. For each image, GLM is calculated. And these feature vectors are indexed in the database. This system is implemented on the Wang database which consists of 1000 images spread across 10 categories (Wang Image Database). The algorithm for the proposed system is given by,

Step 1: User has to select rectangular block of size 2×2 , 4×4 or 16×16 .

Step 2: Then depending on the selection of block size image divided into non overlapped blocks.

Step 3: the Preprocessing process starts where each image in the database applied by BTC.

Step 4: After application of BTC, GLCM applied to an image.

Step 5: For every image, feature extracted and indexed in the database with the feature vector.

Step 6: After preprocessing done user has to provide a query image.

Step 7: Using Euclidean distance similar images are displayed

Figure 4: Sample Results for the method I

Method II: Convolutional Neural Network: In this method pre-trained CNN network is used. VGG16 is a pre-trained CNN network. The last layer of the VGG16 network was removed. Each image in the database generates 4096 elements as features. This proposed system is applied to the image database which consists of 1000 images spread across 20 categories (Rashtchian, C et.al.). The algorithm for the proposed system is given by,

Step 1: In an image database several images with different categories are present.

Step 2: These images are fed to CNN for feature extraction.

Step 3: For each image feature vector generated. Each vector consists of 4096 elements.

Step 4: Features are indexed in the database.

Step 6: The user has to provide an input image considered as a query image.

Step 7: Features of a query image extracted using CNN.

Step 8: Using ANN similarity between features of query image matched with index.

Step 9: Finally top 10 similar images are displayed.

Step 10: Repeat from step 6 for another input.

Figure 5: Sample Results for method II

RESULTS AND DISCUSSION

The performance of the system was evaluated using two parameters Precision and Mean Average Precision (MAP). Precision is the ratio of the number of relevant images retrieved to the total number of images retrieved.

Table 3. Results for Precision

Category	Method I
Airplane	87.75
Beach	83.38
Building	86.75
Cat	84.25
Dinosaurs	90.75
Elephants	82.75
Flowers	86.25
Horses	82.25
Mountains	80.25
Food	85.75
MAP	85.01

$$MAP = \sum_k^K \frac{Avg P(k)}{K} \quad (2)$$

Where k = value of precision for query and K is the number of query images.

Table 4. Results for Precision

Category	Method II	Category	Method II
Airplane	96.18	Dining Table	96.28
Bicycle	96.28	Dog	98.40
Bird	96.28	Horse	97.40
Boat	96.70	Motor Bike	96.18
Bottle	95.70	Person	96.50
Bus	97.24	Plant	95.40
Car	98.20	Sheep	96.28
Cat	96.25	Sofa	95.20
Chair	96.90	Train	96.24
Cow	96.85	TV	98.36
MAP		96.64	

Table 5. Comparison with existing methods

Methods	Images in	MAP
Color in YCBCR, Color Moment (Dandotiya, Y. and Atre, A)	Database 1000 Images	82.70
HSV Color Histogram and Hu moment invariant (Rajkumar, R. and Sudhamani, D.M)	1000 Images	84.20
BTC with GLCM (Proposed)	1000 Images	85.01
CBIR Using CNN (Proposed)	1000 Images	96.65

The above tabular data is conclusive for providing a comparison between two alternating methods BTC and CNN which provides better results on comparison with previously reported methods. For the requirements of providing better results of color features one needs to look up to BTC with GLCM providing 85% MAP. Whereas when one requires greater precision one should utilize CNN where one gets precision above 95% MAP as highlighted in table V. Also as categories increase one should opt for CNN for evolving precision as shown in table III and IV.

CONCLUSION

Analysis of color features is done using BTC and analysis of texture utilizes GLCM. The main contribution

of merging the two methods provides us handcraft features that help in increasing the precision for image retrieval. To provide a better solution towards feature extraction as well as the training time of the database we personalized CNN with the view of the generic image database. This helped in our research findings with an enormous improvement in the precision of image retrieval and classification. Research of this decade shows the advantages of CNN in image classification but our current contribution provides us a breakthrough in utilizing CNN for image retrieval as well.

REFERENCES

- Dandotiya, Y. and Atre, A., 2017, April. Image retrieval using edge detection, RLBP, color moment method for YCbCr and HSV color space. In 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA) (Vol. 2, pp. 662-668). IEEE.
- Gahroudi, M.R. and Sarshar, M.R., 2007, February. Image retrieval based on texture and color method in BTC-VQ compressed domain. In 2007 9th International Symposium on Signal Processing and Its Applications (pp. 1-4). IEEE.
- Ganar, A.N., Gode, C.S. and Jambhulkar, S.M., 2014, January. Enhancement of image retrieval by using colour, texture and shape features. In 2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies (pp. 251-255). IEEE.
- Kuo, C., 2016. Understanding CNN with Mathematical Model.
- Li, J. and Wang, J.Z., 2003. Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Transactions on pattern analysis and machine intelligence, 25(9), pp.1075-1088
- Qiu, G., 2003. Color image indexing using BTC. IEEE transactions on image processing, 12(1), pp.93-101.
- Rai, H.G., Shen, X., Deepak, K.S. and Krishna, P.R., 2011, November. Hybrid feature to encode shape and texture for content based image retrieval. In 2011 International Conference on Image Information Processing (pp. 1-6). IEEE.
- Rajkumar, R. and Sudhamani, D.M., 2019. Development of retrieval system using color from multi-segment and shape features from an image as a combination. International Journal of Engineering and Technology, 9(1).
- Rashtchian, C., Young, P., Hodosh, M. and Hockenmaier, J., 2010, June. Collecting image annotations using amazon's mechanical turk. In Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (pp. 139-147).
- Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- Somnugpong, S. and Khiewwan, K., 2016, July. Content-based image retrieval using a combination of color correlograms and edge direction histogram. In 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE) (pp. 1-5). IEEE.
- Song, H., Li, X. and Wang, P., 2009, July. Multimodal image retrieval based on annotation keywords and visual content. In 2009 IITA International Conference on Control, Automation and Systems Engineering (case 2009) (pp. 295-298). IEEE.
- Wang, L., Zhang, Y. and Feng, J., 2005. On the Euclidean distance of images. IEEE transactions on pattern analysis and machine intelligence, 27(8), pp.1334-1339.