

Image Retrieval based Convolutional Neural Network

Nuha Mohammed Khassaf^{1*}, Shaimaa Hameed Shaker²

¹ Iraqi Commission for Computers and Informatics, Informatics Institute for Postgraduate Studies, Baghdad, IRAQ

² Department of Computer Science, University of Technology, Baghdad, IRAQ

*Contact email: nuha313moh@gmail.com

Article Info

Received
28/08/2020

Accepted
26/10/2020

Published
20/12/2020

ABSTRACT

At the present time, everyone is interested in dealing with images in different fields such as geographic maps, medical images, images obtaining by Camera, microscope, telescope, agricultural field photos, paintings, industrial parts drawings, space photos, etc. Content Based Image Retrieval (CBIR) is an efficient retrieval of relevant images from databases based on features extracted from the image. Follow the proposed system for retrieving images related to a query image from a large set of images, based approach to extract the texture features present in the image using statistical methods (PCA, MAD, GLCM, and Fusion) after pre-processing of images. The proposed system was trained using 1D CNN using a dataset Corel10k which widely used for experimental evaluation of CBIR performance the results of proposed system shows that the highest accuracy is 97.5% using Fusion (PCA, MAD), where the accuracy is 95% using MAD, 90% using PCA. The performance result is acceptable compared to previous work.

KEYWORDS: CBIR; preprocessing; feature extraction; PCA; MAD; GLCM; Fusion.

الخلاصة

في الوقت الحاضر، يهتم الجميع بالتعامل مع الصور في مجالات مختلفة مثلًا لخرائط الجغرافية، والصور الطبية، والصور التي يتم الحصول عليها بواسطة الكاميرا، المجهر، التلسكوب، صور المجال الزراعي، اللوحات، رسومات الأجزاء الصناعية، صور الفضاء، إلخ. أسترجاع الصورة القائمة على المحتوى (CBIR) هو أسترجاع فعال للصور ذات الصلة من قواعد البيانات بناء على الميزات المستخرجة من الصورة. يتبع النظام المقترح لأسترجاع الصور المتعلقة بصورة الاستعلام من مجموعة كبيرة من الصور النهج القائم على استخراج ميزات النسيج الموجودة في الصورة باستخدام الأساليب الإحصائية (PCA، MAD، Fusion+GLCM) بعد المعالجة المسبقة للصور. تم تدريب النظام المقترح باستخدام 1D CNN لمجموعة بيانات Corel10k التي تستخدم على نطاق واسع للتقييم التجريبي لأداء CBIR، وتظهر نتائج النظام المقترح أنا على دقة هي 97.5% باستخدام Fusion(PCA,MAD) حيث بلغت الدقة 95% باستخدام MAD و 90% باستخدام PCA. تم مقارنة الأداء بالدراسات السابقة، وتبين ان نتائج النظام المقترح أعلى من غيرها.

INTRODUCTION

Content Based Image Retrieval (CBIR) is the technique that use retrieve images based on their content. Visual information retrieval is the use of computer vision to the image retrieval problem of searching for digital images in large databases. The basic thought of CBIR in analysis image information with low-level feature of an image, including color, texture, shape, distance relationship between images, etc., and set feature vector for the image as an indicator it [1]. CBIR system includes preprocessing, extracting features, dataset training, classification of test images and evaluating the system according to the techniques that be used in the proposed system.

METHODOLOGY

The methodology of this paper consists four phases, the first phase is preprocessing includes five steps; image resize, convert image to double Precision, convert image to HSV (V), remove noise from image, and compression image, the second phase is extract the feature of the image using three methods (PCA, MAD, GLCM) and fusion these features, the third phase, dataset training using deep learning algorithm Convolutional Neural Network (CNN) and using the model as classifier to testing images test. The last phase is evaluation the proposed system using evaluations criteria. **Error! Reference source not**

found.illustrates content image retrieval for proposed system.

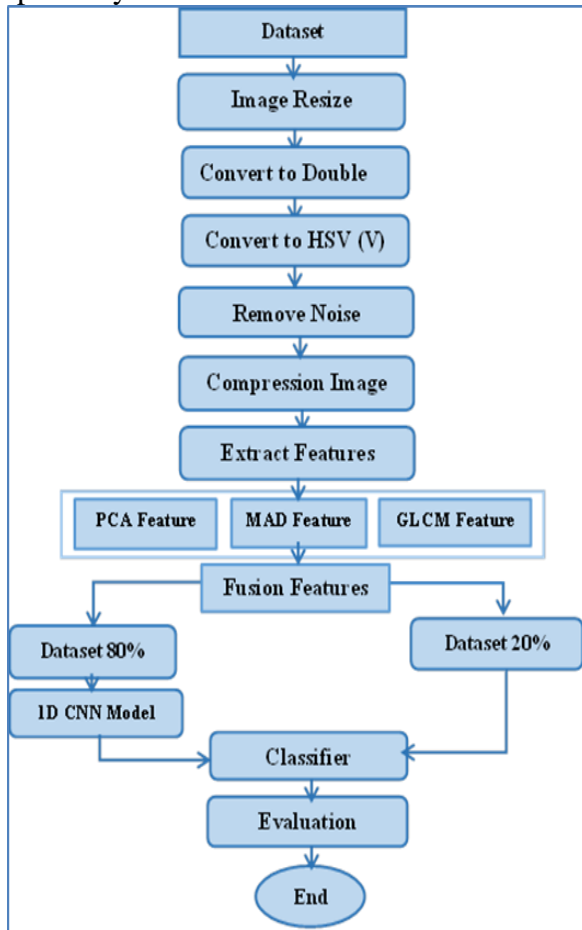


Figure 1. diagram of the CBIR system.

LITERATURE SURVEY

There are several techniques have reported in CBIR system which have used for retrieve image depended on features color, texture, and shape in an image, some of these techniques have been done on cloud computing:

R. Madana M., A. Rama M. R. [2] in 2015, suggested a technique that could be used to retrieve images relevant from a wide range of discriminative images by approach segmentation for extraction the various features by used Discrete Wavelet Transform technique (DWT) that as storage in vectors and then are compared with feature vectors in query image for sort there in descending order, by used K-Means classification and Euclidean Distance. CBIR is an application built on the cloud by using Windows Azure was a parallel process where big databases of images must be run to arrange them depended on similarity to the query image. The performance of system, average percentage of accuracy of CBIR: 82.5%.

Anas F. Ahmed [3] in 2016, the proposal presented a comparative study of human face recognition using two feature extraction techniques: Principle Component Analysis (PCA) and Linear discriminant analysis (LDA). Describe of the participating of each eigenfaces in representing the entry face image to deal with eigenfaces as the basis for mapping the facial image. The drop vector is used in the typical pattern recognition method to determine which pre-defined facial categories, if any. The class can be determined by averaging the results of the eigenface representation on a tiny number of facial images one people. Rating was done by comparing the projection vectors of the training face images with the vectors of the put face image projection. The comparing was depended on the Euclidean distance both the facial categories and the input face image, the highest recognition rate is recorded for the LDA technique was 95.981%. On the other side, the highest recognition rate that recorded for PCA technique was 94.027% where the train images/test images ratio is (80/20).

Manisha V., Balasubramanian R. [4] in 2017, suggested a technique using local neighborhood difference pattern (LNDP) for local features. The traditional Local Binary Pattern (LBP) converts each image pixels into a binary pattern depended on its relation to neighboring pixels. The suggested feature descriptor differs from the LBP because it converts the reciprocal relationship of all adjacent pixels into a binary pattern. Both LBP and LNDP are complementary to each other because they extract various information used local pixels intensity. The performance of system for Corel10k database, precision: 42.81, recall: 17.01, MIT database, precision: 70.24 recall: 39.

S. Selvam, S. Thabasu Kannan [5] in 2017, proposed a technique to combine the genetic algorithm (GA) with the HARP aggregation algorithm to improve system retrieval accuracy for less computational time and to restore the relevant image and possible resolution using CBIR. The efficiency of the system is improved by looking at the image filter for the purpose of calculating the similarity i.e. not taking into account the entire database images. The filter image resides in the same group as the query image, as the usefulness of the assembly process clearly demonstrated retrieval accuracy, that performance of model is good in speed and decreasing retrieval time. The performance of

system, average precision: 89.1% and average recall: 69.8%.

K. Meenakshisundaram, G. Vijai Prabhu [6] in 2017, proposed technique Center Symmetric Local Binary Pattern (CSLBP) was depended on retrieval of images using the feature; the extraction was depended on features such as power, contrast, and homogeneity. Separately calculated threshold value saved in the features in databases where the feature matching was done via the Euclidean Distance used to calculate the distance for two images, the CSLBP produced retrieval accuracy: 77.87%.

S. Govindaraju, G. P. Ramesh K. [7] in 2017, proposed a method Hierarchy and Fuzzy C-Means (HFCM) to collect unsupervised images using continuous probabilistic models and principles of information theory. Image compilation relates to content based image recovery systems. It enables the implementation of efficient recall algorithms and the creation of an easy user interface for the database. The performance of system, precision: 0.81, min precision: 0.69 and recall: 0.71, min recall: 0.31.

Umesh D. Dixit, M. S. Shirdhonkar [8] in 2018, presented a technique to retrieve documents from a person's database based on a face photo. The technique included three stages: (1) Face image detection from the test document. (2) Extracting facial image features. (3) Retrieving all documents containing the face image like to the test document. The proposed technology extends the concept of the GLCM array to include color images for feature extraction and Mahalanobis distance to find a similarity scale. The performance of system for average mean precision (MAP): 82.66%.

Mamta M, et al. [9] in 2018, suggested technique to solve the problem of efficient image retrieval through a huge storage space on Cloud Computing. In Several guides from the images, finding the required photo may cost a lot of time and effort. It is easy to search for an image through its feature (Kerker conversion for global features and wavelet conversion for local features) and thus list short pictures, among many in the database, thus reducing the cost of calculation. An image can be better remembered by its contents such as its color, shape, texture. That

making is a key to a list of images possessing similar features out of a group of images in the database, efficient performance, scalability and min cost, flexible system that operate huge amount of data.

Sidi A. Mahmoudi, et al. [10] in 2019, proposed efficiently combining Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Feature (SURF) technologies that allowed the extraction and matching of image features that improved the image recovery process. The proposed technologies were applied to the Central Processing Unit (CPU) and also adapted to take full advantage of the force of Graphics Processing Unit (GPUs). Their platform was provided with reacting web solutions that provide users with access to image recovery, testing and evaluation technologies, demonstrated the efficiency of system in terms of recall/precision, computation time and multi-user exploitation.

Vo Thi H. Tuyet, et al. [11] in 2019, suggested technology Shearlet to retrieve content-based medical images. It consists two steps: the offline job and the online job in the medicinal image databases: (1) were extracted the features of local objects from the medicinal images in the shearlet field. Were discovered the circumference of the object in the photos with an active contour model. (2) Were undertaken an online mission to retrieve content-based images in the databases. Their system received the test image and showed what looked like images compared to the similarity with the information gathered from the first step. The performance of system, precision: 69.326% and Recall: 81.233%.

Image Retrieval Ways

An image retrieval system is a computer system for browsing, searching and retrieving images in an image database as follows:

1-Text-based

In text retrieval, images are indexed using keywords, subject headings or classification codes, which in turn are used as retrieval keys during search and retrieval. Text retrieval is nonstandard because different users use different keywords for annotation. Text descriptions are sometimes subjective and incomplete because they cannot portray the intricate picture features

very well. Examples are texture images that cannot be described with text. In text retrieval, humans are required to describe each image in the database personally, so for a large image database the technique is complex and expensive. [12].

2- Content Based Image Retrieval (CBIR)

CBIR is a system for retrieving images by automatically- extracted features such as color, texture and shape. In typical content-based image retrieval system the contents of the images are extracted by multi-dimensional feature vectors. These feature vectors form a feature database are used for retrieval system. The system then transforms the query image into its internal representation of the feature vectors. The similarities between these feature vectors and the query of the images is then calculated and performs the retrieval with the help of an indexing scheme [13].

The advantages of CBIR:

A. The features such as color, texture, shape and spatial can be retrieved automatically.

B. Similarities of images are based on the distances between features different query techniques and implementations of CBIR make use of different types of user queries.

In the following, types of low-level descriptors which use to the images retrieval [7]:

A. Retrieval based on Color Feature: The color is more sensitively and obviously features in the image, and is usually described by histograms graphics. The color graph method has the advantages of speed, less space for memory and without sensitivity to images.

B. Retrieval Based on Texture Feature: Generally, it depend the statistical feature of the material and the structure feature in addition to the features based on the spatial field being changed to a frequency band. Homogeneous texture descriptor describes an accurate statistic distribute texture of image. It enables classification of images in high resolution and is used for similarity recovery applications.

C. Retrieval Based on Shape Feature: Object shape can consist of one or more areas, and it may also have perforations. Extraction feature of the shape using the Edge graph descriptor. An edge can be described as image positions where the local density changes in a specific direction. The higher the intensity, the more evidence there is of an advantage in this position. Edges are classified

into five kinds: vertically, horizontally, 45 diagonally, 135 diagonally and no directionally [14].

In general, the CBIR system has following stages:

A. Pre-Processing

Filters and Wavelet transforms play an important role in image processing, analysis and compression part. Nowadays the filters and the wavelet algorithms are very popular and active methods in image processing, de-noising and compression [15].

B. Extracting Features: Features of dataset images and feature of query image is extracted depended on different features of image as color, texture, and shape features [16].

C. Images Classification: Classification is the supervision and un-supervision for categorize images to groups. In supervision classifier is to give a group of labeled images and the problem is the labeling of newly encountered images but without labeling. In a non-supervised classification is to group a certain group of unlabeled images into meaningful groups according to the image content without prior knowledge [16].

D. Image Matching: Measuring the similarity between the test image and the images stored in databases depended on distances like Manhattan, Mahalanobis, chi square, Euclidian and other distances. The features closest to the query image are checked, and the image retrieval according to those features [6].

General Preprocessing ways of Image Retrieval

Image pre-processing is an algorithm challenge that uses image processing algorithms to enhance the representation quality of images in order to improve the quality of image retrieval process. The goal of CBIR image preprocessing is [17]:

- 1- Determine the method to store the images in databases.
- 2- Determine the filtering method to improve the quality of the database images.
- 3- Avoid problems created by different sized images.
- 4- Normalize the database images to remove the effect of color variations.
- 5- Perform color transformation or color space conversion.
- 6- Segmentation the image into meaningful regions.

Image Preprocessing

In order to make the feature extraction step easier, a set of functions to pre-process the input image has been developed.

Resize Image

The amount of features that returns some feature extraction functions depend on the number of pixels in the image. Like in cropping, the image size is concern point in the project develop by values are set up. The values of output image size are 128 pixels in height and 128 pixels in width where the resize is done with the matlab function [18].

```
image_r=imresize(image_original,[128,128])
```

Convert to Double Precision

Double is used to converting uint8 and uint16 data to double precision and also rescales the values in the range 0-1 by dividing all the entries of the input by the maximum possible value of the input data type. So if the input data type is uint8 whose maximum possible value is 255, therefore all the values will be divided by 255. It is possible that the maximum value in image data may not be 255 but since the maximum possible value of uint8 is 255, so all the values will be divided by 255. It is better to work with double when working with image processing and computer vision. In matlab, the covert to double is done with the matlab function [19].

```
image_d=im2double(image_r)
```

Hue Saturation Value model (HSV) Image

The RGB color is the default color space used by computers, representation of color images is more complex and varied. The common way of storing color image contents, each pixel is usually represented by a 24-bit number containing the amount of its red (R), green (G), and blue (B) components [20].

Perceptual color spaces are created from those models that treat the color in a more human intuitive form. In order to achieve this objective, any color is represented by a specific tone or hue, a level of saturation for the hue, and the amount of light available or illumination.

Inside this category, it can be found Hue Saturation Value model (HSV), Hue Saturation

Lightness (HSL) and Hue Saturation Intensity (HSI). All of these color models have a similar description about the color hue but differ in the saturation and illumination definition.

HSV is a cylindrical coordinate system predicated upon three variables and HSV can be computed from the RGB standard model:

Hue (H) is ranging from 0 degrees to 360 (or zero to 2π radians).

Saturation (S) is amount of hue or departure from neutral. 0 indicates a neutral shade, whereas 1 indicates maximum saturation.

Value (V) is Maximum value among the red, green, and blue components of a specific color.

The major advantage of using this model is the fact that it presents immunity to illumination changes. The other feature of this model can be found in the color hue changes; they are continuous and linear. HSV is preferred due to the geometric representation, which is usually more natural than HSL, which allows a better color hue manipulation [21].

Noise Removal

Salt and pepper noise randomly changes a certain amount of pixels to two extremes, either 0 or 255, for an 8-bit image. Noise greatly damages the image information which leads to difficulties in success. Median filter is used to remove noise while preserving edges. To apply a median filter, 3 x 3 size window have been selected. The median filter works by moving the window through the entire image pixel by pixel.

The output is determined by arranging all the pixel value from the window into numerical order and after that substituting the pixel being considered with the center (middle) pixel value [22].

Image Compression

Discrete Haar Wavelet Transform is used to reconstruct the image efficiently by transforming pixel value and get back the image from corresponding wavelet coefficients.

Decomposition the image to 4 sub bands indicated by (LL, HL, LH, HH). Each of those sub bands could be thought of as a smaller version of the image which represents various image properties.

The Low-Low is approximation to the original image, and offers robust signal analysis tools, which is used to compression images [23].

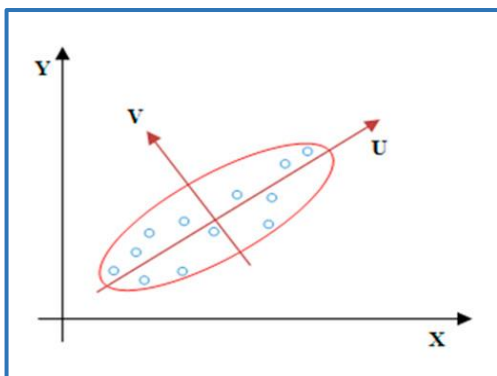
This step of preprocessing is used with PCA method only where the size of image 128X128 is compressed to 64X64.

Image Feature Extraction

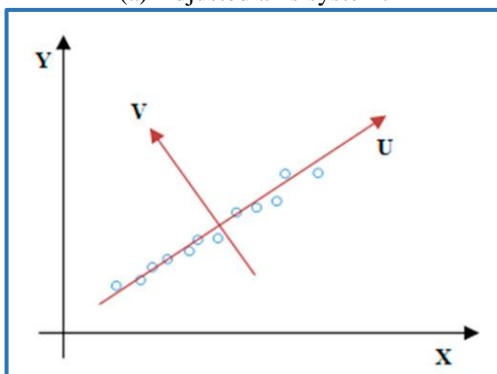
The low level texture feature is extracted which used for comparison to retrieval images.

Principal Component Analysis (PCA)

PCA is unsupervised learning technique and a statistical approach to find the principal features of a distributed dataset based on the total variance. Given a set of multivariate distributed data in X-Y coordinate system, PCA first finds the maximum variations of the original datasets. These data points are then projected onto a new axis called U-V coordinate system. The direction of U and V-axis is known as principal components. The principal direction in which the data varies is shown by U-axis following by its orthogonal direction, V-axis. In the case where all the data points on V-axis are very close to zero as shown in Figure 2, the data set can be represented by only one variable U and the variable V can be discarded [24]. The algorithm 1 illustrates the Principal Component Analysis algorithm steps [25].



(a) Adjusted axis system.



(b) Variable v is discarded.

Figure 2. PCA for dimension reduction.

The advantages and limitations of PCA as following [26]:

Advantages of PCA:

- 1- Reducing complexity.
- 2- Smaller database representation, only images are stored in the form of their projections on a reduced base.
- 3- Reducing noise because the basis for maximum difference so the small differences in the background are automatically ignored.
- 4- 2DPCA over 1DPCA is that the feature carrier is now, so the dimensional problem is greatly reduced.

Limitations of PCA:

- 1-The covariance matrix is difficult to be evaluated in an accurate manner.
- 2- PCA is a less sensitive to different training data set.
- 3- Computationally expensive with the increase in data size.

Algorithm 1: Principal Component Analysis

Inputs: a collection of images as a set of coordinates in a high-dimensional (n) dataset.

Outputs: results a lower dimensional image (eigen images) of given m images.

Begin

1. Conversion each m input images (I_1, I_2, \dots, I_m) of size $N \times N$ into $N \times 1$ one dimensional column vectors ($\Gamma_1, \Gamma_2, \dots, \Gamma_m$).
2. Normalization $N \times 1$ column vector. That means all the common features from every M images are removed so that each image has only unique features.
 - A. Calculation an average/mean image vector (Ψ).
 - B. Subtraction the average/mean image vector (Ψ) from each M one dimensional column vectors Γ .

Normalized Image Vectors $\Theta_i = \Gamma_i - \Psi$.
3. Calculation K significant Eigen Vectors (principal components / axes) and eigen values (variance) from a covariance with reduced dimensionality, Eigen vectors \Rightarrow Determines the direction of the new feature space. Eigen Values \Rightarrow Determines the magnitude/variance of the data along the novel feature space. Calculation the Covariance matrix $C = A^T A$. Where $A = [\Theta_1, \Theta_2, \Theta_3, \dots, \Theta_M]$. Dimension of $A = N^2 \times M$ & $A^T = M \times N^2$ Hence $C = (M \times N^2) (N^2 \times M) \Rightarrow (M \times M)$.

4. Generation M eigen vectors (Principle component).
5. Selection K best Eigen Vectors $K < M$.
6. onversion lower dimensional K eigen vectors into original dimensionality and do projection along these eigen images to find novel features for classification.

End.

Multivariate Alteration Detection (MAD)

The Features are signal representatives. Therefore, extracting the appropriate feature is an important task before classification. The appropriate feature extraction method converts one or more signals to the feature vector. Specific features are expected to contain characteristics to represent the signal and distinguish from others. There are several types of features depending on the time domain and the frequency range of the signal. The widely used domain and frequency features such as MAD, mathematical equations to extract MAD features are as follows [27]:

1. Mean Absolute Deviation:

In dataset, n is number of observations like $x_1, x_2, x_3 \dots x_n$. The mean of the dataset illustrate in equation (1)

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k \quad (1)$$

Therefore, the absolute distances from the mean of the data (\bar{x}) is

$$\sum_{k=1}^n |x_k - \bar{x}| \quad (2)$$

From this viewpoint, the definition of mean absolute deviation is average of absolute distances from the mean of the data and that is why MAD is calculated as following:

$$MAD = \frac{\sum_{k=1}^n |x_k - \bar{x}| f_k}{n} \quad (3)$$

Where, f_k depicts the frequencies corresponding to the given values or observations.

2-Standard Deviation: This feature quantifies the deviation of the signal points with respect to its mean value. Mathematically can be illustrate in the following equation,

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

3-Median Absolute Deviation (MedAD):

MedAD is an amount of a data set quite similar to MAD except that the absolute deviation is measured from the average value point of view

rather than the average of that data as shown in equation (5).

$$MedAD = \text{median}(|x_k - \text{median}(x)|) \quad (5)$$

4-L1-Norm: L1-norm also known as mean norm or least absolute deviations. It is calculated from the sum of the absolute values of the dataset.

$$L1\text{-norm} = \sum_{k=1}^n |x_k| \quad (6)$$

5- L2-Norm: L2-norm represented as mean-square norm or least-squares norm and illustrates as the square root of the sum of the absolute values of the dataset.

$$L2\text{-Norm} = \sqrt{\sum_{k=1}^n |x_k|^2} \quad (7)$$

6- Max Norm: Max norm is also represented as infinity norm or uniform norm and described as the maximum of the absolute values of the dataset.

$$L_\infty\text{-Norm} = \max |x_i| \quad (8)$$

7- Power Spectral Density: P is the average power of a signal $x(t)$, T is the total time period,

$$P = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T |x(t)|^2 dt \quad (9)$$

8- Spectral Entropy: The spectral entropy represents the distribution of the spectral power of the signal. This concept is built on the foundation of Shannon's entropy. This feature is widely used in medical signal. Suppose, the considered signal is $x(n)$ and its power spectrum is $s(m) = |X(m)|^2$, $X(m)$ where is the discrete Fourier transform of $x(n)$. Then, the probability distribution $P(m)$ is calculated as,

$$P(m) = \frac{s(m)}{\sum_i s(i)} \quad (10)$$

Consequently, the spectral entropy H is presented as,

$$H = -\sum_{m=1}^N p(m) \log_2 p(m) \quad (11)$$

Gray Level Co-Occurrence Matrix (GLCM):

Statistical approach characterizes stochastic properties of the spatial distribution of grey levels in the image. GLCM (Gray Level Co-Occurrence Matrix) is statistical approaches which are commonly used to extract the texture feature and find the spatial dependencies. In GLCM, texture is represented as matrix by using grey levels from the image. It implies the roughness and repetition of the texture in image. Statistical approach is easy to implement as compared to other approaches [28]. In 1973 Haralick introduced 14 statistical features. These features are generated by calculating the features for each concurrent matrix obtained using the directions $0^\circ, 45^\circ, 90^\circ,$

and 135°, then calculating the average of these four values. The distance parameter can be specified as one or higher.

A vector of these 14 statistical features is used to characterize the co-occurrence matrix contents [29], only 13 of which are defined as follows [30]:

1- Mean: Mean is the measure of the average intensity value of the pixels present in the region.

$$\frac{1}{n} (\sum_{i=1}^n x_i) \quad (12)$$

1- Standard deviation: Standard deviation is the measure of how much that gray levels differ from its mean.

$$\sqrt{\frac{1}{n} (\sum_{i=1}^n (x_i - \bar{x})^2)} \quad (13)$$

2- Contrast: Contrast is the measure of the difference between the brightness of the objects or regions and other objects within the same field of view.

$$\sum_{i,j} |i - j|^2 p(i, j) \quad (14)$$

4- Correlation: Correlation is the measure of degree and type of relationship between adjacent pixels.

$$\sum_{i,j} \frac{(i-\mu)(j-\mu)p(i,j)}{\delta_i \delta_j} \quad (15)$$

5- Energy: Energy is the sum of squared elements in the Gray level co-occurrence of Matrix.

$$\sum_{i,j} p(i, j)^2 \quad (16)$$

6- Homogeneity: Homogeneity is the closeness of the distribution of elements in the GLCM.

$$\sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (17)$$

7- Entropy: Entropy characterizes the texture of the image.

$$E = \sum (p \cdot \log_2(p)) \quad (18)$$

8- RMS: RMS is the measure of root mean square value of an image.

$$x_{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^N |x_n|^2} \quad (19)$$

9- Variance: Variance is the measure of variance value of an image.

$$\frac{1}{n} (\sum_{i=1}^n (x_i - \bar{x})^2) \quad (20)$$

10- Smoothness: Smoothness is a measure of relative smoothness of intensity in an area.

11- Kurtosis: Kurtosis is a measure of peaks distribution related to the normal distribution.

$$K = \frac{E(x-\mu)^4}{\delta^4} \quad (21)$$

12- Skewness: Skewness is a measure of asymmetry in a statistical distribution.

$$S = \frac{E(x-\mu)^3}{\delta^3} \quad (22)$$

13- IDM: Inverse difference moment is a measure of image texture usually called homogeneity. IDM features get the measure of the closeness of the distribution of GLCM elements to the GLCM diagonal.

$$M_K = E(X - \mu)^K \quad (23)$$

Features Fusion

Different feature combination can provide different information. Selective feature fusion focuses on selecting the most relevant feature for image [31] using methods (PCA, MAD, and GLCM) and merging these features.

Size of (PCA) feature vector: 1x192.

Size of (MAD) feature vector: 1x16916.

Size of (GLCM) feature vector: 1x13.

Fusion feature vector (MAD+PCA): 1x17108.

Fusion vector (PCA+GLCM): 1x205.

Fusion vector (PCA+MAD+GLCM): 1x17121.

Fusion vector (MAD+GLCM): 1x16929.

After Fusion, the dataset is split into 80% for training and 20% for testing.

Training

In training process, the network is trained using the architecture defined by layers Convolution Neural Network (CNN), one vector of these methods PCA or MAD or GLCM or fusion is put in the first layer of the CNN network and then pass through the overall network layers.

Only two layers of CNN are used [32]:

1- Convolutional Layer

Convolutional Layer is the first step in the convolution neural network model that takes one of the vectors (fusion vector, PCA vector, GLCM vector, MAD vector) as matrix 1D of the images as an input to extract another feature. This performed by using convolution layer with filter receptive field size of [3x3] and the number of the filters are 50 filters, this results in 50 different features maps for a single input. The convolution layer followed by rectified linear unit (Relu).ReLU is an element wise operation (applied per pixel) and replaces all negative pixel values in the feature map by zero. The purpose of ReLU is to introduce non-linearity in CNN, since most of the real-world data CNN to learn would be non-linear.

2- Fully Connected Layer

The Fully Connected layer is a traditional multi-Layer perceptron that uses a softmax activation

function in the output layer where every neuron in the previous layer is connected to every neuron on the next layer. The output from the convolutional represents high-level features of the input image. The purpose of the Fully Connected layer is to use these features for classifying the input image into various classes. The image classification task a set out to perform has four possible outputs. Fully Connected Layer -each node is connected to every other node in the adjacent layer. The optimizer (ADAM) is used to update weight parameters to minimize the loss function using with Initial Learn Rate 0.001[33].

The epoch is a measurement of how long the network will continue to process, 50 epochs is used in paper. One epoch means that the entire dataset is being utilized to train the network that passed via the network once [34]. Table 1 shows steps of CNN.

Table 1. Steps of CNN

No.	Steps	Purpose
1	Convolutional Layer and Relu layer	Extract features by using filter size of [3x3] with 50 filters
2	Fully Connected Layer and softmax layer	Classifying the features of images in various classes (4 class) based on the training dataset
3	Adaptive Moment Estimation	Update weight parameters to minimize the loss function using with Initial Learn Rate 0.001
4	Epochs	Measurement of how long the network continue to process (50 epochs)

Testing

In testing process, the testing images uses the trained model architecture and goes through all of its layers and by using the saved parameters included the weights that the network reached and filter size and number, then the classifier predicts class labels for testing dataset using the trained network.

EXPERIMENTS AND EVALUATION OF RESULT

In this paper, tested Corel-10K dataset which are freely available for research works, where 400 samples were used for dataset. Type of images are vehicles which include four categories (Cars,

Ships, Agricultural machines, Buses), for each category there are (100) samples. In deep learning, dataset is a very important element when training a model. After training, it is necessary to test the performance of the model by using test images. So training dataset and testing dataset are needed. For this, the dataset is divided into two sets. Training dataset is consisted of 320 images for training the network that means it is with percentage of 80% of the total dataset. On the other hand, the testing dataset consisted of 80 images of the dataset for testing that means it is with percentage of 20% of the total dataset.

All images in datasets are in RGB color space, with size 192x128 and 128 x192 respectively. The first phase in the proposed system is preprocessing. This phase contains five steps which are image resize , convert image to double Precision, convert image into HSV model and select value channel (gray image), remove noise by median filter and using Haar Wavelet transform for images compression.

Second phase is concern with extracting features of images using three method (PCA, MAD, and GLCM) as vectors and fusion of these vectors(PCA+MAD),(PCA+GLCM),(MAD+GLCM)and(PCA+MAD+GLCM) as shown in Figure 3.



Figure 3. Interface of the proposed system.

Third phase is training using deep learning algorithm 1 D Convolution Neural Networks (CNN) that have filter size [3x3], 50 number of filter, Learn Rate 0.001and 50 epochs for training data in four classes. The results of dataset training using Fusion (PCA, MAD) illustrates in Table 2 and Figure 4.

Table 2. Training dataset result using (PCA+MAD)

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:12	22.66%	9.2667	0.0010
25	50	00:05:31	86.72%	0.8169	0.0010
50	100	00:10:14	100.00%	0.0009	0.0010

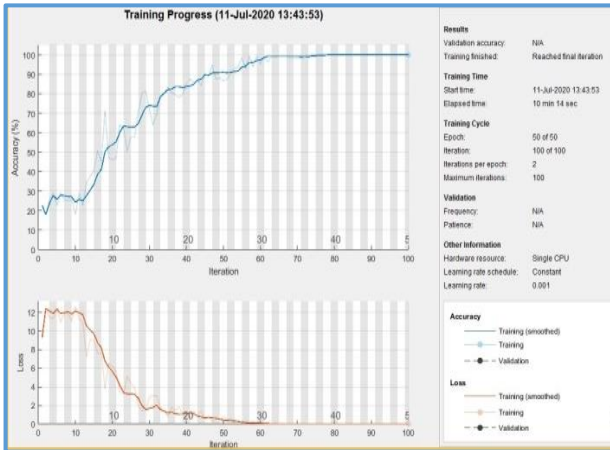


Figure 4. Graph of training dataset result using (PCA+MAD)

Then classifier where predicts class labels for testing dataset using the trained network.

The fourth phase is evaluation where accuracy, precision, recall and F-measure are used to evaluation [35].

1- Accuracy (Ac): is the capability of the classifier to choice all cases that necessity to be selected and reject all cases that necessity to be rejected.

$$Ac = (TP+TN)/(TP+TN+FP+FN) \quad (24)$$

2- Precision: Precision (P) is described as the retrieved similar images' count to the retrieved similar a dissimilar images' count in database.

$$P = TP/ (TP+FP) \quad (25)$$

3- Recall: Recall is described as the total count of retrieved similar images to the whole count of similar images in database.

$$Recall=TP/(TP+FN) \quad (26)$$

4- F-measure (FM) is used to measure efficiency and success based on the values of precision and recall. In order to have a single measure with higher effectiveness F-measure is calculated by combining precision and recall. It is a harmonic mean that gives a precise result.

$$FM= 2 *(P*Recall)/ (P+Recall) \quad (27)$$

For evaluation the proposed system, Figure 5 illustrates the confusion matrix using Fusion

(PCA+MAD) and the performance of each method illustrated inTable3.

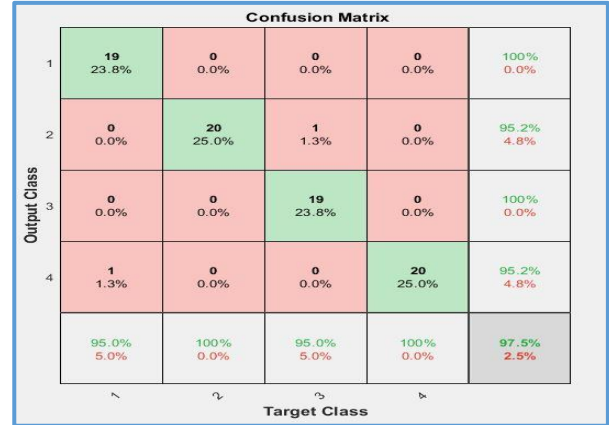


Figure 5. Confusion Matrix using (PCA+MAD).

Table 3. performance of proposed system.

Evaluation					
Method	Accuracy	Average of Precision	Average of Recall	Average of F-measure	Time in Second
MAD	95%	0.95	0.9518	0.95	861
PCA	90%	0.9	0.9122	0.9	62
GLCM	55%	0.55	0.5225	0.5182	53
PCA+GLCM	88.75%	0.8875	0.8950	0.8877	59
PCA+MAD	97.5%	0.975	0.9762	0.975	683
MAD+GLCM	93.75%	0.9375	0.9375	0.9372	673
PCA+MAD + GLCM	96.25%	0.9625	0.9637	0.9628	742

Experimental results shows that (MAD+PCA) gave the best results, then (PCA+MAD+GLCM) got second order but requires more computational time than (MAD+PCA), MAD got third order but requires more time than (MAD+PCA), (PCA+MAD+GLCM), while PCA based dimension reduction method got fourth order where gives high performance and less time, while GLCM give weak results in very little time. Based on classifier where predicts class labels for test image using the trained network, the similar images of testing image can be retrieve from database where 10 images is retrieved as shown in Figure 6.



Figure 6. Retrieve Images using (PCA+ MAD).

CONCLUSIONS

One of the most critical issues in CBIR systems is how to set an optimal method to retrieval images similar (query image vs. indexed image in Database) based on training phases for the given dataset. In this paper, performance tested using the Precision, Recall, F-Measure, and Accuracy for evaluation.

The performance has been compared with other previous studies; the results concluded that the accuracy of the proposed system is higher than other and It can also be combined could combined with other techniques to enhance such computer vision applications like object detection, object recognition, and image classification. PCA is dimension reduction method by calculating the eigenvectors and eigenvalues of the covariance matrix, principal components of the dataset can be obtained. These principle components hold the feature factor information of the original image where this method allows significant savings of storage space, which can be critical in applications and in processing large dataset.

Features of different methods can be combined to improve results at the expense of time. For future work, improvement training of dataset by increasing efforts to change filters number and size, epochs and patch size of CNN.

Repeat the experiments using multiple dataset because of the challenges that pose on the system results.

REFERENCES

[1] Benchikh H.,Razi S., and Korichi H.,“ Content-based image retrieval,”, 2017.

[2] R.MADANA M., and A.RAMA M. R., “CCBIR: A Cloud Based Implementation of Content Based Image Retrieval”, E-ISSN:2224-2872 333 Volume 14, 2015.

[3] Anas F. A.,“A COMPARATIVE STUDY OF HUMAN FACES RECOGNITION USING PRINCIPLE COMPONENTS ANALYSIS AND LINEAR DISCRIMINANT ANALYSIS TECHNIQUES”, Vol. 20, No.05, ISSN 2520-0917, 2016.

[4] Manisha V., and Balasubramanian R.,“Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval,” DOI10.1007/s11042-017-4834-3, 2017.

[5] S.Selvam, and S.Thabasu K.,“A New Architecture for Image Retrieval Optimization with HARP Algorithm,” Asian Journal of Computer Science and Technology,ISSN: 2249-0701 Vol. 6 No.1, 2017.

[6] Dr.K.Meenakshisundaram, and G.Vijaiprabhu,“AN ENHANCED CENTER SYMMETRIC LOCAL BINARY PATTERN TECHNIQUE FOR IMAGE RETRIEVAL USING EUCLIDEAN DISTANCE”, (IJCRCT) ISSN: 2395-5325Volume 3, Special Issue 1,2017.

[7] Prof.S.Govindaraju, and G.P.Ramesh K., “Content Based Image Retrieval Using Hierarchical and Fuzzy C-Means Clustering” (IJCTT) – Volume 44 Issue 2- February 2017.

[8] Umesh D. D., and M. S. Sh., “Face-based Document Image Retrieval System,” ICCIDS,2018.

[9] Mamta M., Shrutika M., and Shruti T.,“Effectiveness of Saas Cloud Model For Retrieving Images From CBIR System”, (IOSRJEN) ISSN (e): 2250-3021, ISSN (p): 2278-8719 PP 70-74, 2018.

[10] Sidi A. M., Mohammed A. B., El.Wardani D., Saïd Mahmoudi and Mohammed B. , “Cloud-Based Image Retrieval Using GPU Platforms,” doi: 10.3390/computers8020048,2019.

[11] Vo T. H. T.,Nguyen M. H. , Pham B. Q., Nguyen T. S., and Nguyen T.,“ Adaptive Content-Based Medical Image Retrieval Based On Local Features Extraction In Shearlet Domain”, EAI, 2019.

[12] Nitika S., and Sonika J.,“ A REVIEW ON CONTENT BASED IMAGE RETRIEVAL,” ISSN2277-3061 Volume15Number14INTERNATIONAL JOURNAL OF COMPUTERS &TECHNOLOGY,2017.

- [13] A.Rama, and C.Nalini, "A REVIEW ON VARIOUS IMAGE RETRIEVAL TECHNIQUES AND ITS APPLICATIONS," ISSN: 0975-766X CODEN: IJPTFI, 2016.
- [14] Naveena AK, and N.K.Narayanan, "Sub-Block based Color Moments, Wavelet and Edge Histogram for Image Retrieval," Volume 4 Issue 3; IEEE Network, 2017.
- [15] Mamta M., Vinayak A. B., and Krunali V., "Hybrid Wavelet Based CBIR System using Software as a Service (SaaS) Model on public Cloud," *Procedia Computer Science* 79 -278 – 286, 2016.
- [16] Mr. Yogen M. L., and Prof. Sushant J. P., "A novel method for Content Based Image retrieval using Local features and SVM classifier," (IRJET) e-ISSN: 2395-0056 Volume: 04 Issue: 07, 2017.
- [17] Priya R., "Machine Learning with Optimized Features for Enriched Performance of Content Based Image Retrieval," *Shodhganga*, 2015.
- [18] Shikha K., "An Hybrid Algorithm to Currency Recognition using Digital Image Processing and Feature Extraction," *IJAST*, Vol. 27, No. 1, pp. 244-250, 2019.
- [19] Higham, Nicholas J. Pranesh, and Srikara "SIMULATING LOW PRECISION FLOATING-POINT ARITHMETIC," The University of Manchester Manchester, M13 9PL, UK, 2019.
- [20] Asia M. N., "Color to Grayscale Image Conversion Based Dimensionality Reduction with Stationary Wavelet Transform," DOI: 10.1109/AIC-MITCSA, 7759946, 2018.
- [21] Muthana H., "Digital Image Processing," 2017.
- [22] Qin X., Qiang Z., Duo H., and Jinpei L., "Removal of Salt and Pepper Noise in Corrupted Image Based on Multilevel Weighted Graphs and IGOWA Operator," doi.org/10.1155/7975248, 2018.
- [23] Raniah A. M., Kawther T. S., and Haitham S. C., "Feature Extraction Based on Wavelet Transform and Moment Invariants for Medical Image," DOI: doi.org/10.31695/IJERAT, 3315, Volume 4, Issue 8, 2018.
- [24] S.C.Ng, "Principal component analysis to reduce dimension on digital image," *Procedia Computer Science* 113–119, 10.1016/j.procs.06.017, 2017.
- [25] N.Santhi, K.Annbuselvi, and S. Sivakumar, "Performance Analysis of Feature Extraction Techniques: PCA and LDA for Face Recognition," ISSN (Online) 2394-2320, (IJERCSE) Vol 5, Issue 3, 2018.
- [26] Ningthoujam S. D., and K.Hemachandran, "Face Recognition Using Principal Component Analysis," (IJCSIT), Vol. 5 (5), 6491-6496, 2014.
- [27] Md.Asadur R., Md.Foisal H., Mazhar H., and Rasel A., "Employing PCA and t-statistical approach for feature extraction and classification of emotion from multichannel EEG signal,"/doi.org/10.1016/j.eij.2019.10.002, *Egyptian Informatics Journal*, 2020.
- [28] Imran M., Qasim, and Nasrullah I., "A Novel Technique for Region-Based Features Similarity for Content-Based Image Retrieval," DOI: 10.22581/muet1982.1802.14, *Mehran University Research Journal of Engineering & Technology*, 2018.
- [29] Faleh H. M., and Wafaa A. A., "Texture Features Analysis using Gray Level Co-occurrence Matrix for Abnormality Detection in Chest CT Images," *Iraqi Journal of Science*, Vol. 57, No.1A, pp: 279-288, 2016.
- [30] Asuntha, P. B., Ainthaviasari, Bharath K., and Srinivasan, "Feature extraction to detect Bone Cancer Using Image Processing," ISSN: 0975-8585, *RJPBCS*, 2017.
- [31] T.Saikumar, S.Venkatesh, and Md.Ameen U., "Comparison Techniques of Image Fusion in Image Segmentation," Volume 3, Issue 1; Spl. Issue of IC3T @ISSN: 2248-9584, 2014.
- [32] Ujjwalkarn, "An Intuitive Explanation of Convolutional Neural Networks," *nlp, data science*, 2016.
- [33] Renu K., "Overview of different Optimizers for neural networks," 2019.
- [34] Y.Jones "Deep Learning for Image Enhancement and Visibility Improvement," *Eps.Hw.Ac.Uk*, 2018.
- [35] Ning L., Martin S., and Yuchen G., "A Systematic Review of Unsupervised Learning Techniques for Software Defect Prediction," arXiv:1907.12027v4 [cs.SE], 2020.