Received: November 2, 2022. Revised: December 18, 2022.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Unsupervised Content Based Image Retrieval Using Pre-Trained CNN and PCNN Features Extractors

Mahmoud S. Sayed[1*] **Ahmed A. A. Gad-Elrab**[1, 2] **Khaled A. Fathy**[1] **K. R. Raslan**[1]

[1] *Faculty of Science, Al-Azhar University - Cairo, Egypt*
[2] *Faculty of Computing and Information Technology, King Abdul-Aziz University, Jeddah, Saudi Arabia*
*Corresponding author's Email: mahmoud.saeed@azhar.edu.eg

Abstract: Content-based image retrieval (CBIR) is a popular approach for searching and retrieving digital images from labeled or unlabeled image collections using image content features. The most important parts of the CBIR system are the computational complexity and the retrieval accuracy. Many studies have been conducted to increase the accuracy of image retrieval systems. In this paper, we propose a new fully unsupervised content-based image retrieval (CBIR) technique to increase the efficacy of image retrieval using a clustering approach on unlabeled image dataset. The proposed method combines features extracted from both pre-trained AlexNet model and pulse-coupled neural networks (PCNN) to extract high and low level features. Then principal component analysis (PCA) is performed on AlexNet's features and these combinations are fed to the K-means algorithm after normalization process. Then Euclidean distance is used to measure the similarity between query and stored images within the same cluster. Finally top similar images are ranked and retrieved. Experimental results on the benchmark Corel-1k and Corel-10k datasets show that the proposed method achieves high precision values of 92.8% and 64.8% respectively, on the top 20 retrieval levels compared to other methods.

Keywords: Content based image retrieval, Convolutional neural networks, AlexNet, CNN, Pulse coupled neural network, Principal components analysis, K-means, Euclidean distance.

1. Introduction

On the internet, images are widely used, sharing them on various social media has produced a large number of images. Several ways of effectively searching and retrieving images have been developed. The most typical approach for retrieving images is known as text-based image retrieval system (TBIR), in which the system uses meta-data associated with the images, such as keywords, tags, labels, or descriptions, to perform the retrieval process. Although the TBIR takes significant time and effort, since the images must be manually annotated, which is a time-consuming process, it usually doesn't work effectively. This is because picture labelling and tagging with describing text does not always accurately reflect what an image represents because the same word can have multiple meanings in different contexts. Another image retrieval approach must be developed to solve these drawbacks of the text-based technique.

CBIR (content-based image retrieval) is a system that finds related images of a query image in the absence of a caption or image description from an image database. Thus, any CBIR system depends on the extracted features, which are used to evaluate the similarity of the contents between the two images, e.g., textures, colours, shapes, etc. To recognise images, humans use high-level semantics, whereas machines use low-level visual features. The difference between high-level semantics (individuals, things, actions, etc.) and low-level features (e.g., pixels) is called semantic gap. Furthermore, due to the fact that different items may have the same colour,

texture or shape, all retrieved images are not actually similar to the query image during similar image retrieval. It's difficult to distinguish items that are the same colour and texture. There is no general algorithm that can understand images as well as a human.

The convolution neural network (CNN) is recently considered as one of the most effective learning algorithms for understanding image content, with efficient performance in image classification [1], face recognition [2–3], and image retrieval [4]. One significant limitation that faces this new generation of neural networks is that they require a lot of training data, which isn't always accessible across various domains of knowledge [5]. Because of this, the most recent generation of pre-trained CNN has been determined to meet the requirements for well-trained CNNs and to provide highly accurate performance. Since these pre-training CNN models were trained on large-scale annotated natural image data collections in ImageNet [6], they have the ability to transfer their knowledge and they have been effectively utilised in several image processing application areas [7-10].

Combining several features in a CBIR often increases accuracy but also increases retrieval time. It is important to provide fast retrieval with high accuracy in an online CBIR. The clustering algorithm is a powerful processing method that can quickly classify massive resources in a short period of time [11]. Clustering is the process of separating a dataset into groups, whereas images that are similar to each other will be clustered together in the same cluster.

In this paper, a new fully unsupervised method is proposed to obtain high precision and recall scores on un-labeled image dataset. The proposed method is based on a combination of features extracted from pre-trained AlexNet Convolutional neural network followed by principal components analysis (PCA) for dimensionality reduction integrated with features extracted from Pulse coupled neural network (PCNN) to provide a robust feature representation for image retrieval tasks and then feed these combinations to the K-means algorithm after normalization in order to group the unlabeled image dataset into different clusters, whereas images that are similar to each other will be clustered together in the same cluster. Finally, Euclidean distance is used to measure the similarity between a query and stored images grouped together in the same cluster in order to get the most similar images relevant to the query images.

This paper is organised as follows, section 2 presents a related work on unsupervised CBIR systems. Section 3 presents the proposed image retrieval method. Section 4 presents experiments and results. Section 5 provides a conclusion and future work.

2. Related work

Many methods have been proposed to improve the performance of the CBIR system. In this section, we briefly survey some existing CBIR work in the unsupervised image retrieval domain. Authors in [12] proposed the CBIR method by clustering binary signatures of images. The cluster graph is used in this paper to improve the speed of the clustering process. This approach presents the segmentation method based on low-level visual features, including colour and texture of the image. In order to find similar images, the similarity measure between the images is based on binary signature. The accuracy obtained by this approach, which is equal to the precision score, was 82.6% on the Corel-1k dataset, which is still quite low.

Authors in [13] proposed an unsupervised CBIR approach using both global and local features to describe the content of an image. This method uses a combination of SURF detector and descriptor with color moments as local features and modified GLCM as global features. Color moments are computed in the region surrounding the SURF blob points, yielding local colour information. Both local and global features are used because only local features are insufficient when the variety of images is large. The accuracy, which is equal to the precision score, obtained by this approach was 70.48% for the top 20 retrieval images on the Corel-1k (WANG) dataset, which is still quite low.

Authors in [14] proposed a CBIR system by using ordered-dither block truncation coding (ODBTC) and a phase congruency feature (PCF). Combining the PCF and ODBTC features improves CBIR system usage in various visual data processing domains. This yields a better CBIR system, which assists in the reduction of storage space, decreases retrieval time, and increases the accuracy of the system. The experiments were tested on the Corel 1K dataset and achieved a mean average precision of 88.21% which is still quite low.

Authors in [15] proposed several CBIR models with several features in a combination of two and three. The models with two features are labelled as CS, CT, and ST models based on the combination of color with shape, color with texture, and shape with texture, respectively. The three-feature based model is developed using color, shape, and texture and is labeled as the CST model. The open-source image dataset COREL is used to evaluate the performance of these various models for image retrieval. With the

data subset evaluated, it is found that among two features-based models, the CT model gives the best average performance of 73.8 %, 72.5%, and 64.7 % for the top 20, 30, and 100, respectively. It is also found that combining three features in the CST model improves image retrieval accuracy by 80.8 %, 79.9%, and 69.4% for top 20, 30, and 100, respectively. The precision score still needs to be improved.

Authors in [16] proposed a CBIR system that divides or groups the image collection into a subset of relevant images. This paper uses the hybrid K-means moth flame optimization technique for image clustering (KMFO), which overcomes the drawbacks of the conventional K-means clustering algorithm by assigning the optimum number of clusters and cluster centroids using the number of flames and flame values obtained in MFO. It uses the HSV colour histogram, colour moments, the colour correlogram, the GLCM, the wavelet transform, the dominant colour, and region-based descriptors as feature vectors. The experiments were tested on the Corel 1K dataset and achieved a mean average precision of 81.3% on the top 20 retrieval images, also achieved an average recall score of 16.20% on the top 20 retrieval images which are still quite low.

Authors in [17] proposed a CBIR (content-based image retrieval) method based on hierarchical clustering on low-level features. Low-level features consisting of color, texture, and shape are extracted and then clustered hierarchically. The resulting clusters are then validated to obtain their optimal number. In the retrieval process, the query image features are extracted and compared with the cluster centroid for each feature. The scores of query results on each feature are normalized, and then the normalized scores are weighted to get the total score. The experiments were tested on the Corel 1K and Corel 10k datasets. Based on the experimental results on the Corel-1k dataset, the average precision scores obtained by this method are 0.88, 0.81, 0.76, 0.71, 0.68, and 0.56 for L of 10, 20, 30, 40, 50, and 100, respectively and the average recall scores obtained are 0.09, 0.16, 0.23, 0.29, 0.35, and 0.60 for L of 10, 20, 30, 40, 50, and 100. On the experiment on the Corel-10k dataset, the average precision and recall scores were also measured on the number of retrieval L images of 10, 20, 30, 40, 50, and 100. The average precision score obtained by the proposed method was 0.64, 0.62, 0.57, 0.53, 0.50, and 0.40 for L of 10, 20, 30, 40, 50, and 100, respectively. Furthermore, the average recall scores are 0.06, 0.11, 0.15, 0.19, 0.22, and 0.35 for L of 10, 20, 30, 40, 50, and 100, respectively. The precision and recall scores still need to be improved.

Authors in [18] present a deep convolutional neural network-based model called MaxNet for content-based image retrieval. The proposed MaxNet model consists of twenty-one convolution layers that are iterated in a structured manner to extract the most information from the images. The dataset is fed into the network after the MaxNet model has been trained, and the feature vectors are taken from the last layer of the proposed MaxNet model. During the retrieval process, the query image is feed-forwarded through the network. The extracted features of the query image are compared to feature vectors from the full dataset. The experiments were performed on the Corel 10K dataset and obtained mean average precision and recall scores of 45.8% and 9%, respectively, on the top 20 retrieval images, which are quite low.

Authors in [19] proposed a CBIR approach that detects and extracts consistent zones from images using texton templates. The consistent zone is then used to obtain the dominant colour descriptor (DCD). Furthermore, the Hu moments feature's translation and rotation invariance are used to extract shape information in the same consistent zone of the image. This study used various levels of quantization in extracting CDCs for retrieval during the experimental stage. The experiments were tested on the Corel 10K dataset and achieved mean average precision and recall scores of 46.2% and 9% respectively, on the top 20 retrieval images which are still quite low.

Based on the observations of the above CBIR methods, this paper introduces a new CBIR method based on features extracted from both the pre-trained AlexNet model and pulse-coupled neural networks (PCNN) to extract high and low-level features with a clustering approach in order to obtain high precision and recall scores.

3. Methodology

In this paper, a new unsupervised content-based image retrieval method called clustering content-based image retrieval (CCBIR) based on unsupervised learning is proposed. In the rest of this section, the basic idea of CCBIR is introduced, then the steps of the proposed approach are described.

3.1 Basic idea

The simple mechanism is to extract some useful features from the query and database images, and find images that have a similar set of features in order to retrieve similar images from a large database based on a query image using clustering approach, which is one of the unsupervised learning methods that can be

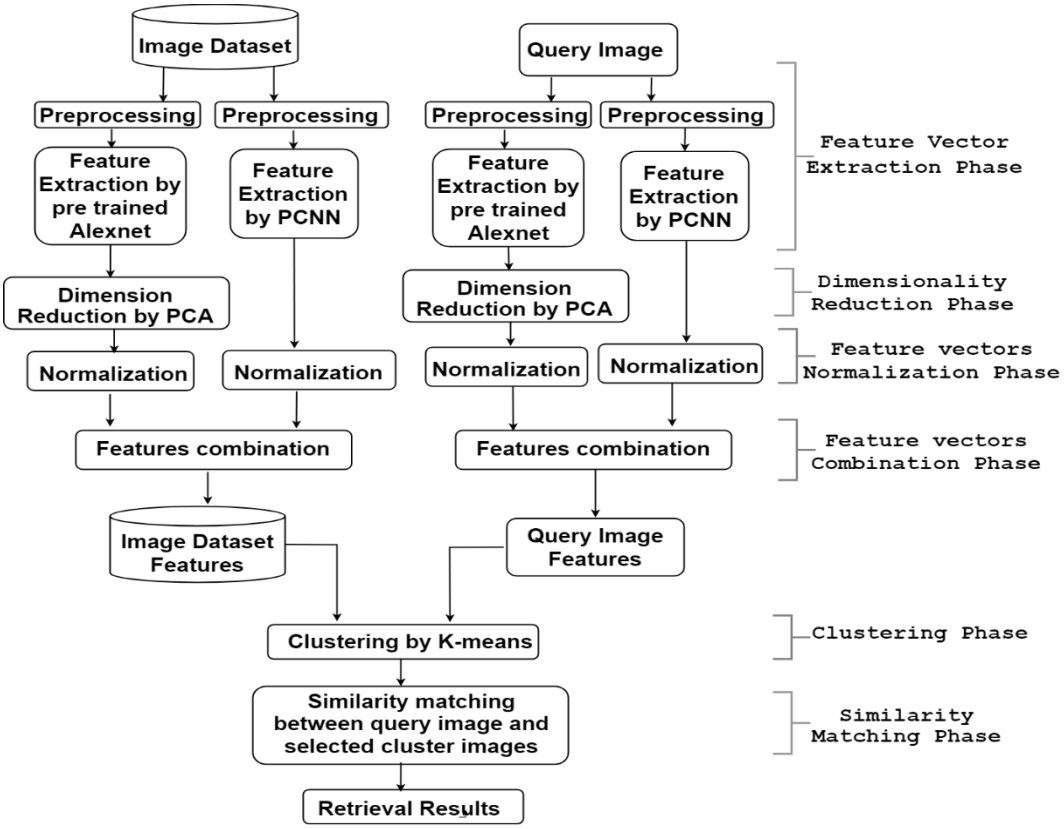

Figure 1 Six phases of the proposed CCBIR approach

used in CBIR systems to reduce the search space of retrieval tasks. To satisfy this goal, the basic idea of CCBIR is based on four main issues, which are:

(1) Using a pre-trained AlexNet model combined with a pulse coupled neural network (PCNN) to extract high-level and low-level features. This combination provides a robust feature representation of images.
(2) Using principal components analysis (PCA) to reduce the features extracted from the pre-trained AlexNet model.
(3) Using the K-means algorithm to group the un-labeled image dataset into different clusters using these combined feature vectors after the normalization process, whereas images that are similar to each other will be clustered together in the same cluster.
(4) Using Euclidean distance as a similarity measure between query and stored images that are grouped together in the same cluster to reduce processing time and improve retrieval accuracy. Then they are ranked to get the top similar images relevant to the query image.

3.2 The proposed approach

The proposed CCBIR approach consists of six phases, which are feature extraction, dimensionality reduction, feature vector normalization, feature vectors combination, unsupervised learning-clustering and similarity determination. Fig. 1 shows these six phases of the proposed CCBIR approach. These phases are described as follows:

3.2.1. Features extraction phase

In this phase, in order to reduce the rich content and large data input of the images while preserving the entire image content representation, features must be extracted carefully. Thus, the feature extraction process can reduce processing time and increase retrieval accuracy. In our proposal, features are extracted from both the pre-trained AlexNet convolutional neural network (CNN) and the Pulse coupled neural network (PCNN) for robust feature representation of images, which are described as follows.

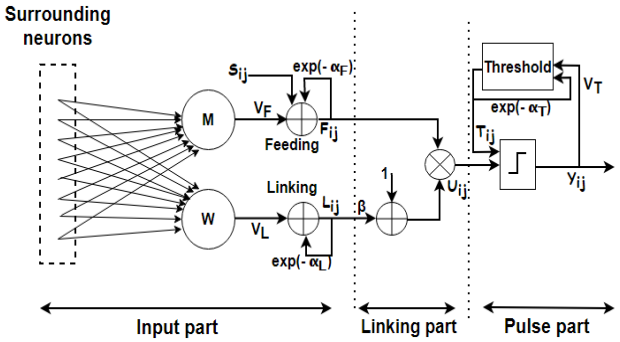

Figure. 2 PCNN's neuron model

3.2.1.1. Pre-trained AlexNet model

AlexNet [20] is a convolutional neural network that has a significant effect in the machine learning area, especially in the deep learning application. In this work, we are utilising pre-trained CNN's AlexNet to extract the important and high-level features from the images. The AlexNet has already been trained on the ImageNet dataset. The full ImageNet dataset has more than 15 million images and 22,000 class labels, making it significantly larger than the typical training dataset. This can provide a more accurate classifier when working with images that the ImageNet dataset may have already seen. Transfer learning [21] is a method for feature representation from a pre-trained model, facilitating us to use the learnt weights of an already trained model to solve similar problems rather than starting the training process of a model from scratch. In this way, we save time by using past learning through learnt weights. Additionally, results obtained are usually much better compared to training from scratch. Thus, we use a pre-trained CNN's AlexNet for feature extraction in this work.

The AlexNet model [20] consists of eight trained layers. The first five are convolutional layers, while the remaining three are fully connected layers. To accelerate the train, the rectified linear unit (ReLU) is applied after all convolutional and fully connected layers. Dropout is applied before both the first and the second fully connected layer. The main goal of using deep CNNs for retrieval is to extract feature representations from a pre-trained model by feeding images into the input layer and taking activation values from either the convolutional layers or the fully connected layers, which are intended to capture high-level semantic information.

In this work, images are read and resized to K x K x z (e.g., 227 × 227 × 3) and the 7th layer of the architecture is used for feature extraction, with 4096 dimensions/features per image. CNN's Alex Net process starts over the image dataset for feature extraction. After that, the features that were extracted are saved for later processing.

3.2.1.2. Pulse coupled neural network

The pulse coupled neural network (PCNN) is a synthetic model designed from studies of the visual cortex of small mammals conducted by e.g. Eckhorn et al [22]. The PCNN model has a big advantage in that it can run without any training [23]. Also A 2D image can be transformed by the PCNN into a 1D periodic time signal known as the image's signature. The implementation of the PCNN was first carried out by Johnson [24].

The PCNN is a single-layer, two-dimensional neural network. An input image pixel is represented by one network neuron. Because of this, the structure of the PCNN is determined by the structure of the input picture. The PCNN's neuron structure is shown in Fig. 2 [25]. This neuron consists of three components: an input part, a linking part, and a pulse generator. Feeding and linking inputs supply the neuron with input signals. Feeding input is the primary input from the neuron's receptive area. The neuron receptive area consists of the neighbouring pixels of the corresponding pixel in the input image. Linking input is the secondary input of lateral connections with neighbouring neurons.

The following equations define the standard PCNN model as an iteration:

$$F_{ij}[n] = S_{ij} + F_{ij}[n-1]\, e^{-\alpha_F} + V_F \sum_{kl} M_{ijkl} * Y_{kl}[n-1] \quad (1)$$

$$L_{ij}[n] = L_{ij}[n-1] e^{-\alpha_L} + V_L \sum_{kl} W_{ijkl} * Y_{kl}[n-1] \quad (2)$$

$$U_{ij}[n] = F_{ij}[n](1 + \beta L_{ij}[n]) \quad (3)$$

$$T_{ij}[n] = T_{ij}[n-1]\, e^{-\alpha_T} + V_T Y_{ij}[n-1] \quad (4)$$

$$Y_{ij}[n] = \begin{cases} 1 & U_{ij}[n] > T_{ij}[n] \\ 0 & Otherwise \end{cases} \quad (5)$$

The two main components F_{ij} and L_{ij} are called feeding and linking in (i,j) position, respectively. The internal activity of a neuron is denoted by U_{ij}, while the dynamic threshold is denoted by T_{ij}. Y_{ij} is the neuron's output. The synaptic weight coefficients are W_{ijkl} and M_{ijkl}, the symbol * is the convolution operator, and S_{ij} represents the intensity of pixel (i,j) in the input matrix, typically this value is normalized. The PCNN neuron's decay constants are α_F, α_L and α_T. The magnitude scaling constants

are V_F, V_L and V_T. The linking coefficient constant is β.

If $U_{ij}[n] > T_{ij}[n]$, the neuron generates a pulse, called a firing time; If not, called unfiring time. After PCNN firing, the total firing number generates the output of PCNN. The PCNN converts the multilevel input image that's represented by a two-dimensional matrix into a series of temporary binary images. Each of these binary images is a matrix with the same dimension as the input matrix, and they are generated by groups of pixels with similar intensity. The aggregate of all activities in a single iteration step produces one value, which represents one feature. We get N features if we have N iteration steps. The length of the vector G represents the vector's quality. The greater N is required to build a high-quality feature vector. But as a result, it will take a long time to generate a single feature vector.

For each iteration step n, the one-dimensional time signal generated from the values of outputs Y_{ij} is defined as:

$$G(n) = \sum_{ij} Y_{ij}(n) \tag{6}$$

This generated time signal is invariant to rotation, dilatation or translation of images [24] and it is a significant advantage of the PCNN. So, in this process, the images are read as grayscale images and resized to K × K × z (e.g., 227 × 227 × 1), Then, each image is fed to PCNN to generate a feature vector with a dimension 1 × n (e.g., 1 × 64) to represent an image.

3.2.2. Dimensionality reduction phase

To accelerate the image retrieval process and improve its performance, dimension reduction on the features extracted from the 7th trained layer (FC layer) of the pre-trained AlexNet model is applied. This will be achieved by using principal components analysis (PCA), which is a powerful technique in data analysis for reducing dimension and obtaining the most variance from data.

PCA is a versatile technique and has been widely used, achieving good results in various applications such as dimensionality reduction, data compression, and feature extraction [26]. The use of the PCA technique has the advantage of reducing the dimensionality of a data set by identifying a new set of variables that are smaller than the original set of variables and helps in the classification of data. By computing the eigenvectors and eigenvalues of the data covariance matrix, principal components can be obtained. The following gives details about the PCA method. Assume that we have a matrix A that includes term weights that were obtained by using feature extraction methods.

$$A = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2k} & \cdots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} & \cdots & X_{nm} \end{bmatrix}$$

Where x_{jk} $(j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, m)$ are the terms of weight that exist in the collection of vectors. Where n is the number of images to be classified and m is the number of weights produced by feature extraction. The steps used by PCA to reduce the dimensionality of matrix A are described as follows:

Step 1: Find the mean of m variables in the matrix A:

$$\overline{X_k} = \frac{1}{n}\sum_{j=1}^{n} x_{jk} \tag{7}$$

Step 2: Calculate the covariance S_{ik} of m variables in the matrix A:

$$S_{ik} = \frac{1}{n}\sum_{j=1}^{n}(x_{ji} - \overline{X_i})(x_{jk} - \overline{X_k}) \tag{8}$$

Where $i = 1, \ldots, m$. Eigenvectors and eigenvalues of the covariance matrix are calculated, and principal components are selected. Then we select the first $d \leq m$ Eigen vectors where d is the desired value corresponding to the d largest eigenvalues of the covariance matrix C. Finally, a matrix M with dimension n×d is represented as:

$$M = \begin{bmatrix} f_{11} & f_{12} & f_{13} & \cdots & f_{1d} \\ f_{21} & f_{22} & f_{23} & \cdots & f_{2d} \\ f_{31} & f_{32} & f_{33} & \cdots & f_{3d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & f_{n3} & \cdots & f_{nd} \end{bmatrix}$$

Where f_{ij} is a reduced feature vectors from the n×m original data size to n×d size.

The PCA algorithm is used in our work to reduce the feature vector size of each image that is extracted from the 7th trained layer (FC layer) of AlexNet CNN from 1× 4096 to 1× M (e.g., 1 x 64) and obtain the maximum variance of data.

3.2.3. Feature normalization

Normalization is necessary to give equal weight to all features in a data set and thus be useful for classification algorithms. The normalization process is done by considering the values in the vector. For instance, if the vector is of size 1×4: $[4, 6, 9, 11]$, we

need to calculate the l2-norm for this vector in order to normalize it, which is $\sqrt{4^2 + 6^2 + 9^2 + 11^2}$ = 15.93. Then divide each of the vector values with this l2-norm: $[\frac{4}{15.93}, \frac{6}{15.93}, \frac{9}{15.93}, \frac{11}{15.93}]$ that is, equal to $[0.25, 0.37, 0.69, 0.56]$.

3.2.4. Feature vectors combination phase

In this phase, the normalized feature vector, which is extracted from pre-trained AlexNet after applying PCA, is combined with a normalized feature vector that is extracted from PCNN, and finally a robust feature vector representation of the image is created. For example, the AlexNet-PCA feature vector with a dimension of 1 × M (e.g., 1 × 64) and the PCNN feature vector with a dimension of 1 × M (e.g., 1 × 64) are combined, and an efficient image descriptor with a dimension of 1 × 2M (e.g., 1 × 128) is created.

3.2.5. Unsupervised Learning - clustering phase

In this phase, in order to reduce the search space and improve the results of the proposed approach, the unsupervised K-means algorithm is used to group the unlabeled dataset into different clusters using the combination of normalized feature vectors. Images that are similar to each other will be clustered together in the same cluster. The k-means algorithm [27] is effective in producing good clustering results for many applications. Using the K-means clustering algorithm, the data is divided into K groups based on features or attributes.

The K-Means algorithm is used to determine the position of a cluster of each image by firstly calculating the image distance with all the centroids using the Euclidean distance method. Cluster mapping is done by selecting the closest distance to all the existing centroids. The distance is calculated using an equation.

$$D(H,I) = \sqrt{\sum_{k=1}^{n}(X_{Hk} - C_{Ik})^2} \qquad (9)$$

Where $D(H,I)$ is the distance of image H to centroid I, X_{Hk} is the value of the feature number k of image H, C_{Ik} is the value of the feature number k of centroid I and n refers to the total number of dimensions.

3.2.6. Similarity determination phase

In this phase, the Euclidean distance Eq. (10) is used for a similarity measure between the query and stored images, and it was performed only on one cluster that the query image mapped to it by calculating the closest distance to all the existing centroids. Euclidean distance is the best choice for a similarity metric due to its ease of calculation and popularity.

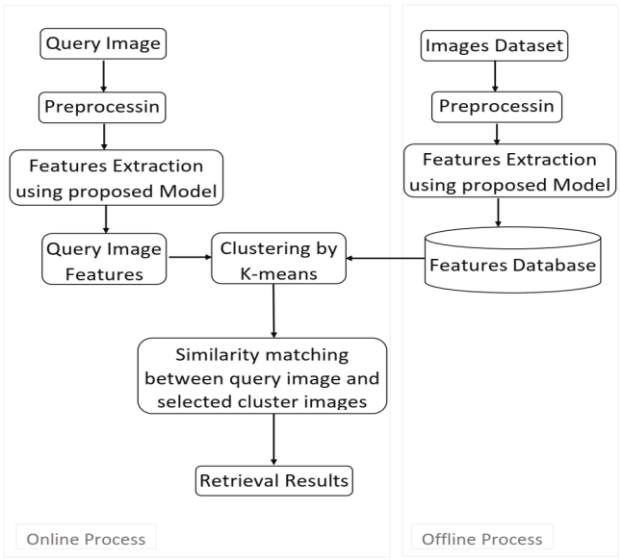

Figure. 3 Overview of the online and offline processes of the proposed CCBIR approach

$$dist(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (10)$$

Where X and Y are the feature vectors of the query image and the feature vector of the image in the database, while x_i and y_i are the value of the feature number k in these vectors and n refers to the total number of dimensions. A set of relevant images were collected and then sorted to rank the images based on their similarities, and then the top ranked images were retrieved.

3.3 The online and offline processes proposed approach

The proposed CCBIR model is described in Fig. 3, including two process modes. These are online processes (on the left in Fig. 3) and offline processes (on the right in Fig. 3). In the offline process, a feature database for each and every image in the database is created. On the other hand, in the online process mode, which is designed based on a user interface, the feature extraction process is done on the query image given by users. From this, the inputted image feature is clustered with all the image dataset features, whereas images that are similar to each other will be clustered together in the same cluster. Then Euclidean distance is used to measure the similarity between query and stored images grouped together in

Figure. 4 From left to right, samples of images from each category in the Corel-1k dataset

Table 1. COREL image database with index values

Class No.	Index Value	Class Name
1	(0-99)	Village & People
2	(100-199)	Beach
3	(200-299)	Buildings
4	(300-399)	Buses
5	(400-499)	Dinosaurs
6	(500-599)	Elephants
7	(600-699)	Flowers
8	(700-799)	Horses
9	(800-899)	Mountains
10	(900-999)	Food

the same cluster. These distance measurements are sorted to rank the images based on their similarities, and then the top ranked images are retrieved.

4. Experiments and results

In this section, we have presented brief information about datasets and experimental results performed using the proposed CCBIR method.

4.1 Datasets

Most of the research in image retrieval uses the Corel-1k [28] and Corel-10k [29] datasets as the benchmark for image retrieval and classification. Therefore, the performance of the proposed CCBIR is examined using these datasets. The detail of each dataset is described as follows:

4.1.1. Corel-1k dataset description

The Corel-1k dataset consists of ten categories, each of which contains 100 images with a resolution of 384 × 256 × 3 pixels or 256 × 384 × 3 pixels, for a total of approximately 1,000 images shown in Table 1. Six samples of each type are shown in Fig. 4.

Figure. 5 From left to right, six samples of images from some categories in the Corel-10k dataset

4.1.2. Corel-10k dataset description

Corel-10k is the largest dataset used in this study. It has 100 categories, including art, aviation, cats, dogs, owls, tigers, lions, and etc. There are a total of 10,000 images, 100 in each category. The Corel-10k dataset's sample images are shown in Fig. 5.

4.2 Performance evolution

In this section, the performance measures used to evaluate the efficiency of the proposed method are precision and recall [30]. Precision is a metric that measures how many correct positive predictions are made, while recall is a metric that measures how many correct positive predictions are made out of all possible positive predictions. Precision and recall can be computed as:

$$Precision = \frac{No.relevant\ images\ retrieved}{Total\ No.images\ retrieved} \quad (11)$$

$$Recall = \frac{No.relevant\ images\ retrieved}{Total\ No.relevant\ images\ in\ the\ collection} \quad (12)$$

Each image in this work is treated as a query image from each of the 10 groups of images. The top L results were selected from the proposed CCBIR system, known as precision at L. The average precision values (AP) for every class at different precision levels (L = 10, 20 ….. 100) are computed and then the mean of all these average precision values is calculated at each L level, which is called mAP and is computed as follows:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (13)$$

Where i is the class or category number, N is the total number of classes and AP_i is the average

Table 2. The precision percentage score of the proposed method on the Corel-1k dataset

Categories	L					
	10	**20**	**30**	**40**	**50**	**100**
Village & People	81.6	75.7	69.4	64.4	59.6	38.2
Beach	77.1	75.2	74.6	74.3	74.2	65.2
Buildings	93.9	93.2	92.5	91.7	90.8	84.8
Buses	100	100	100	100	100	99
Dinosaurs	100	100	100	100	100	99
Elephants	99.8	99.9	99.8	99.7	99.7	99
Flowers	98.7	98.7	98.5	98.2	97.4	84.5
Horses	100	99.8	97.6	92	87.3	56.2
Mountains	91.6	89.4	88.5	87.9	87	80.8
Food	96.5	96.3	96.1	95.6	95.5	90.8
Average	93.9	92.8	91.7	90.4	89.2	79.8

Table 3. The recall percentage score of the proposed method on the Corel-1k dataset

Categories	L					
	10	**20**	**30**	**40**	**50**	**100**
Village & People	8.2	15.1	20.8	25.7	29.8	38.2
Beach	7.7	15.0	22.4	29.7	37.1	65.2
Buildings	9.4	18.6	27.8	36.7	45.4	84.8
Buses	10.0	20.0	30.0	40.0	50.0	99.0
Dinosaurs	10.0	20.0	30.0	40.0	50.0	99.0
Elephants	10.0	20.0	30.0	39.9	49.9	99.0
Flowers	9.9	19.7	29.5	39.3	48.7	84.5
Horses	10.0	20.0	29.3	36.8	43.6	56.2
Mountains	9.2	17.9	26.6	35.2	43.5	80.8
Food	9.7	19.3	28.8	38.2	47.8	90.8
Average	9.4	18.6	27.5	36.2	44.6	79.8

precision of class i.

4.2.1. The experiment on the Corel-1k dataset

On the experiment on the Corel-1k dataset, the average precision and recall percentage scores were evaluated on the number of retrieval L images of 10, 20, 30, 40, 50, and 100. Table 2 shows the average precision AP percentage scores by the proposed CCBIR for each class on the Corel-1k dataset.

Table 3 shows the recall percentage scores by the proposed CCBIR for each category on the Corel-1k dataset.

Table 4 shows the comparison between the proposed CCBIR method and other methods CST [15], [16] and [17] on the Corel-1k dataset in terms of precision and recall scores at the top 20 retrieval levels.

Table 4. Mean average precision and recall comparison of CCBIR and other methods on Corel-1k dataset at the top 20 retrieval levels

Method	Prec %	Rec %
CST Model [15]	80.8	–
Authors in [16]	81.3	16.2
Authors in [17]	81	16
The Proposed CCBIR	92.8	18.5

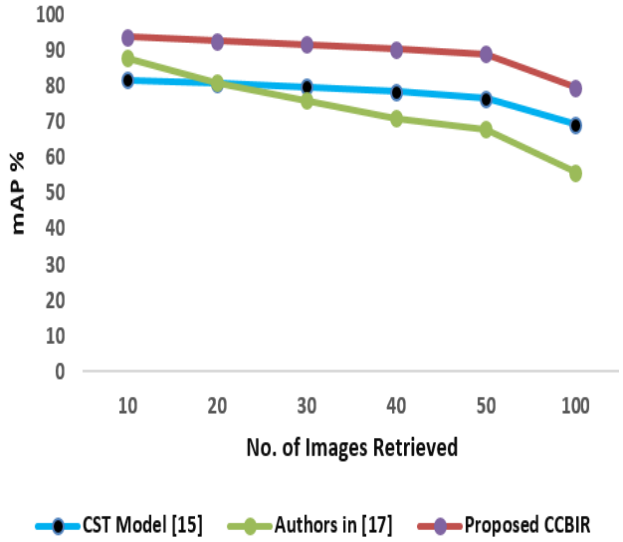

Figure. 6 Mean average precision score comparison with varying number of retrieved images for Corel-1K

Fig. 6 shows the mean average precision comparison with different numbers of images retrieved on Corel-1K dataset at 10, 20, 30, 40, 50 and 100 retrieval levels. The proposed approach shows high precision values along all retrieval levels compared to methods CST model [15] and [17] in terms of the mAP on the Corel-1k dataset.

Fig. 7 shows the average recall comparison for Corel-1K with varying numbers of retrieved images compared to method [17]. From both Fig. 6 and Fig. 7, it indicates that the proposed method is better than methods CST model [15] and [17] in terms of precision and recall scores.

Fig. 8 shows most five similar images from each category retrieved by the query image in our proposed CCBIR on Corel-1k dataset.

4.2.2. The experiment on the Corel-10k dataset

On the experiment on the Corel-10k dataset, the mean average precision and recall percentage scores were also evaluated on the number of retrieval L images of 10, 20, 30, 40, 50, and 100. Table 5 shows

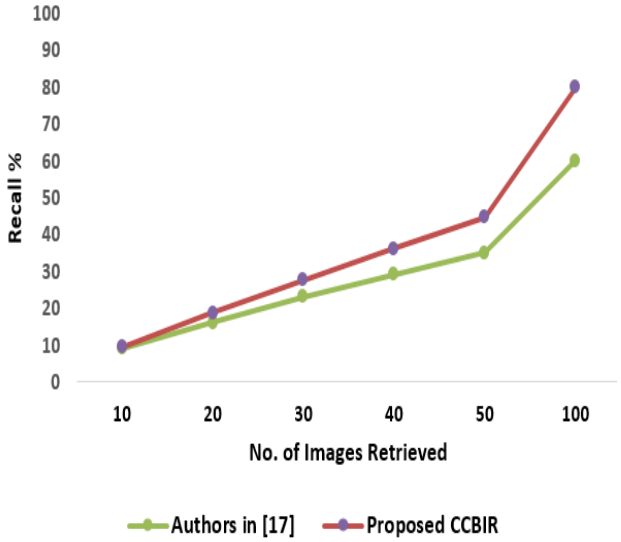

Figure. 7 Retrieval average recall comparison for Corel-1K

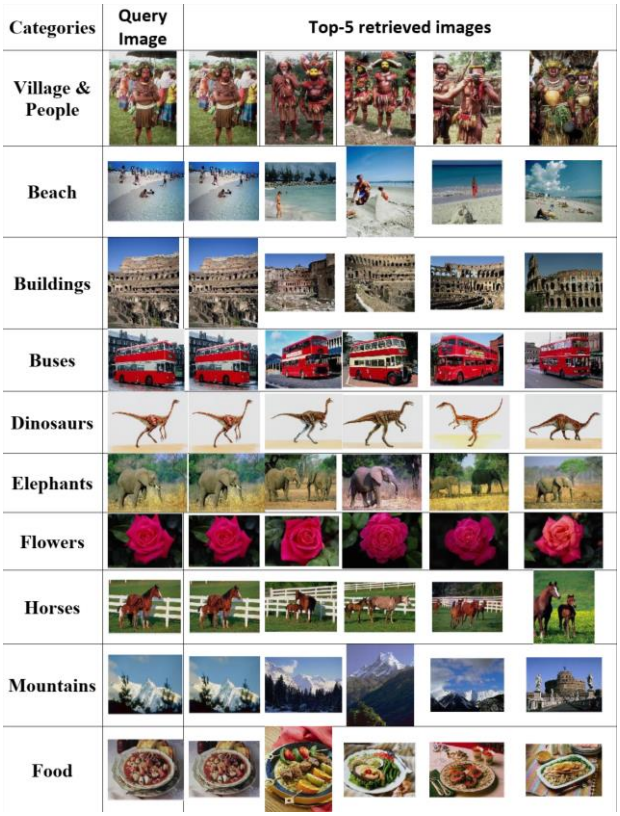

Figure. 8 The most five similar images retrieved by query image using proposed CCBIR on Corel-1k dataset

the average precision and recall percentage scores by the proposed CCBIR on the Corel-10k dataset.

Table 6 shows the comparison between the proposed CCBIR method and other methods [17-19] on the Corel-10k dataset in terms of precision and recall scores at the top 20 retrieval levels.

Table 5. The precision and recall percentage scores of the proposed method on the Corel-10k dataset

Performance	**L**					
	10	**20**	**30**	**40**	**50**	**100**
Precision	70.2	64.8	61.7	59.4	57.2	45.5
Recall	7.0	13.0	18.5	23.8	28.6	45.5

Table 6. Mean average precision and recall comparison of CCBIR and other methods on Corel-10k dataset at the top 20 retrieval levels

Method	Prec %	Rec %
Authors in [19]	46.2	9
Authors in [18]	45.8	9
Authors in [17]	62	11
The Proposed CCBIR	64.8	13

Figure 9 Mean Average Precision Score Comparison with varying number of retrieved images for Corel-10K

Fig. 9 shows the mean average precision comparison with different numbers of images retrieved on Corel-10K dataset at 10, 20, 30, 40, 50 and 100 retrieval levels. The proposed approach shows high precision values along all retrieval levels compared to method in [17] in terms of the mAP on the Corel-10k dataset.

Fig. 10 shows the average recall comparison for Corel-10K with varying numbers of retrieved images compared to method [17]. From both Fig. 9 and Fig. 10, it indicates that the proposed method is better than method in [17] in terms of precision and recall scores.

Fig. 11 shows the top 10 similar images from the categories dinosaur, cat and flower retrieved by the query image in our proposed CCBIR on the Corel-10k dataset.

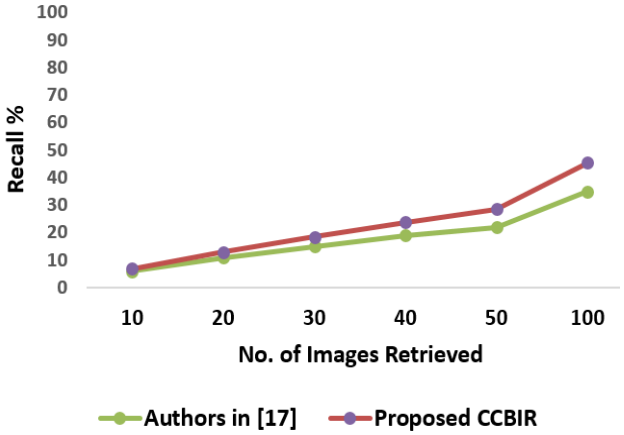

Figure. 10 Retrieval average recall comparison for Corel-10K

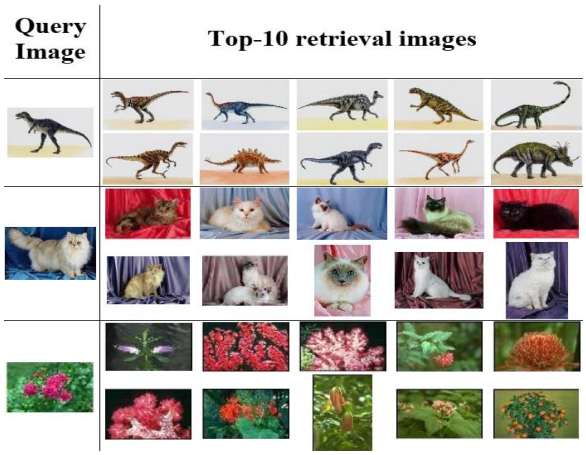

Figure. 11 The top 10 similar images from the categories dinosaur, cat and flower retrieved by query image on Corel-10k dataset

5. Conclusion and future work

In this paper, a new method is proposed to retrieve similar images by unsupervised learning using a clustering approach called CCBIR. This proposed method is based on the integration of pre-trained AlexNet CNN features combined with features extracted from PCNN to provide a robust feature representation of images. The K-means algorithm is used to group the un-labeled image dataset into different clusters using these combined feature vectors after the normalization process, whereas images that are similar to each other will be clustered together in the same cluster. In addition, the Euclidean distance measure was used as the similarity metric to retrieve images like the query image from the cluster that the query image mapped to it by calculating the closest distance to all the existing centroids. The results of conducted experiments on the Corel-1K and Corel-10K datasets showed that CCBIR achieves high precision and recall values compared to other existing systems in the unsupervised image retrieval domain.

On the Corel-1k dataset, the average precision percentage score is 92.8% at the top 20 retrieval levels, an improvement of around 11% over the precision percentage score obtained by [17], which is 81% at the top 20 retrieval levels. Moreover, the proposed method also obtained the best recall percentage score, which is 18.5%, increased by approximately 2% compared to the recall percentage score obtained by [17] which is 16% at the top 20 retrieval level. On the Corel-10k dataset, the average precision percentage score is 64.8% at the top 20 retrieval levels, an improvement of around 3% over the precision percentage score obtained by [17] which is 62% at the top 20 retrieval levels. In addition, the recall percentage score obtained by the proposed method, which is 13%, increased by approximately 2% compared to the recall percentage score obtained by [17] which is 11% at the top 20 retrieval levels. In future work, the proposed CCBIR will be improved through implementing CCBIR in parallel computation using cloud computing to speed up the training and retrieval processes from huge databases.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, and writing—original draft preparation, Ahmed. A. A. Gad-Elrab and Mahmoud Saeed; writing—review and editing, Ahmed. A. A. Gad-Elrab; visualization, Ahmed. A. A. Gad-Elrab and Mahmoud Saeed; supervision by Ahmed. A. A. Gad-Elrab, K. R. Raslan, and Khaled Fathy.

Acknowledgments

This research was supported by the Faculty of Science, Al-Azhar University, Cairo, Egypt. I thank it for providing us with the capacity to conduct and complete this research.

References

[1] M. Farag, M. Mohie, E. Din, and H. Elshenbary "Deep learning versus traditional methods for parking lots occupancy classification", *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 19, No. 2, pp. 964-973, 2020.

[2] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification", In: *Proc. of International Conf. on Computer Vision and*

Pattern Recognition (CVPR), Columbus, Ohio, USA, pp. 1701–1708, 2014.

[3] D. Triantafyllidou, P. Nousi, and A. Tefas, "Lightweight Two-Stream Convolutional Face Detection", In*: Proc. of International Conf. on 25th European Signal Processing Conference (EUSIPCO)*, Kos island, Greece, pp. 1190–1194, 2017.

[4] N. Passalis and A. Tefas, "Spatial Bag Of Features Learning For Large Scale Face Image Retrieval", *In: Proc. of International Conf. on Big Data*, Thessaloniki, Greece, pp. 8-17, 2016.

[5] D. Han, Q. Liu, and W. Fan, "A new image classification method using CNN transfer learning and web data augmentation", *Expert Systems with Applications*, Vol. 95, pp. 43-56, 2018.

[6] ImageNet. http://www.image-net.org

[7] Y. Lou, G. Fu, Z. Jiang, A. Men, and Y. Zhou, "PT-NET: Improve Object And Face Detection Via A Pre-Trained CNN Model", In: *Proc. of International Conf. on Signal and Information Processing (GlobalSIP)*, pp. 1280-1284, 2017.

[8] J. Dinesh and K. Rajesh, "Cybernetic microbial detection system using transfer learning", *Multimedia Tools and Applications*, Vol. 79, No. 7-8, pp. 5225-5242, 2020.

[9] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep Learning Earth Observation Classification Using Imagenet Pretrained Networks", *IEEE Geoscience and Remote Sensing Letters*, Vol. 13, No. 1, pp. 105-109, 2015.

[10] M. Hussain, J. Bird, and D. Faria, "A Study On CNN Transfer Learning For Image Classification", In: *Proc. of International Conf. on UK Workshop on Computational Intelligence*, Nottingham, United Kingdom, pp. 191-202, 2018.

[11] W. Dai, Z. Zhu, and F. Wu, "Image Clustering Algorithm and Its Application in Human Resources Management in Colleges", *IEEE Access*, pp. 1-8, 2020.

[12] T. Van and T. Le, "Content-based image retrieval based on binary signatures cluster graph", *Expert Systems*, Vol. 35, No. 1, p. e12220, 2018.

[13] M. Sadique, B. Biswas and S. Haque, "Unsupervised Content-Based Image Retrieval Technique Using Global And Local Features", In: *Proc. of International Conf. on Advances in Science, Engineering and Robotics Technology (ICASERT*), Dhaka, Bangladesh, pp. 1-6, 2019.

[14] S. Gupta, S. Modem, and V. Thakre, "Phase Congruency And ODBTC Based Image Retrieval", *IET Image Processing*, Vol. 14, No. 10, pp. 2195-2203, 2020.

[15] S. Zakariya and M. Jamil, "Unsupervised Content based Image Retrieval at Different Precision Level by Combining Multiple Features", In: *Proc. of International Conf. on Mechatronics and Artificial Intelligence (ICMAI)*, Gurgaon, India, p. 012059, 2021.

[16] A. Joseph, E. Rex, S. Christopher, and J. Jose, "Content-based image retrieval using hybrid k-means moth flame optimization algorithm", *Arabian Journal of Geosciences*, Vol. 14, No. 8, pp. 1-14, 2021.

[17] R. Hidayat, A. Harjoko, and A. Musdholifah, "A Robust Image Retrieval Method Using Multi-Hierarchical Agglomerative Clustering and Davis-Bouldin Index", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 2, pp. 441-453, 2022, doi: 10.22266/ijies2022.0430.40.

[18] S. Hussain, M. Zia, and W. Arshad, "Additive deep feature optimization for semantic image retrieval", *Expert Systems with Applications*, Vol. 170, No. 15 May, p. 114545, 2021.

[19] G. Xie, B. Guo, Z. Huang, Y. Zheng, and Y. Yan, "Combination of Dominant Color Descriptor and Hu Moments in Consistent Zone for Content Based Image Retrieval", *IEEE Access*, Vol. 8, pp. 146284-146299, 2020.

[20] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", *Advances in Neural Information Processing Systems*, Vol. 25, pp. 1106-1114, 2012.

[21] L. Torrey and J. Shavlik, "Handbook of research on machine learning applications and trends: algorithms, methods, and techniques", *Hershey*, PA, 2010.

[22] R. Eckhorn, H. Reitboeck, M. Arndt, and P. Dicke, "Feature Linking via Synchronization among Distributed Assemblies: Simulations of Results from Cat Visual Cortex", *Neural Computation*, Vol. 2, No. 3, pp. 293-307, 1990.

[23] R. Lippmann, "Neural networks, a comprehensive foundation", *International Journal of Neural Systems*, Vol. 5, No. 4, pp. 363-364, 1994.

[24] J. Johnson, "Pulse-coupled neural nets: translation, rotation, scale, distortion, and intensity signal invariance for images", *Applied Optics*, Vol. 33, No. 26, pp. 6239-6253, 1994.

[25] Z. Wang, Y. Ma, F. Cheng, and L. Yang, "Review of pulse-coupled neural networks", *Image and Vision Computing*, Vol. 28, No. 1, pp. 5-13, 2010.

[26] M. M. E. Din, M. E. Nahas, and H. E. Shenbary, "Hybrid Framework For Robust Multimodal Face Recognition", *International Journal of Computer Science Issues (IJCSI)*, Vol. 10, No. 2, pp. 471-476, 2013.
[27] J. Hartigan and M. Wong, "Algorithm As 136: A K-Means Clustering Algorithm", *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 28, No. 1, pp. 100–108, 1979.
[28] COREL-1K Image Database is obtained from: http://wang.ist.psu.edu/docs/related/
[29] COREL-10K Image dataset is obtained from: https://www.kaggle.com/datasets/michelwilson/corel10k/
[30] H. Müller, W. Müller, D. Squire, S. M. Maillet, and T. Pun, "Performance Evaluation In Content-Based Image Retrieval: Overview And Proposals", *Pattern recognition letters*, Vol. 22, No. 5, pp. 593-601, 2001.