Indonesian Journal of Electrical Engineering and Computer Science
Vol. 29, No. 3, March 2023, pp. 1577∼1589
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v29.i3.pp1577-1589

A deep learning content-based image retrieval approach using cloud computing

Mahmoud S. Sayed[1], Ahmed A. A. Gad-Elrab[1,2], Khaled A. Fathy[1], Kamal R. Raslan[1]
[1]Department of Mathematics, Faculty of Science, Al-Azhar University, Cairo, Egypt
[2]Faculty of Computing and Information Technology, King Abdul-Aziz University, Jeddah, Saudi Arabia

Article Info

Article history:

Received May 20, 2022
Revised Oct 15, 2022
Accepted Oct 24, 2022

Keywords:

AlexNet
Cloud computing
Content-based image retrieval
Convolutional neural networks
Discrete cosine transform
Principal components analysis
Support vector machine

ABSTRACT

Due to the rapid growth in multimedia content and its visual complexity, content-based image retrieval (CBIR) has become a very challenging task. Existing works achieve high precision values at first retrieval levels such as top 10 and top 20 images, but low precision values at subsequent levels such as top 40, 50, and 70, so the goal of this paper is to propose a new CBIR approach that achieves high precision values at all retrieval levels. The proposed method combines features extracted from the pre-trained AlexNet model and discrete cosine transform (DCT). Then principal components analysis (PCA) is performed on AlexNet's features and feeding these combination to multiclass support vector machine (SVM). The euclidean distance is used to measure the similarity between query and stored images features within the predicted class by SVM. Finally top similar images are ranked and retrieved. All above techniques require huge computational power which may not be available on client machine thus, the processing of these tasks is processed on cloud. Experimental results on the benchmark Corel-1k show that the proposed method achieves high precision value 97% along all retrieval levels top 10, 20, and 70 images and requiring less memory compared to other methods.

Corresponding Author:

Mahmoud S. Sayed
Department of Mathematics, Faculty of Science, Al-Azhar University
Nasr city, 11884, Cairo, Egypt
E-mail: mahmoud.saeed@azhar.edu.eg

1. INTRODUCTION

People are increasingly coming into contact with a large amount of image information as a result of the rapid development and popularisation of digital technology, computer and network technology, and images have become a common carrier to describe and store the information. Image retrieval (lR) is one of the most popular image processing research areas. At the moment, the majority of web based image search engines rely solely on metadata connected with images, such as keywords, tags, or descriptions, this method called text based image retrieval (TBIR) and this may result in a large number of false detections. Furthermore, manually adding keywords for images in a huge database can be wasteful and may not catch every keyword that characterises the image. As a result, the performance of these systems is unsatisfactory.

Content-based image retrieval (CBIR) has recently become essential due to its ability to overcome the existing challenges. The main purpose of CBIR is to extract key visual features of images, such as texture, colour and shape and determine the degree of similarity among images using similarity measures. As a result, the two most critical elements impacting CBIR efficiency are feature representations and similarity measurements. Several low-level feature descriptors for image representation have been proposed in the past, ranging from global features such as colour [1], texture [2], and shape [3]. Interest points detectors like histogram of oriented gradients (HOG) [4], scale-invariant feature transform (SIFT) [5] and speeded-up robust features (SURF) [6]. Image representations relying on one type of feature could result in unsatisfying CBIR performance because of insufficient representation of the images' visual contents. Jabeen *et al.* [7] proposed an image retrieval system rely on the features fusion of speeded-up robust features-fast retina keypoint (SURF-FREAK) feature descriptors on the basis of the bag-of-visual-words (BoVW) model, to overcome the semantic gap and increase image retrieval efficiency. Elnemr [8] proposed an image retrieval system that combines the SURF and maximally stable extremal regions (MSER) approaches. The SURF detector can recognise features such as blobs and corners, but it is unable to detect keypoints respect to regions. It is also noise sensitive and rotation and scale invariant, however it is not affine. MSER, on the other side, can detect features surrounding an object's region but cannot detect corner or blob features. MSER is also rotation, scaling, and affine transformation invariance.

In recent years, machine learning algorithms have been widely used and have produced good results. Deep learning is a significant subfield of machine learning. Deep learning techniques, specifically convolutional neural networks (CNN), has widely used and achieved great improvement in image processing field. A CNN is composed of several hidden layers that execute mathematical computations on the input given by the previous layer and produce an output that is fed into the next layer. Over the past recent years, CNNs have improved the performance of computer vision systems, including feature extraction [9], image classification [10]-[15], pattern recognition [16] and speech recognition [17].

There are many researchers, who have used CNNs to improve CBIR and achieving significant improvements. Shah *et al.* [18], trained a deep CNN's AlexNet framework, where the authors utilized eight trained layer network with The first five layers of the network are convolutional, and the remaining layers are fully connected. They have utilized the features extracted from the seventh trained layer to obtain similar images. However, CNN features have higher dimensionality and unskillfulness of resemblance calculation between a pair of vectors with 4,096 dimensions. Later, dimensionality mitigation was proposed in order to reduce the dimensionality of the features where in [19], proposed a combination of AlexNet CNN features, local binary pattern (LBP), and HOG features, The principal components analysis (PCA) used to reduce the dimensions of the HOG descriptor to 1x59. Then, The feature vectors of HOG-PCA and LBP are combined to create a new handcrafted feature vector with a dimension of 1x118. To match the dimension of the handcrafted feature vector with the dimension of the deep feature vector, the handcrafted feature vector is processed by PCA, and 64 of the 118 features are selected to create a handcrafted-PCA feature vector with a 1x64 dimension. Finally, a combination of the deep feature vector and the handcrafted-PCA is performed and an efficient image descriptor with a dimension of 1×128 is created.

Recently, pre-trained CNN models with transfer learning approach have the ability to produce and extract effective and descriptive features from image data and achieving high accuracy result as in [20]. Maji and Bose [21] proposed a new CBIR approach in which features are obtained from pre-trained network models from a deep learning convolution network trained for a large image classification problem. Ahmed [22] proposed CBIR systems based on features extracted using pre-trained CNN models ResNet18 and SqueezeNet. They employed these pre-trained CNN models to extract two groups of features that are stored separately and then later are used for online image searching and retrieval. Experimental results on the popular image dataset Core-1K show that ResNet18 features based on the CBIR method have overall accuracy of 95.5% on top 10 retrieval images. Jiang [23] suggested a new approach for CBIR based on image feature fusion and fisher encoding (FV). First, image blocks are used to extract low-level image content features such as hue-saturation-value (HSV) histograms, uniform LBP, and dual-tree complex wavelet transform (DTCWT). In contrast, high-level features are retrieved using the AlexNet CNN. The LBP and DTCWT were subjected to the singular value decomposition (SVD). Second, low-level features are merged using normalisation and weights. Finally, after utilising the FV encoding, the fused fisher vectors are utilised to quantify the similarity of picture pairings. The testing findings on the benchmark Corel-1k reveal that the accuracy on the top 10, 12, and 20 images returned are 93.4%, 92.8%, and 91.4%, respectively.

Keisham and Neelima [24] proposed efficient content-based picture retrieval strategies, which are discussed with machine learning (ML) algorithms. Pre-processing, multiple feature extraction, feature fusion, clustering, and classification are all processes in the proposed deep neural network-synthetic aperture radar (DNN-SAR). In the pre-processing step, a fast average peer group (FAPG) filter is utilised to reduce noise.

Then, numerous features such as colour, shape, and texture are extracted, and feature vectors are computed. Using average and weighted average approaches, all three characteristics are combined into a single feature. Following that, the fused features are grouped using the adaptive sunflower optimization (SFO) method. Finally, the appropriate photos are extracted using the DNN-SAR optimization process. mAP value of suggested (DNN-SAR) in terms of Corel-1k (93.91%) on top 10 retrieval images.

The drawback of the existing works that they achieving good precision value at first retrieval levels e.g. top 10 and top 20 retrieval images but achieving low precision at the remaining levels e.g. top 40, 50 and 70 and that drawback will be overcome in our research. The utilisation of new technology, like cloud computing, is active in its successful application. Cloud computing is defined as "transferring the process from the user's machine to servers on the internet, and storing the user's data to be accessible from any location and any machine," the software becoming services, and the user's computer becoming only an interface as in [25]. The design of CBIR through cloud computing is shown in Figure 1.

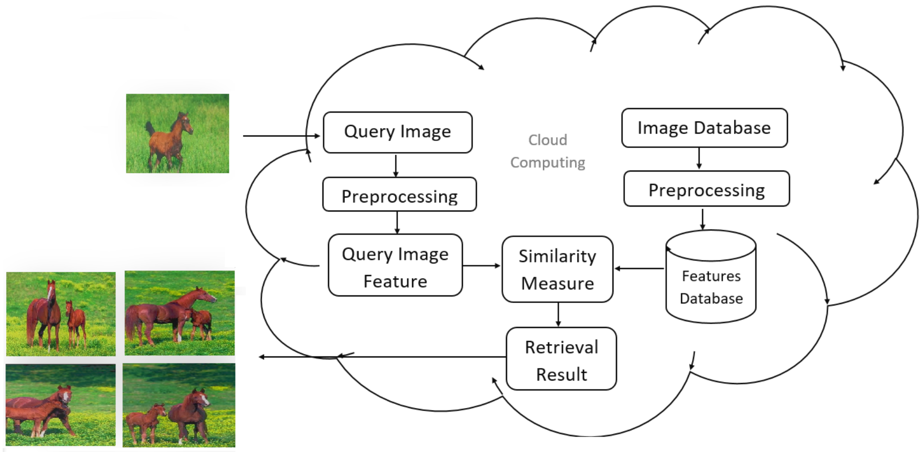

Figure 1. The design of CBIR through cloud computing

In this paper, a new CBIR method is proposed to achieve high precision value along all retrieval levels with less calculation complexity through advantages of cloud computing. The proposed approach is based on a combination of pre-trained AlexNet CNN for features extraction followed by PCA for dimensionality reduction integrated with discrete cosine transform (DCT) of entire image and feeding these combination to multiclass support vector machine (SVM) for classification and finally euclidean distance is used for similarity measure between query and stored images using the extracted features. This paper is organized as follows. Sections 2 presents the proposed image retrieval method. Section 3 presents results and discussion. Section 4 provides conclusion.

2. METHOD

In this paper a new content-based image retrieval method called deep learning content-based image retrieval using cloud computing (DLCBIR) is proposed in order to achieve better retrieval results for the CBIR system. In the rest of this section, the basic idea of DLCBIR is introduced, and then the steps of the proposed approach are described. A CBIR system typically has two phases, the offline phase and the online phase, which will be described at the end of this section.

2.1. Basic idea

The basic idea behind CBIR is to find similar images in a large database based on a query image. Typically, some useful features are extracted from query and database images, and retrieve images which have similar set of features. In our work we utilize a deep learning in order to extract these features integrated with DCT of entire image. Also in order to accelerate features similarity process, we apply dimension reduction approach on extracted features. We also use multiclass SVM in order to improve accuracy result. All above techniques requires huge amount of computing power, which may not be available with client machine, thus this processing is done on cloud.

2.2. The proposed approach

DLCBIR consists of six phases which are i) geatures extraction, ii) dimensionality reduction, iii) feature vector normalization, iv) feature vectors combination, v) multiclass classification, and vi) similarity determination. There are six phases of the proposed DLCBIR approach. These phases are described as Figure 2.

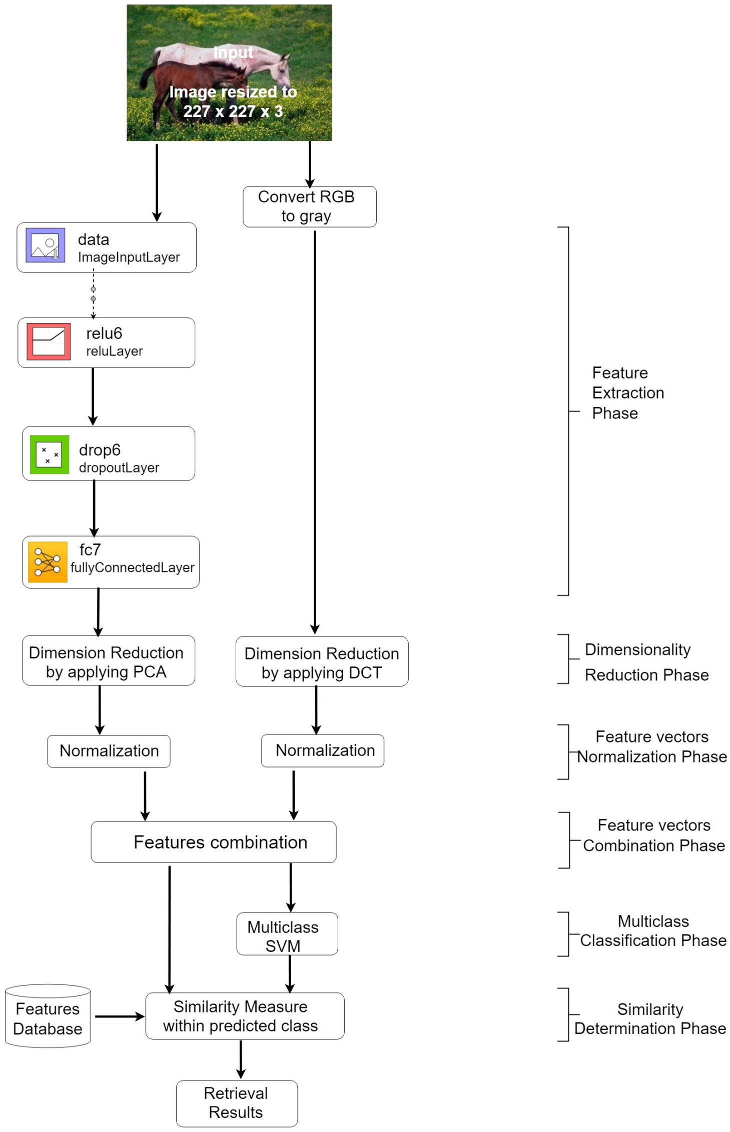

Figure 2. Six phases of the proposed DLCBIR approach

2.2.1. Features extraction phase

In this phase, the AlexNet CNN is used for extracting all features from images dataset. AlexNet CNN is a modified version of CNN. CNN is a valuable research topic in the field of machine learning and computer vision. CNN is consisted of multiple hidden layers that execute mathematical computations on the input given by the previous layer and produce an output that is fed into the next layer, as shown in Figure 3 a CNN varies from neural networks in that it has convolutional layers, which can be a good model to detect correlations between neighbouring pixels rather than fully connected layers. The training stage is typically very expensive in terms of computing and can take a long time to accomplish. The time for prediction is quite fast and efficient once the network training step is completed and the classifier has been initialised appropriately.

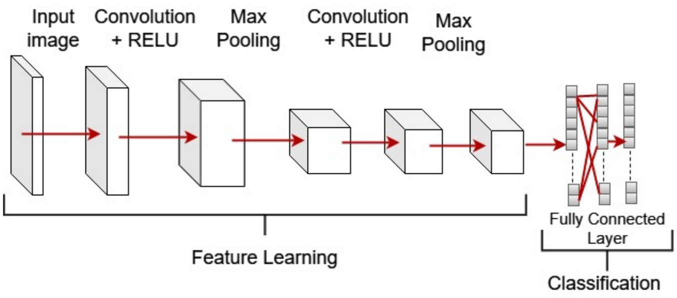

Figure 3. The architecture of CNN

AlexNet [26] is a CNN which had a significant impact on the area of machine learning, especially in terms of applying deep learning to machine vision. The AlexNet has already been trained on the ImageNet Dataset, which has over 15 million pictures and 22,000 class labels, significantly more than a normal training dataset. When working with images of popular items from the ImageNet dataset, this can indeed result in a somewhat good classifier. Thus, we use a pre-trained CNN's AlexNet for feature extraction in this work.

The AlexNet is composed of eight trained layers. The first five layers are convolutional, whereas the last three layers are fully connected. To accelerate the train, the rectified linear unit (ReLU) is applied after all convolutional and fully connected layers. Dropout is used before the first and second fully connected layers. So, in this phase, the images are read and resized to xKxz (e.g., 227×227×3). This work use the pre-trained 7th layer for feature extraction with a feature vector of length 4096 per image [18]. The CNN AlexNet process started by extracting features from the image dataset. Then stores the extracted features for further processing. Figure 4 shows an example of the AlexNet architecture.

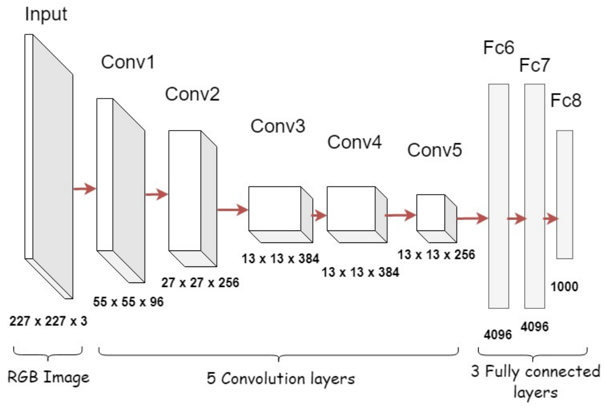

Figure 4. AlexNet architecture

2.2.2. Dimensionality reduction phase

In this phase, to accelerate image retrieval process and improving its performance, dimension reduction on the features extracted of 7th pre-trained layer (FC layer) of AlexNet CNN is applied by using PCA. PCA is a useful method in data analysis for reducing dimension and to obtain maximum variance of data. On the other hand, DCT is used for entire image features compression without losing too much performance. This process is described as follows.

DCT [27] has good energy accumulation characteristics and can still maintain performance during dimensionality reduction [28]. The 1D discrete cosine transform X(k) of a finite sequence x(n) of data with length N is defined as (1).

$$X\left(k\right) = \alpha \sum_{n=0}^{N-1} x\left(n\right) \cos\left(\frac{\pi\left(2n+1\right)k}{2N}\right) \tag{1}$$

where

$$\alpha(k) \begin{cases} \sqrt{\frac{1}{N}} & k = 0 \\ \sqrt{\frac{2}{N}} & k \neq 0 \end{cases}$$

The two-dimensional transform is equivalent to a one-dimensional DCT performed along a single dimension followed by a one-dimensional DCT in the other dimension. One of the main characteristic of DCT is its ability to convert the energy of the image into a few coefficients [29] by cluster high value coefficients in the upper left corner and low value right of the image. Thus, applying DCT on the image and taking the first K significant coefficients extracted in a zigzag order started from the upper left corner from the transformed image can be used as feature vector that represent the image with a few coefficients without losing too much performance. The number K of coefficients to keep is determined experimentally. The higher number of taken coefficients makes high quality of the representation.

PCA is a versatile technique and has been widely used and achieving a good result in various applications such as dimensionality reduction, data compression, and feature extraction [30]. The advantages of using PCA method are reduce the dimensionality of a data set by finding a new set of variables smaller than the original set of variables, retains most of the sample's information and help in classification of data. Principal components can be identified by calculating the eigenvectors and eigenvalues of the data covariance matrix. Following give details about PCA method. Suppose we have matrix A which contains the term weights obtained by feature extraction techniques:

$$A(k) = \begin{bmatrix} X_{11} & X_{12} & .. & X_{1k} & .. & X_{1m} \\ X_{21} & X_{22} & .. & X_{2k} & .. & X_{2m} \\ . & . & .. & .. & .. & . \\ . & . & .. & .. & .. & . \\ X_{n1} & X_{n2} & .. & X_{nk} & .. & X_{nm} \end{bmatrix} \tag{2}$$

where x_{jk} (j=1,2,...,n; k=1,2,...,m) is the terms weight that exists in the collection of vectors. Where n is the number of images to be classified and m is the number of term weights obtained from feature extraction. The used steps by PCA to reduce the dimensionality of matrix A are described as follows:
step 1: calculate the mean of m variables in matrix A:

$$\bar{X}_k = \frac{1}{n} \sum_{j=1}^{n} x_{jk} \tag{3}$$

step 2: calculate the covariance S_{ik} of m variables in matrix A:

$$S_{ik} = \frac{1}{n} \sum_{j=1}^{n} (x_{ji} - \bar{X}_i)(x_{jk} - \bar{X}_k) \tag{4}$$

where i = 1, ... ,m. Eigenvectors and eigenvalues of the covariance matrix are computed, 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 (5).

$$M = \begin{bmatrix} f_{11} & f_{12} & f_{13} & \cdots & f_{1d} \\ f_{21} & f_{22} & f_{23} & \cdots & f_{2d} \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ f_{n1} & f_{n2} & f_{n3} & .. & f_{nd} \end{bmatrix} \tag{5}$$

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 feature vector size of each image that extracted of 7^{th} pre-trained layer (FC layer) of AlexNet CNN from 1× 4096 to 1× M (e.g., 1 x 64) and obtain maximum variance of data.

2.2.3. Feature vectors normalization phase

Normalization gives equal weight to all features in a data set and thus be useful for classification algorithms. Normalization can improve classification model prediction performance as in [31]. The normalization process is done by considering the values in the vector. For example, if the vector is of size 1×4: [4,6,9,11]. To normalize it we need to calculate the l2-norm for this vector, 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 $[0.25, 0.37, 0.69, 0.56]$.

2.2.4. Feature vectors combination phase

In this phase, a combination of a normalized feature vector produced by DCT and a feature vector produced by PCA is done and created a finally features vector that represent the images. For example, the DCT feature vector with dimension 1 x M (e.g., 1x10) and PCA feature vector with dimension 1 x N (e.g., 1x64) are combined and an efficient image descriptor with a dimension of 1x(M+N) (e.g., 1x74) is created.

2.2.5. Multiclass classification phase

In this phase, multiclass SVM is used for classification to increase the accuracy of the proposed approach along all retrieval levels. While categorizing a particular image, there are N different classes to which the given image can be placed. Therefore, it is required to construct a function which can effectively predict the class to which the given image belongs. SVMs are primarily designed for binary classification that is for only two classes possibility. For more than two classes, there is no SVM equivalent to multinomial regression. Rather, the outputs of individual two-class SVMs are combined. There are several ways to accomplish this. Our implementation applies the “one-against-one” approach as shown in Figure 5, the support vector classification procedure (for a k number of classes) is executed k (k -1)/2 times for each possible pair of these classes. For each pair, the winning class is the one with the highest points among all two-class SVMs [32].

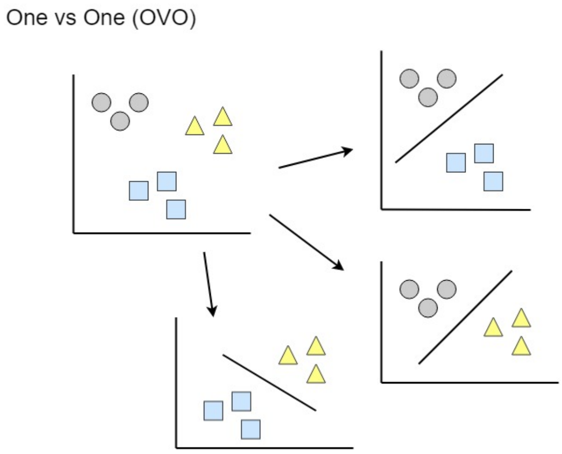

Figure 5. An example illustrating one-against-one” approach a multiclass SVM

2.2.6. Similarity determination phase

In this phase, the euclidean distance is used for similarity measure between query and stored images. Euclidean distance is the most appropriate measure for determining similarity due to its popularity and simplicity of computation. Query image feature vector is compared with dataset feature vectors within the class that predicted from multiclass SVM using euclidean distance. A set of relevant images is selected then they arranged in descending order of their euclidean distance score to retrieve top N.

$$dist(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{6}$$

In (6) X and Y feature vector of query image and feature vector of image in the database while x and y are the element in these vectors.

2.3. The online and offline processes proposed approach

The proposed CBIR framework as shown in Figure 6 includes two types of process modes. Online process (on the left side of Figure 6) and offline process (on the right side of Figure 6). During the offline phase, a feature database is created for each image in the database and the multi-class SVM is trained on these features. The online process mode, on the other side, is based on the user interface, where the features are extracted from the query image given by users. From this, the distance measure is then used to compare the provided image feature to the features database within the predicted class using the trained SVM. These distance measurements are sorted to rank the images based on their similarities and then retrieved top ranked images. Figure 6, shows an overview of the online and offline processes of the proposed DLCBIR system.

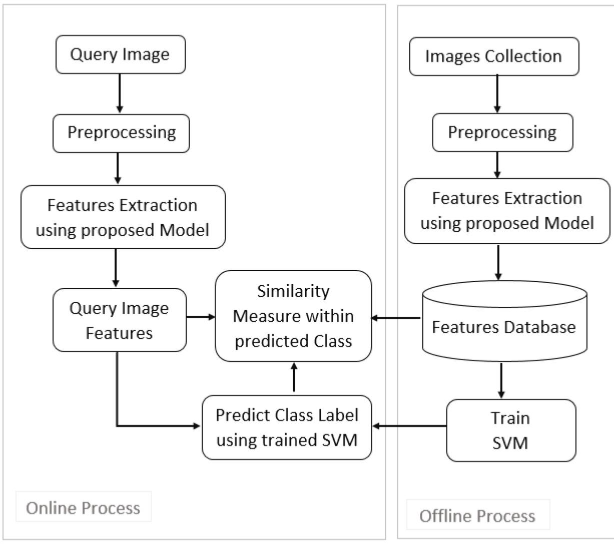

Figure 6. Overview of the online and offline processes of the proposed DLCBIR approach

3. RESULTS AND DISCUSSION

3.1. Dataset description

The commonly dataset used in most image retrieval and classification research is Corel-1k dataset [33]. Therefore, the performance of the proposed DLCBIR is examined using Corel-1k dataset. The Corel-1k dataset composed of 10 categories, each one contain 100 images with a resolution of 256×384×3 or 384×256×3 pixels. We used 70% of images per class for training and 30% for evolution. Six samples of each type are shown in Figure 7.

Figure 7. Images samples from each category in the Corel-1k dataset from left to right

3.2. Performance evolution

In this section, the performance of the proposed DLCBIR is evaluated and measured against the existing systems in [7], [8], [18]-[24]. The proposed DLCBIR retrieves a set of relevant images from the dataset based on their euclidean distance score. The performance of all methods is measured using precision [34]. Precision is a metric that quantifies the number of correct positive predictions made. Precision can computed as (7):

$$Precision = \frac{No.\,relevant\,images\,retrieved}{Total\,No.\,images\,retrieved} \tag{7}$$

For a given query q, the corresponding average precision AP is calculated, and then the mean of all these APs scores is calculated which is called mAP and is computed as (8).

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AveragePrecision \tag{8}$$

Table 1 shows the average precision AP of the proposed DLCBIR for each class by using Corel-1k dataset. The proposed system achieved high average precision AP on each category by using Corel-1k dataset. Figure 8 show the confusion matrix for the proposed DLCBIR on corel-1k dataset and Figure 9 shows most five similar images from each category retrieved by the query image in our proposed DLCBIR on Corel-1k dataset.

Table 1. Category-wise average precision results of DLCBIR at the top 10 retrieval images

Categories	AP%
Beaches	93.33
Bus	100
Dinosaurs	100
Elephants	100
Flowers	100
Foods	90
Horses	100
Monuments	100
Mountain and snow	90
People and villages in Africa	96.67

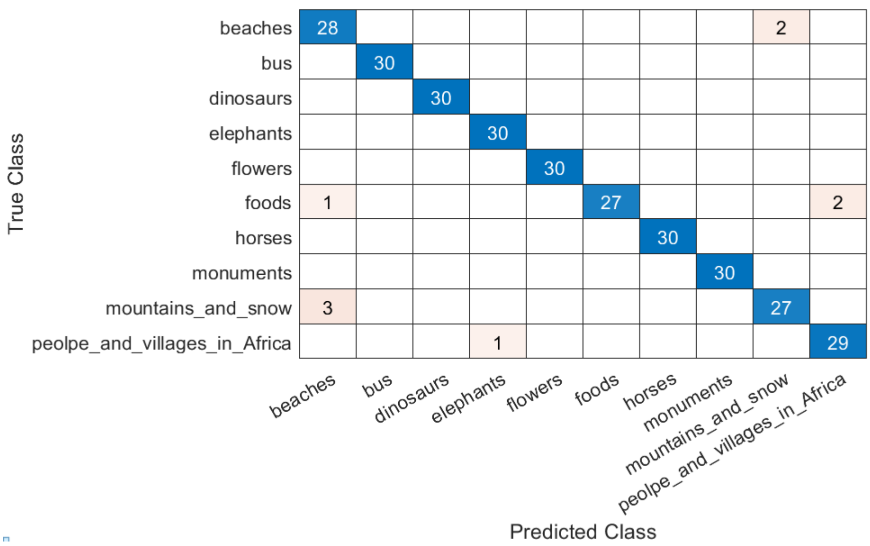

Figure 8. Confusion matrix for the proposed DLCBIR on corel-1k dataset

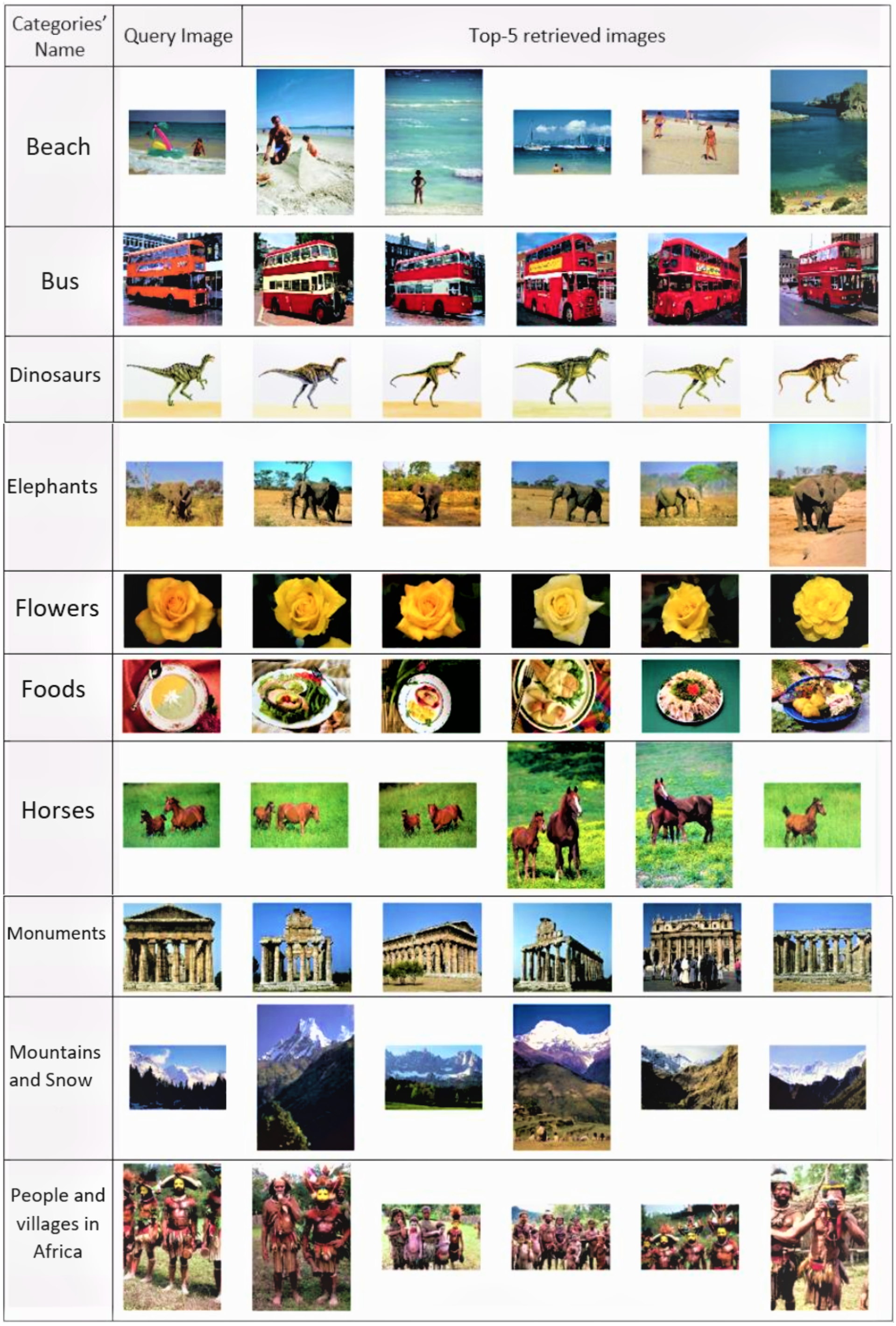

Figure 9. The visual results of query image using proposed DLCBIR on Corel-1k dataset

Figure 10 show comparison results of our proposed method with other method for Corel-1k dataset in term of precision for top 10 retrieved images. Figure 11 show precision graph with varying number of retrieved images for Corel-1k. Numbers of retrieved images are 10, 20, ..., 70. The proposed method is showing high precision value along all levels among all compared methods. Table 2 shows the comparison of proposed DLCBIR with the existing methods in terms of the mAP by using Corel-1k dataset. The achieved results proof that the proposed DLCBIR can achieve higher precision value and requiring less memory compared to existing retrieval systems [7], [8], [18]-[24].

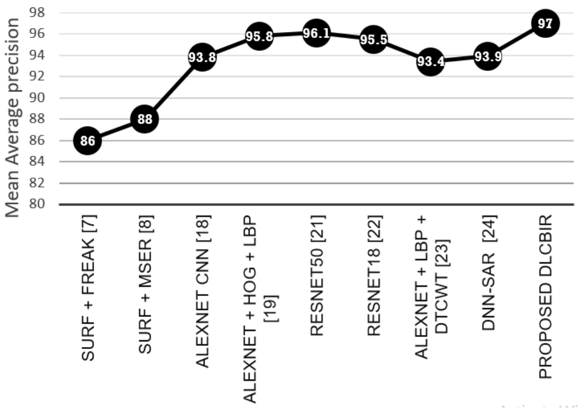

Figure 10. Comparison results of our proposed method with other method for Corel-1k dataset in term of precision for top 10 retrieved images

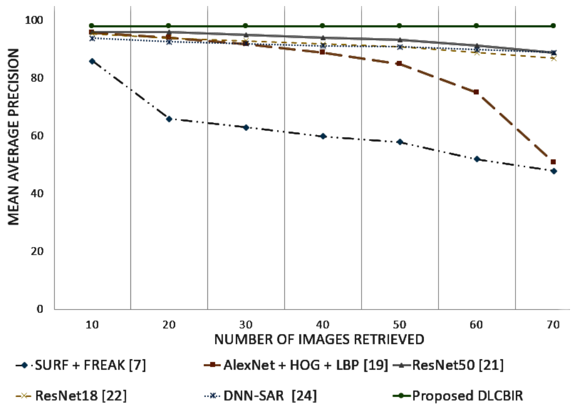

Figure 11. Precision Graph with varying number of retrieved images for Corel-1k

Table 2. The mAP results of DLCBIR and other methods on Corel-1k dataset at the top 10 retrieval images

Method	mAP %	Dimension
SURF + FREAK [7]	86	1 × 128
SURF + MSER [8]	88	1 × 128
AlexNet CNN [18]	93.80	1 × 4096
AlexNet + HOG + LBP [19]	95.80	1 × 128
ResNet50 [21]	96.11	1 x 100
ResNet18 [22]	95.50	1 x 512
AlexNet + LBP + DTCWT [23]	93.4	–
DNN-SAR [24]	93.91	–
Proposed DLCBIR	**97.00**	**1 × 74**

4. CONCLUSION

In this paper, a new algorithm to retrieve similar images through advantages of cloud computing called DLCBIR is proposed. DLCBIR is based on the pre-trained AlexNet CNN features followed by PCA method integrated with features extracted from DCT of entire image and feeding these combination after normalization process to Multiclass SVM method. The combination of features extracted from DCT and the features extracted from AlexNet-PCA was used because it will give a good precision compared to use one of them separately. The multiclass SVM used to increase the performance where the similarity measure between query and stored images occurred within the class which is predicted by it. In addition, the euclidean distance measure was used as the similarity metric to retrieve images that is most like the query image from the database. The results of conducted experiments on the Corel-1k dataset showed that DLCBIR achieves high precision value at different precision level which was 97% compared to other existing systems, for the correctly classified and retrieved images in the test data. In future work, the proposed DLCBIR will be improved through implementing DLCBIR in parallel computation to retrieve images from large databases while decreasing the time necessary for training and extracting features from the databases.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] G. Lu, "Image retrieval based on color," *Storage and Retrieval for Still Image and Video Databases IV*, vol. 2670, pp. 310-320, 1996, doi: 10.1117/12.234811.

[2] N. Chaturvedi, S. Agarwal, and P. K. Johari, "A novel approach of image retrieval based on texture," *International Journal of Enhanced Research in Management and Computer Applications*, vol. 3, no. 1, pp. 42-48, 2014.

[3] G. Lu, "An approach to image retrieval based on shape," *Journal of Information Science*, vol. 23, no. 2, pp. 119–127, 1997, doi: 10.1177/016555159702300203.

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004, doi: 10.1023/B:VISI.0000029664.99615.94.

[6] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust features," in *Computer Vision – ECCV 2006. ECCV 2006. Lecture Notes in Computer Science*, Berlin: Springer, 2006, pp. 404–417, doi:10.1007/11744023_32.

[7] S. Jabeen, Z. Mehmood, T. Mahmood, T. Saba, A. Rehman, and M. T. Mahmood, "An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model," *PLOS ONE*, vol. 13, no. 4, p. e0194526, 2018, doi: 10.1371/journal.pone.0194526.

[8] H. A. Elnemr, "Combining SURF and MSER along with color features for image retrieval system based on bag of visual words," *Journal of Computer Science*, vol. 12, no. 4, pp. 213–222, 2016, doi: 10.3844/jcssp.2016.213.222.

[9] P. S. Nair, T. A. Berihu, and V. Kumar, "An image-based gangrene disease classification," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 6, pp. 6001–6007, 2020, doi: 10.11591/ijece.v10i6.pp6001-6007.

[10] M. S. Farag, M. M. M. El Din, and H. A. Elshenbary, "Deep learning versus traditional methods for parking lots occupancy classification," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 19, no. 2, 964–973, 2020, doi: 10.11591/ijeecs.v19.i2.pp964-973.

[11] K. Kusrini, M. R. A. Yudianto, and H. Al Fatta, "The effect of Gaussian filter and data preprocessing on the classification of Punakawan puppet images with the convolutional neural network algorithm," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 3752–3761, 2022, doi: 10.11591/ijece.v12i4.pp3752-3761.

[12] W. Suwarningsih *et al*., "Ide-cabe: chili varieties identification and classification system based leaf," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 1, pp. 445–453, 2022, doi: 10.11591/eei.v11i1.3276.

[13] A. Julianto and A. Sunyoto, "A performance evaluation of convolutional neural network architecture for classification of rice leaf disease," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 4, pp. 1069–1078, 2021, doi: 10.11591/ijai.v10.i4.pp1069-1078.

[14] M. A. Rasyidi, R. Handayani, and F. Aziz, "Identification of batik making method from images using convolutional neural network with limited amount of data," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 3, pp. 1300–1307, 2021, doi: 10.11591/eei.v10i3.3035.

[15] S. Z. M. Zaki, M. A. Zulkifley, M. M. Stofa, N. A. M. Kamari, and N. A. Mohamed, "Classification of tomato leaf diseases using MobileNet v2," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 2, pp. 290–296, 2020, doi: 10.11591/ijai.v9.i2.pp290-296.

[16] M. A. Rasyidi and T. Bariyah, "Batik pattern recognition using convolutional neural network," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1430–1437, 2020, doi: 10.11591/eei.v9i4.2385.

[17] J. O. Pinzón-Arenas and R. Jiménez-Moreno, "Comparison between handwritten word and speech record in real-time using CNN architectures," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 4, pp. 4313–4321, 2020, doi: 10.11591/ijece.v10i4.pp4313-4321.

[18] A. Shah, R. Naseem, Sadia, S. Iqbal, and M. A. Shah, "Improving CBIR accuracy using convolutional neural network for feature extraction," *2017 13th International Conference on Emerging Technologies (ICET)*, 2017, pp. 1–5, doi: 10.1109/ICET.2017.8281730.

[19] A. Shakarami and H. Tarrah, "An efficient image descriptor for image classification and CBIR," *Optik*, vol. 214,2020, doi: 10.1016/j.ijleo.2020.164833.

[20] H. A. Razak, M. A. M. Saleh, and N. M. Tahir, "Review on anomalous gait behavior detection using machine learning algorithms," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 5, pp. 2090–2096, 2020, doi: 10.11591/eei.v9i5.2255.

[21] S. Maji and S. Bose, "CBIR using features derived by deep learning," *ACM/IMS Transactions on Data Science*, vol. 2, no. 3, pp. 1–24, 2021, doi: 10.1145/3470568.

[22] A. Ahmed, "Pre-trained CNNs models for content based image retrieval," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021, doi: 10.14569/IJACSA.2021.0120723.

[23] D. Jiang, "Image feature fusion and fisher coding based method for CBIR," *2021 International Conference on Communications, Information System and Computer Engineering (CISCE)*, 2021, pp. 503–508, doi: 10.1109/CISCE52179.2021.9445921.

[24] N. Keisham and A. Neelima, "Efficient content-based image retrieval using deep search and rescue algorithm," *Soft Computing*, vol. 26, no. 4, pp. 1597–1616, 2022, doi: 10.1007/s00500-021-06660-x.

[25] K. V. Vartak and V. A. Bharadi, "Cloud based cbir software as a service," *International Journal of Applied Information Systems (IJAIS)*, vol. 2, pp. 1-6, 2012.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1-9, 2012.

[27] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. C–23, no. 1, pp. 90–93, Jan. 1974, doi: 10.1109/T-C.1974.223784.

[28] V. Naidu, "Discrete cosine transform-based image fusion," *Defence Science Journal*, vol. 60, no. 1, pp. 48–54, 2010, doi: 10.14429/dsj.60.105.

[29] J. H. AlKhateeb, J. Ren, J. Jiang, S. S. Ipson, and H. El Abed, "Word-based handwritten Arabic scripts recognition using DCT features and neural network classifier," in *2008 5th International Multi-Conference on Systems, Signals and Devices*, 2008, pp. 1–5, doi: 10.1109/SSD.2008.4632863.

[30] M. M. El-Din, M. El Nahas, and H. El Shenbary, "Hybrid framework for robust multimodal face recognition," *International Journal of Computer Science Issues (IJCSI)*, vol. 10, no.2, pp. 471-476, 2013.

[31] A. B. A. Graf and S. Borer, "Normalization in support vector machines," in *Pattern Recognition. DAGM 2001. Lecture Notes in Computer Science*, Berlin: Springer, 2001, pp. 277–282, doi: 10.1007/3-540-45404-7_37.

[32] N. Guenther and M. Schonlau, "Support vector machines," *The Stata Journal: Promoting communications on statistics and Stata*, vol. 16, no. 4, pp. 917–937, 2016, doi: 10.1177/1536867X1601600407.
[33] "Corel1k", [Online]. Available: https://drive.google.com/drive/folders/19aCYwRF7DW5NBhtU1sz1vcVzhhJkF09y
[34] H. Müller, W. Müller, D. M. Squire, S. Marchand-Maillet, and T. Pun, "Performance evaluation in content-based image retrieval: overview and proposals," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 593–601, Apr. 2001, doi: 10.1016/S0167-8655(00)00118-5.

BIOGRAPHIES OF AUTHORS

Mahmoud S. Sayed received the B.Sc., pre-M.Sc. and M.Sc. degrees in Mathematics and Computer Science from the faculty of Science, Al-Azhar University, Egypt, in 2012, 2013 and 2017 respectively. His current position is a teaching assistant of mathematics and computer science, Faculty of Science, Al-Azhar University, Cairo, Egypt. 2014-till now. His research interests include image processing and machine learning. He can be contacted at email: mahmoud.saeed@azhar.edu.eg.

Ahmed A. A. Gad-Elrab is Associate Professor of ubiquitous and mobile computing at the Department of Mathematics, Faculty of Science, Al-Azhar University, Cairo, Egypt. He received B.S. Degree in Computer Science, from Faculty of Science, Alexandria University, Egypt in 1999. He received MS. Degree in Computer Science from Faculty of Science, Cairo University, Egypt in 2008. He received his Ph.D. at the Nara Institute of Science and Technology (NAIST), a national corporation university located in NARA, Japan in 2012. He received the NAIST best PhD student award in March 2012, he received the outperformance Award from Graduate School of Information Science, NAIST in March 2012, and he received 2011 IPSJ Yamashita Memorial Award (given to only one or two papers among all papers presented in one year in each IPSJ SIG, Japan. His research interests include cloud computing, mobile computing, Internet of Things applications, smart home, data science, sensor networks, dynamic distributed systems, big data, and mobile crowd sensing. Currently, he is an associate professor at Faculty of Computing and Information Technology, King Abdul-Aziz University, Jeddah, Saudi Arabia. In addition, he is a Consultant in Vice Presidency of Development, King Abdul-Aziz University, Jeddah, Saudi Arabia. He can be contacted at email: asaadgad@azhar.edu.eg, aaahmad4@kau.edu.sa.

Khaled A. Fathy was born in 1981. He received the B.Sc., pre-M.Sc., M.Sc., and Ph.D. degrees in Mathematics and Computer Science from the faculty of Science, Al-Azhar University, Egypt, in 2004, 2007, 2012, and 2018 respectively. His current position is a lecturer of mathematics and computer science. His research interests include algorithms (sequential and parallel), cryptography and image processing. He can be contacted at email: khaledfathy66@yahoo.com.

Kamal R. Raslan received the M.Sc. and Ph.D. degrees from the Faculty of Science, Menoufa University and Al-Azhar University, Egypt, in 1996 and 1999, respectively. He is currently a full Professor of Mathematics with the Faculty of Science, Al-Azhar University, Egypt. He has authored/coauthored over 114 scientifc papers in top-ranked International Journals and Conference Proceedings. His research interests include Numerical Analysis, Finite Difference Methods, Finite Element Methods, Approximation Theory, and Computational Mathematics. He can be contacted at email: kamal_raslan@yahoo.com.