Embedding neural networks for semantic association in content based image retrieval

**Aun Irtaza · M. Arfan Jaffar ·
Eisa Aleisa · Tae-Sun Choi**

Abstract Content based image retrieval (CBIR) systems provide potential solution of retrieving semantically similar images from large image repositories against any query image. The research community are competing for more effective ways of content based image retrieval, so they can be used in serving time critical applications in scientific and industrial domains. In this paper a Neural Network based architecture for content based image retrieval is presented. To enhance the capabilities of proposed work, an efficient feature extraction method is presented which is based on the concept of in-depth texture analysis. For this wavelet packets and Eigen values of Gabor filters are used for image representation purposes. To ensure semantically correct image retrieval, a partial supervised learning scheme is introduced which is based on K-nearest neighbors of a query image, and ensures the retrieval of images in a robust way. To elaborate the effectiveness of the presented work, the proposed method is compared with several existing CBIR systems, and it is proved that the proposed method has performed better then all of the comparative systems.

Keywords CBIR · Partial supervised learning ·
Neural network based semantic association

A. Irtaza
National University of Computer & Emerging Sciences, Islamabad, Pakistan
e-mail: aun.irtaza@gmail.com

M. A. Jaffar (✉)
Al-Imam Muhammad Ibn Saud Islamic University, Riyadh, Saudi Arabia
e-mail: arfan.jaffar@ccis.imamu.edu.sa

E. Aleisa
Imam Muhammad Ibne Saud University, Riyadh, Saudi Arabia
e-mail: aleisa@ccis.imamu.edu.sa

T.-S. Choi
Gwangju Institute of Science and Technology, Gwangju, Korea
e-mail: tschoi@gist.ac.kr

1 Introduction

Modern image search engines which retrieve images on the base of actual visual contents are referred to as Content Based Image Retrieval (CBIR) systems. CBIR systems have found vast applications in many fields like architectural design, surveillance systems, geographical information systems, data mining, remote sensing, fabric design, medical image retrieval, Internet image search, video search, and communication systems [30]. Efforts are carried out for determining the ways through which retrieval of semantically correct images in response of any query image can be assured. But it is a challenging task, as understanding an image and interpreting semantics is a difficult thing for both humans and computers. For example as shown in Fig. 1, in response of query image all three response images are relevant images, depending on user's search intention. If we become able to develop an efficient and robust CBIR system we can overcome the linguistic problems as well which appears while describing and sharing images, as suffered by all keyword based image retrieval systems.

Main working theme of any CBIR system is to calculate the feature vector of repository images and characterize some visual properties against them. These feature vectors (also known as image signatures) serves as feature database and are used for measuring similarity amongst the images in the image repository. User provides a query image, and the CBIR system computes feature vector for it, and then compare it with feature database using any similarity or dissimilarity metric like Manhattan distance. Most similar images are returned to the user as system response. But this approach has two main flaws due to which CBIR systems are unable to produce good results: (1) they rank the retrieved images on the base of distance or similarity with query image and generate the results, in this regard they do not verify their output. The problem with this approach is that many images may appear as the response images while they are not relevant at all, e.g. as shown in Fig. 2: due to the lack of verification, CBIR systems may bring dog's image in response of girl's image as both images are similar in features. (2) Secondly they do not consider the neighbors for the finalization of obtained results [6].

Keeping the above mentioned points in mind; in this paper we have focused on finding the ways through which such kind of shortcomings can be avoided and

Fig. 1 Semantic gap: In above query image some users focus on cars so the best match could be the car images, some users focus on girls so the best match could be the girl images, and similarly some user focus on rock and road images so the best match could be the rock and road images

Fig. 2 Visual features of girl's image and the dog image are very similar, but both images belongs from different semantic class

performance of CBIRs can be enhanced. A backpropagation neural network based structure is proposed which is trained on sub repository of images generated from the main image repository and utilizes the right neighborhood of the query image. The aim of this training is to insure the retrieval of semantically correct images in response of query images. To further enhance the performance of proposed system, a powerful feature extraction and neighbor selection way is introduced which is based on indepth texture analysis of images. For this hybrid texture features are used which are the fused version of features obtained from wavelet packets and Eigen values of Gabor filter. For neighbor selection Pearson correlation is used.

Rest of paper is organized as follows. Section 2 provides the details of related work done in the area of content based image retrieval. Section 3 is focusing on the proposed method introduced in this paper. Experimentation and results are covered in Section 4. Finally we have concluded in Section 5.

2 Related work

In the commercial domain, QBIC system by IBM [9] is considered as the first CBIR system. After QBIC system many additional systems were developed at IBM, NEC AMORA, VIRAGE, Interpix and Bell Laboratory [25]. Common ground of all proposed CBIR systems is their intention to address image searching problem in a more effective way by addressing new feature types and image similarity detection measures. For this the focus of research is on texture [1, 14, 17, 33], color [32, 33], or shape [34] features or any of these combinations [17]. As far as the scope of features is concerned, features can be divided into two main categories: global features and local features. Global feature category includes color histogram [8, 10], texture histogram [12], color layout [19] of the whole image, and features selected from multidimensional discriminant analysis of a collection of images [4]. While local feature category includes color, texture [16], and shape features for sub images, segmented regions [6, 25], or interest points. Color histogram tells the global distribution of colors in the images. It is insensitive to small variations in the image structure. But, it has two major drawbacks. Firstly, they are unable to fully accommodate the spatial

information, and also they are not unique. Images with similar color distribution and different object composition produce very similar histograms. Moreover, similar images of same point of view carrying different lighting conditions create dissimilar histograms. Many researchers suggested the use of color correlogram for avoiding inconsistencies involving the spatial information [30].

In Babu Rao et al. [2], as a first step image is divided into eight coarse partitions. After the coarse partitioning, the centroid of each partition is selected as its dominant color. Texture of an image is obtained by using GLCM (Gray Level Co-occurrence Matrix). Color and texture features of an image are normalized, and after that shape information is captured through edge images obtained by Gradient Vector Flow fields. The combination of the shape color and texture features of an image is provided as a feature set for image retrieval. Weighted Euclidean distance of color, texture and shape features of gray level edge images of RGB space is used in retrieving the similar images. But this method suffers from boundary delocalization and is not able to capture the contours effectively.

In Wang et al. [25], introduced a semantics classification method which uses a wavelet based approach for feature extraction and then for the image comparison they used image segmentation based region matching approach. IRM proposed in this work is not efficient for texture classification due to uncertain modeling. So to address this issue, their idea was further processed by Chen et al. [6]. They introduced an un-supervised clustering based technique, which generate multiple clusters of retrieved results and give more accurate results as compare to the previous work, but their method suffers from issues like numbers of clusters identification, and segmentation uncertainty, due to which the results of this technique are not reliable.

In Lama et al. [12], proposed a content based image retrieval system for computed tomography nodule images. Their system generates feature vector through GLCM, Gabor filters, and Markov random fields. On the extracted features they perform Euclidean, Manhattan and Chebychev distances to find the relevant images from image database. In the same way Bunte et al. [5] has used CBIR in dermatology. They have used two different methods to learn favorable feature representations Limited Rank Matrix Learning Vector Quantization (LiRaMLVQ) and Large Margin Nearest Neighbor (LMNN). Both methods use labeled training data and provide discriminant linear transformation of the original features. But these techniques do not report the comparision with other techniques. Therefore it is a difficult thing to judge their effectiveness.

An important focus of research in content based image retrieval is on relevance feedback [4]. The main theme of relevance feedback is to keep the user in a loop through feedbacks to improve the performance of CBIR. But it exhibits some limitations like over sensitivity, and inability to accumulate knowledge thats why these systems are still not able to give the robust solutions. Historically RF systems use machine learning techniques like EM, KNN to bring semantically similar results in response of any query image [22]. New relevance feedback learning methods have recently been proposed among which SVM [18, 23, 28], RBFs, and Bayesian inference are most popular ones [21]. These methods consider the retrieval process as classification problem, in which relevant and irrelevant images are considered as two separate training sets; these approaches usually fail to produce good results in case of imbalanced feedback samples.

3 Proposed method

Proposed CBIR system considers categorical data for image retrieval purposes. As a first step for the implementation of process, CBIR system generates feature repository. For this it analyzes the images on the base of best nodes of wavelet packets tree; and then it uses smallest approximation image of wavelet packets tree to perform the detail Gabor analysis, and selects a set of Eigen values to completely represent them. Then by fusing these two types of feature vectors it generates the corresponding feature vector for image representation purposes. After feature extraction 'n' bags of images (BOI) [31] are generated by placing $R \geq 2$ example images in every bag to represent all semantic classes. On these BOIs a backpropagation neural network structure is defined for every class which is trained on these example images present in the corresponding BOI. So there are 'n' neural network structures as per the number of bags. These trained neural networks are used for classifying the query image into the corresponding semantic class by considering its top 'K' nearest neighbors obtained through Pearson correlation. Semantic association process is applied on all images present in the image repository and their semantic class is determined. As output of the system, images having same semantic class are returned to the user after sorting against a query image. To further enhance the capabilities of proposed system and to avoid the risk of mis association, system also returns 'M' top neighbors(also generated through Pearson correlation) to the user and enable relevance feedback upon them. Architecture of the proposed method is presented in Fig. 3.

3.1 Bags of images

There are $| D |$ images in image database belonging from $(A = n)$ categories. CBIR system divides the imagebase into $L = \{l_1, l_2, \ldots, l_n\}$ subsets as per the number of categories present in the imagebase. These subsets are known as bags of images(BOI). Each BOI contains $| D |_L = | D | / L * 0.3$ images from each category. For every BOI, system considers the images (feature vectors) present in image bag as positive examples $| d |^+$ and all other images present in other image bags as negatives examples $| d |^-$. This is represented as:

$$| d |^+ = \{i = 1, 2, \ldots, | D_{l_i} | \mid I_i \in | D_{l_i} | \} \tag{1}$$

and

$$| d |^- = \{i = 1, 2, \ldots, | D_{L-l_i} | \mid I_i \in | D_{L-l_i} | \} \tag{2}$$

On these sets system defines category specific neural networks.

3.2 Signature development and similarity calculation

For texture feature extraction and signature development, images are analyzed through wavelet packets and Gabor filters. The detail of the process is as follow.
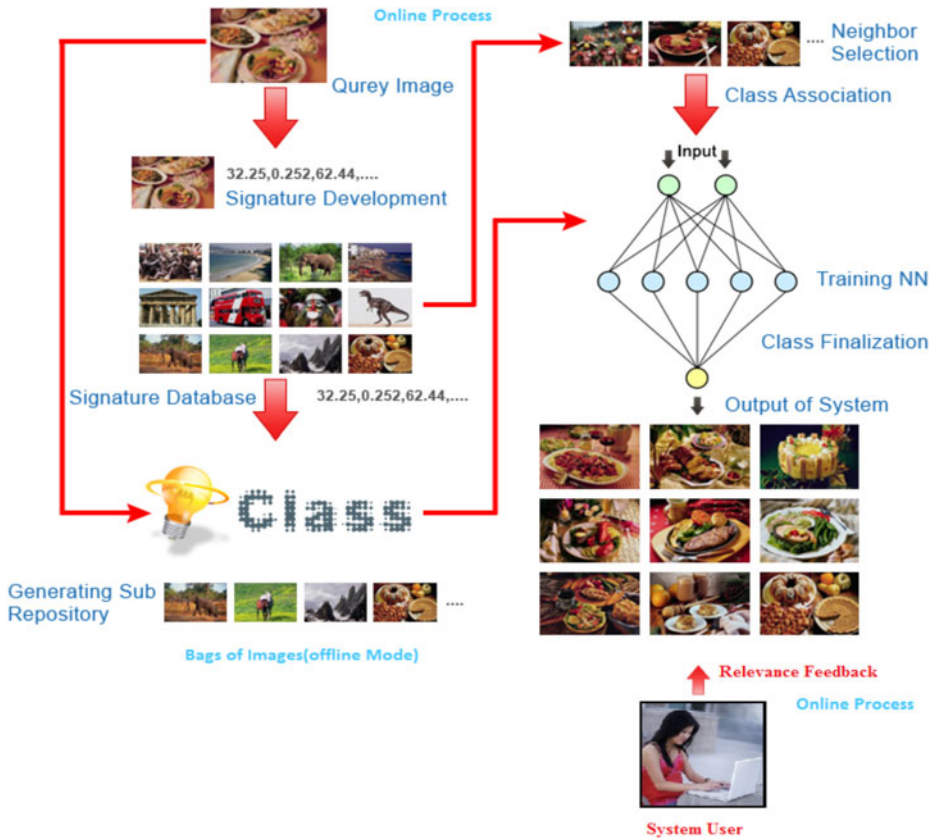
Fig. 3 Architecture of the proposed method

3.2.1 Wavelet packets

To analyze the signals in a rich way the wavelet packet method is used which is a generalization of wavelet decomposition and can be described by the collection of functions $WJ(x)|J \in Z+\}$ obtained from [14]:

$$2^{\frac{p-1}{2}} W_{2n}(2^{p-1}x - l) = \sum_m h_m - 2l^2 \frac{p}{2} W_n/(2^{p_x} - m) \quad (3)$$

$$2^{\frac{p-1}{2}} W_{2n+1}(2^{p-1}x - l) = \sum_m g_m - 2l^2 \frac{p}{2} W_n/(2^{p_x} - m) \quad (4)$$

where 'p' is a scale index and 'l' is the translation index. $W_0(x) = \Phi(x)$ is the scaling function. $W_1(x) = \Psi(x)$ is the basic wavelet function. h_k and g_k are the quadratic mirror filters. Wavelet packets are well localized in both time and frequency and thus provide an attractive alternative to pure frequency (Fourier) analysis. For a given orthogonal wavelet function, a library of bases is obtained called wavelet packet bases. Each of these bases offers a particular way of coding signals, reconstructing

exact features and preserving global energy. The inverse relationship between wavelet packets of different scales can be shown through [14]:

$$2^{\frac{p}{2}} W_{2n+1}(2^{p-1}x - l) = \sum_{l} h_k - 2l^2 \frac{p-1}{2} W_{2n}(2^{p-1}x - l)$$

$$+ \sum_{l} g_k - 2l^2 \frac{p-1}{2} W_{2n+1}(2^{p-1}x - l) \qquad (5)$$

Equation (5) can be used to calculate the wavelet packets. Coefficients of coarser scale can be calculated using (3) and (4)

$$S_{2n,l}^{p-1} = \sum_{m} hm - 2l S_{n,m}^{p} \qquad (6)$$

$$S_{2n+1,l}^{p-1} = \sum_{m} gm - 2l S_{n,m}^{p} \qquad (7)$$

Main difference between normal wavelet decomposition and wavelet packets decomposition is that despite of just splitting the approximation component of an image, wavelet packets decomposes the detail components of image as well. Therefore gives better representation of transient events which occurs at different scales. For feature extraction, wavelet packets of the repository images are calculated up to the 3rd level. This will result in the form of 64 nodes of wavelet packets tree. But we are concerned with only those nodes which have the best entropy values. So for this purpose we have generated the best entropy valued tree as corresponding best tree. Figure 4 is an example input image; below images are showing the corresponding full wavelet packets tree nodes, and entropy based best tree generated from the full

Fig. 4 Top image is the input Image and bottom images are the representation of corresponding full wavelet packets tree decomposition and nodes obtained from best tree of the wavelet packets tree

(a) Wavelet Packets Decomposition

(b) Best tree nodes
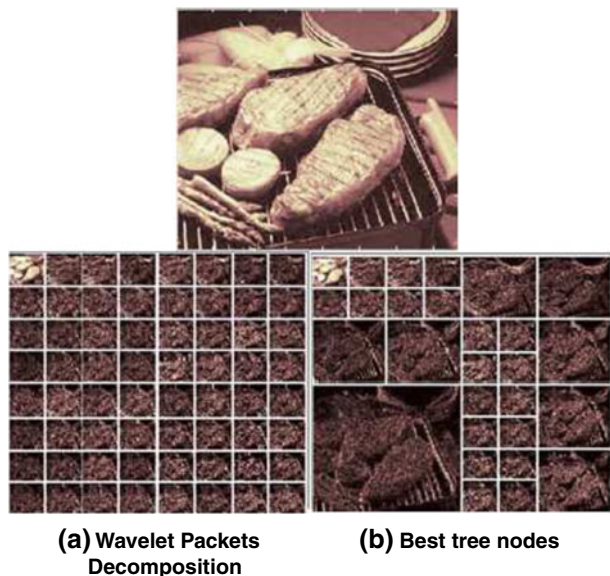
wavelet packets tree. Nodes of the wavelet packets tree are used for texture feature extraction using the following formula:

$$f_r = \sqrt{\frac{\sum c_{ij}^2}{i * j}} \qquad (8)$$

Where f_r is the computed Wavelet packet features of the sub image, C_{ij} is the representation of the intensity value of elements of sub image, and i*j is the size of the sub image.

3.2.2 Gabor features

Now as a second step of feature extraction and signature development, lowest approximation image of wavelet packet decomposition is used for Gabor analysis. In our implementation we have used only the odd components of Gabor filter to avoid the imaginary values. Odd component of Gabor filter is defined by the following equation [12]:

$$G_o(x, y) = \exp\left(\frac{x_\theta^2 - \gamma^2 y_\theta^2}{\alpha^2}\right) \sin\left(\frac{2\pi x_\theta{}^i}{\lambda}\right) \qquad (9)$$

where

$$x_\theta = x \cos_\theta + y \sin_\theta \qquad (10)$$

$$y_\theta = -x \sin_\theta + y \sin_\theta \qquad (11)$$

and σ is the standard deviation of the Gaussian function, λ is the wavelength of the harmonic function, θ is the orientation, and γ is the spatial aspect ratio which is left constant at 0.5. The spatial frequency bandwidth is the ratio σ/λ and is held constant and equal to .56. Thus there are two parameters which changes when forming a Gabor filter θ and λ.

For obtaining the Gabor response, images are divided into 9×9 non-overlapping regions and are convolved by Gabor filter with above mentioned parameters. After generating the response images following scheme is used for feature extraction:

1. Obtain twelve Gabor based response images after applying parameters.
2. Obtain one Eigen vector corresponding to every Gabor response image.
3. Calculate mean of every Eigen vector.
4. Merge mean values in one vector for feature vector generation.

3.2.3 Hybrid features

Application of above mentioned procedure will return two feature vectors, representing texture features obtained from wavelet packets and Gabor filters respectively. Aggregation of these feature vectors in a single vector will represent hybrid texture features against any image (Fig. 5).

Fig. 5 Sample images of each
category of Corel dataset

3.3 Similarity calculation

Pearson correlation [11] in proposed method is used for output generation and
neighbor selection purposes (for semantic association). The reason for our choice
is that Pearson correlation has the ability to describe linear relationship between two
entities. It is widely used in clustering to measure the degree of association between
records, but it is not common to use it in image classification and retrieval purposes.
The similarity results are in the range of -1 to $+1$. Correlation of -1 suggests
that two records (feature vectors) are entirely opposite. While 0 as the correlation
value suggests that there is no correlation between comparative records. While $+1$
suggests that comparative records are highly correlated. Pearson's correlation can be
calculated using following formula [11]:

$$r = \sqrt{\frac{\sum xy - \frac{\sum x \sum y}{n}}{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

where 'x' is feature vector of query image and 'y' is the feature vector of repository
images. Corresponding values in both feature vectors are used for correlation
calculation according to the formula.

3.4 Semantic association using backpropagation neural networks

For every BOI in image sub-repository, one feed-forward backpropagation neural
network structure [26] is defined with one hidden layer and one output unit. Every
network's hidden layer contains 20 neurons. Detail of the neural network structure
is summarized in Table 1. Sigmoid function is used in hidden layer and output layer
as transfer function i.e.

$$f(x) = g(x) = \frac{1}{1 + \exp(-x/x_0)} \tag{12}$$

These networks are trained with the concept of one against all classes (OAA)
classification. All feature vectors present in positive training set of a specific category
or BOI are labeled with '1'; and all other feature vectors present in other BOIs are

Table 1 Summary of neural network structure for every image category used in this work

Input

| Input: | $\vec{a} = (a_1, \ldots, a_n)$ | $\dim(\vec{a}) = N$ | (I.1) |

Middle (hidden) layer

Input:	$\vec{b} = U\vec{a}$	$\dim(\vec{b}) = M$	(I.2)
Output:	$\vec{c} = f(b - \vec{s})$	$\dim(\vec{c}) = M$	(I.3)
U:	M × N weight matrix		
f:	Hidden layer activation function		
$\vec{s}$:	thresholds		

Output layer

Input:	$\vec{d} = W\vec{c}$	$\dim(\vec{d}) = K$	(I.4)
Output:	$\vec{e} = g(d - \vec{t})$	$\dim(\vec{e}) = 1$	(I.5)
W:	1 × M weight matrix		
g:	Output layer activation function		
$\vec{t}$:	Thresholds		

Error correction

MSE:	$E = 1/2(\vec{p} - \vec{e})^2$	(I.6)
ΔW_{ij}	$= -\alpha \partial E/\partial W_{ij} = \alpha \delta_i c_j$	(I.6)
Δt_i	$= \alpha \delta_i$	(I.7)
ΔU_{ji}	$= -\beta \partial E/\partial U_{ji}$	(I.8)

labeled with '0'. In this way training sets are defined for all categories and neural networks are trained upon them.

After training, all images present in image database are tested against neural networks, and their association with all classes is measured using following function:

$$m^* = \bar{y} fm \tag{13}$$

where $m = \{1, 2, \ldots, n\}$ are the total number of neural networks structures, $\bar{y} fm$ returns the association factor of corresponding neural network represented by m^*. Class specific neural network performances for training is measured on mean squared error and are plotted in Fig. 8.
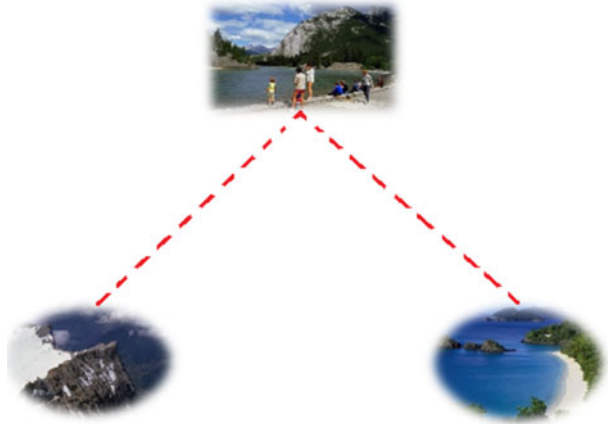
3.4.1 Class finalization through K-nearest neighbors

As shown in Fig. 6, Object composition of many images suggests multiple semantic classes for them. So the chances exists that CBIR system may associate them with undesired semantic class. Therefore for the finalization of semantic class similarity amongst the target images is also considered. For this semantic association process is applied on top K = 5 neighbors of the input image as well, and according to the majority voting rule (MVR), semantic class of the input image is finalized. MVR is represented as [22]:

$$C^*(x) = \text{sgn}\left\{ \sum_i C_i(X) - \frac{K-1}{2} \right\} \tag{14}$$

where $C_i(X)$ is the class wise association factor of input image and its top neighbors. MVR counts the largest number of classifiers that agree with each other [22]. So

Fig. 6 Composition of objects suggest that input image can be associated with mountains and beach classes

according to (14) the class of input image is one which has a maximum combined association value for itself and its nearest neighbors, as represented in (15).

$$C^*(x) = \text{argmax} \left(\sum_i m^* \right) \tag{15}$$

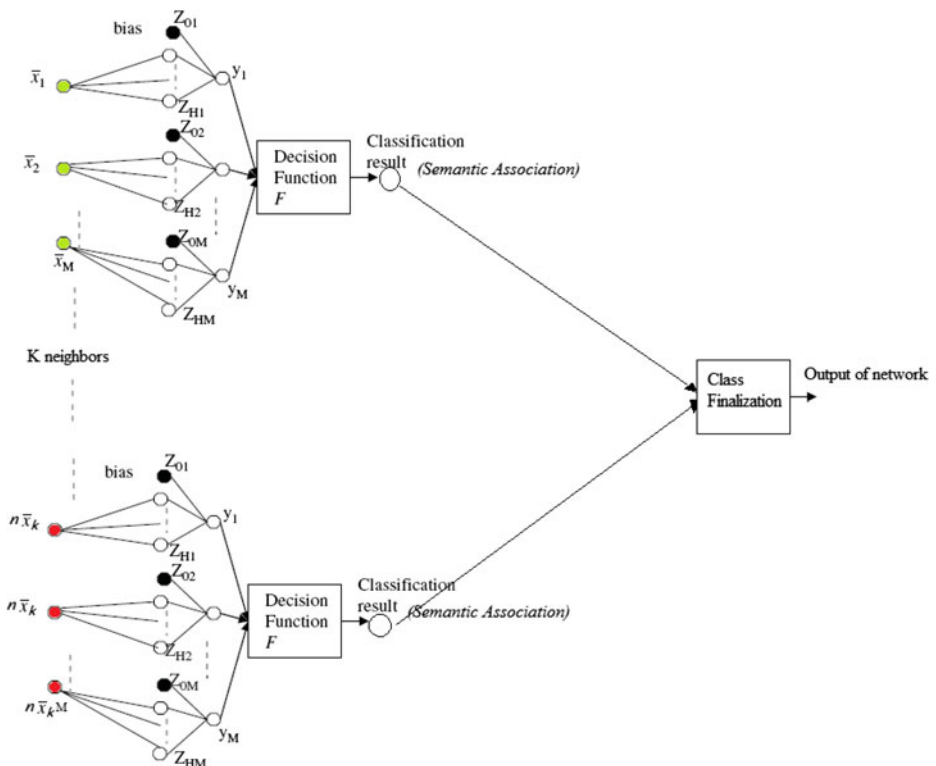Process of class finalization is graphically represented in Fig. 7.

Fig. 7 Process of semantic association and class finalization

3.5 Content based image retrieval

All images present in image database are associated with their semantic class through above mentioned process. Therefore, when system suggests semantic class for a query image, only images of same semantic class are returned to the user after sorting through Pearson correlation. For images, associated with wrong semantic class, system also returns query 'M' top neighbors to the user also obtained through the Pearson correlation (in our implementation M = 20), and enable relevance feedback upon them. Relevance feedback gives freedom to the user to guide image retrieval system in case of complex queries to achieve desired output. Both sets of images are returned to the user in the form of representative images. These representative images are selected from both sets, and are the images which appear most similar to the query image in terms of distance. User has the option to view both sets of images.

4 Experiment and results

This section provides the details of experiments performed to test our technique. The section is composed of following subsections. Section 4.1 gives the details of the image database we have used for experimentation. Section 4.2 aims to illustrate the performance of the system through query examples. Section 4.3 elaborate the effectiveness of proposed system through precision and recall. Section 4.4 is about relevance feedback.

4.1 Database description

For designing an image retrieval system, selection of a suitable image database is a critical and important step. At the present time, there is not a standard image database for this purpose [13]; and also there is no agreement on number of images and type of images in the image database [13]. As CBIR systems are if not developed for domain specific applications, then it must include various semantic groups in the image database. This is the reason that we have used COREL image dataset for this purpose. Corel dataset is publicly available at http://wang.ist.psu.edu/docs/related/, and is widely used in literature for image retrieval purposes [6, 13]. Corel dataset covers a wide range of semantic groups from natural scenes to artificial objects for CBIR experiments. Dataset has 1000 images, and is partitioned into ten semantic categories namely African people and village, beach, buildings, buses, dinosaurs, elephants, flowers, horses, mountains, and food (Fig. 5). Each category has 100 images, and size of each image is either 384×256 or 256×384. Partitioning of the dataset into semantic categories is determined by the creators and reflects the human perception of image similarity [13].

To further elaborate the performance of the proposed method, experiments are also carried out on Columbia object image library (COIL) [30]. Dataset has 7200 images from 100 different categories.

4.2 Query examples for practibility of system

As a first step for the implementation of proposed method, grayscale versions of repository images are generated. This pre-processing is required to perform the feature extraction in a cost effective way. Therefore as a backend process feature extraction is performed on grayscale versions of images and to display the results we display the color versions of the retrieved images. Each image in the image dataset is tested against the trained neural networks and its semantic class is determined (Fig. 8). Results of semantic association is stored in a file which serves as the association database.

For performance evaluation of proposed system, one image from all image categories is randomly selected; and then retrieved results are displayed against them. Query image results are displayed in Fig. 9. The response of the system can be observed from the number of correctly retrieved images in response of query images.

Therefore from query results it can be easily observed that performance of system is very good.

4.3 Retrieval precision and recall evaluation

To evaluate the effectiveness of the proposed method, we determined how many relevant images are retrieved in response of a query image. For this retrieval effectiveness is defined in terms of precision and recall rates. Precision is used to measure specificity of image retrieval, and recall is used for measuring sensitivity or true positive rate. Experiments results are reported after running five times on ten query images randomly selected from each image category. For each query image, relevant images are considered to be those images only, which belong to the same category as that of the query image. Based on this concept, the retrieval precision and recall are defined as:

$$Precision = \frac{No.\ of\ relevant\ images\ retrieved}{Total\ number\ of\ images\ retrieved}$$

$$Recall = \frac{No.\ of\ relevant\ images\ retrieved}{Total\ number\ of\ relevant\ images}$$

Top 20 retrieved images are used to compute the precision and recall rates.

4.3.1 Comparison results on Corel dataset

In order to show the superiority of proposed technique, it is compared with CT-DCIRS [3], CLUE [6], CTCHIRS [15], Babu Rao et al. [2], and EMSVM [29]. The results of comparative techniques are obtained from the original research work reported by the corresponding authors. Table 1 describes the class wise comparison of the proposed system with comparative systems in terms of average precision values. Same results are graphically illustrated in Fig. 10. Similarly Table 2 presents comparison of the proposed system with comparative systems in terms of average recall values (Table 3). Recall results are graphically illustrated in Fig. 11. From results it can be observed that, there is no system which is giving highest results in all
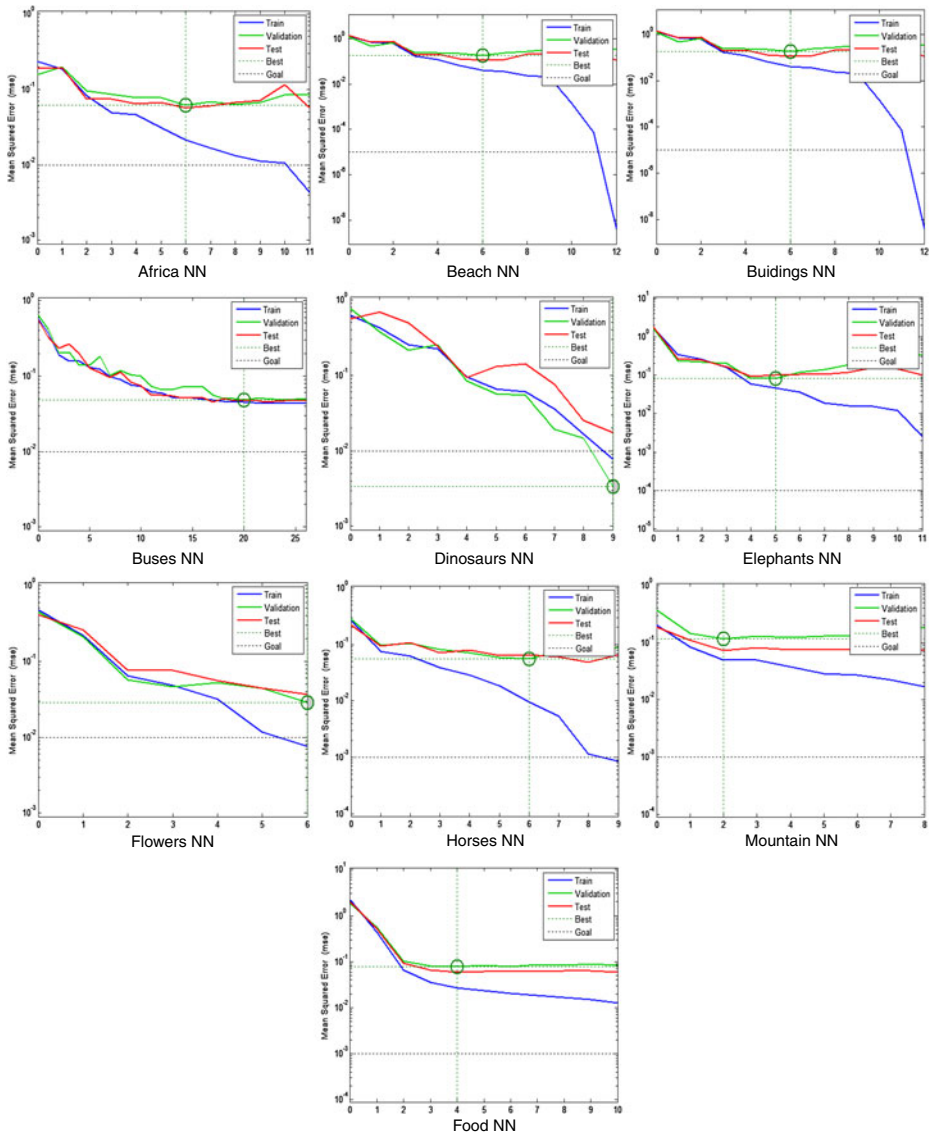
Fig. 8 Class specific neural networks training performance graph. Plotted mean squared error (MSE)

categories. But in terms of overall accuracy our proposed system has outperformed all other systems.

4.3.2 Comparison results on coil dataset

We compared the proposed method with ICTEDCT [30] on Coil dataset. Five images are selected from each image category and then performance of both systems are
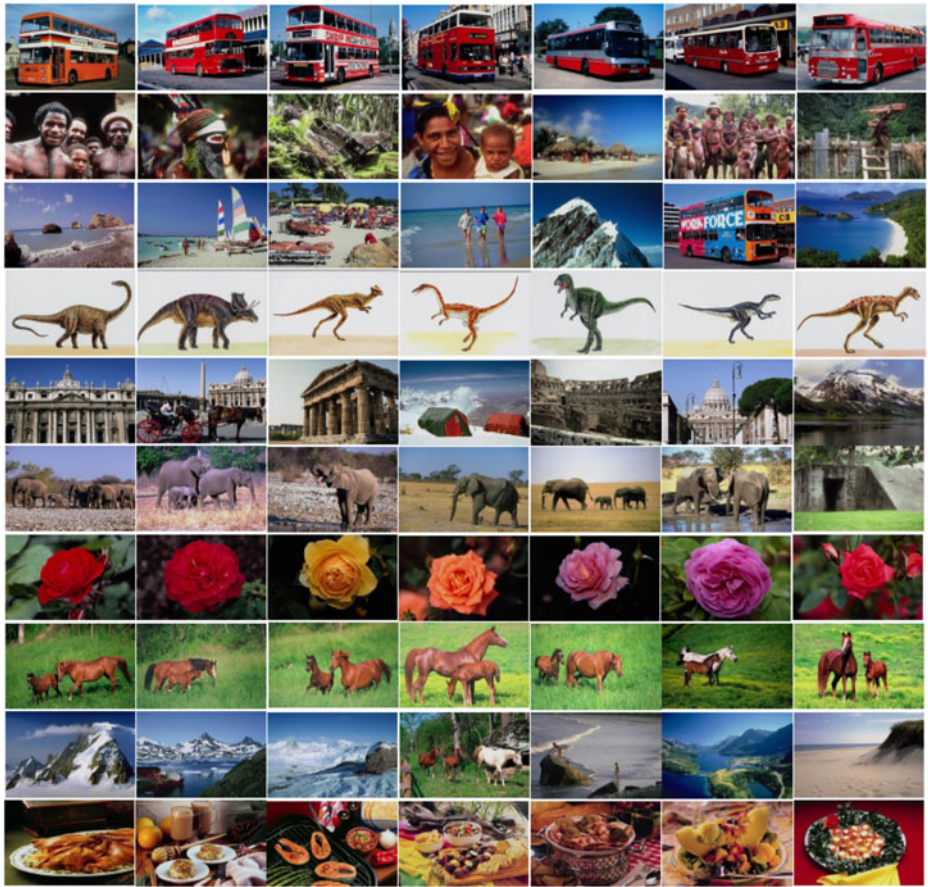
Fig. 9 Image retrieval results for all semantic classes. 1st image in every group is the query image and all other images in that group are retrieved images
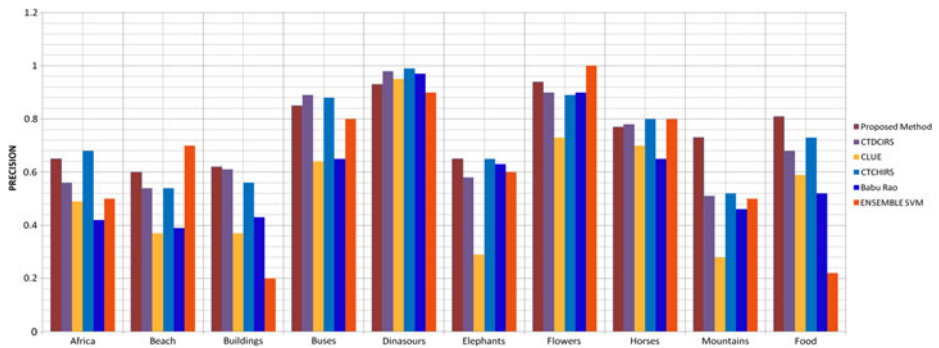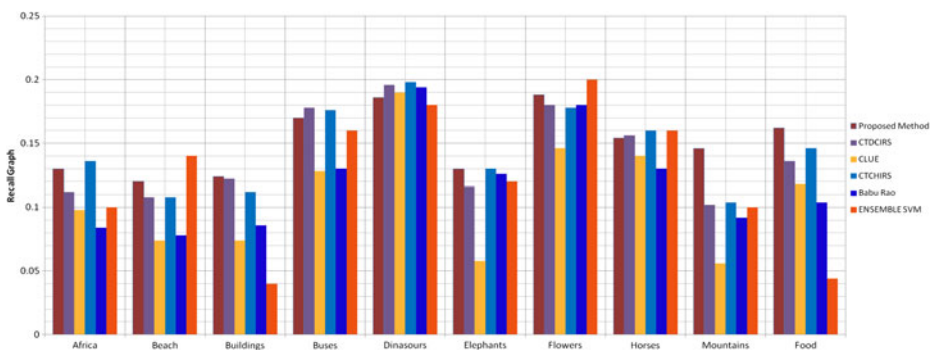

Fig. 10 Comparison of average precision obtained by proposed method with other standard retrieval systems [2, 3, 6, 15, 29]

Table 2 Comparison of average precision obtained by proposed method with other standard retrieval systems [2, 3, 6, 15, 29]

Class	Proposed method	CTDCIRS [3]	CLUE [6]	CTCHIRS [15]	Babu Rao et al. [2]	EMSVM [29]
Africa	0.65	0.56	0.49	0.68	0.42	0.50
Beach	0.60	0.54	0.37	0.54	0.39	0.70
Buildings	0.62	0.61	0.37	0.56	0.43	0.20
Buses	0.85	0.89	0.64	0.88	0.65	0.80
Dinasours	0.93	0.98	0.95	0.99	0.97	0.90
Elephants	0.65	0.58	0.29	0.65	0.63	0.60
Flowers	0.94	0.90	0.73	0.89	0.90	1.00
Horses	0.77	0.78	0.70	0.80	0.65	0.80
Mountains	0.73	0.51	0.28	0.52	0.46	0.50
Food	0.81	0.68	0.59	0.73	0.52	0.22
Average	0.75	0.70	0.54	0.72	0.60	0.62

Table 3 Comparison of average recall values obtained by proposed method with other standard retrieval systems [2, 3, 6, 15, 29]

Class	Proposed method	CTDCIRS [3]	CLUE [6]	CTCHIRS [15]	Babu Rao et al. [2]	EMSVM [29]
Africa	0.13	0.11	0.10	0.14	0.08	0.10
Beach	0.12	0.11	0.07	0.11	0.08	0.14
Buildings	0.12	0.12	0.07	0.11	0.08	0.04
Buses	0.17	0.17	0.64	0.13	0.13	0.16
Dinasours	0.19	0.20	0.19	0.20	0.19	0.18
Elephants	0.13	0.12	0.06	0.13	0.13	0.12
Flowers	0.19	0.18	0.15	0.18	0.18	0.2
Horses	0.15	0.16	0.14	0.16	0.13	0.16
Mountains	0.10	0.51	0.06	0.10	0.09	0.10
Food	0.16	0.14	0.12	0.15	0.10	0.04
Average	0.15	0.14	0.11	0.14	0.12	0.12

Fig. 11 Comparison of average recall obtained by proposed method with other standard retrieval systems [2, 3, 6, 15, 29]
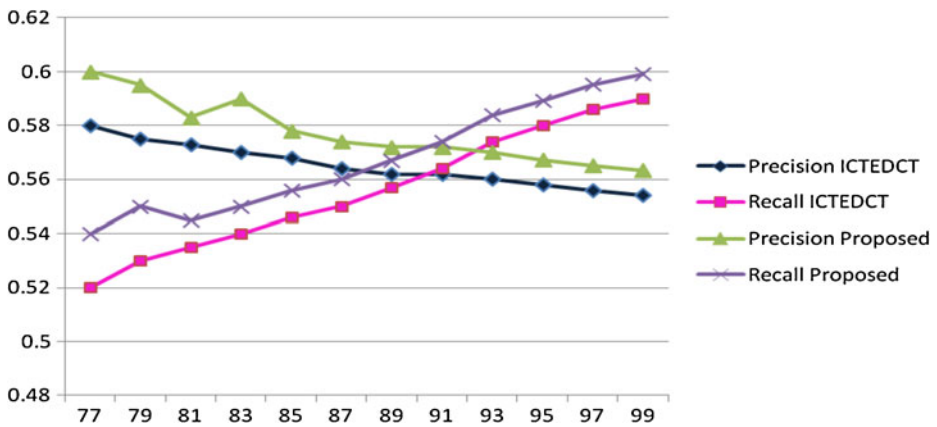
Fig. 12 Precision and Recall results on Coil dataset [30]

compared against each category. From the results elaborated in Fig. 12, it can be clearly observed that proposed method is giving higher recall and precision rates as compare to ICTEDCT [30]. Hence from the results of proposed method on Coil and
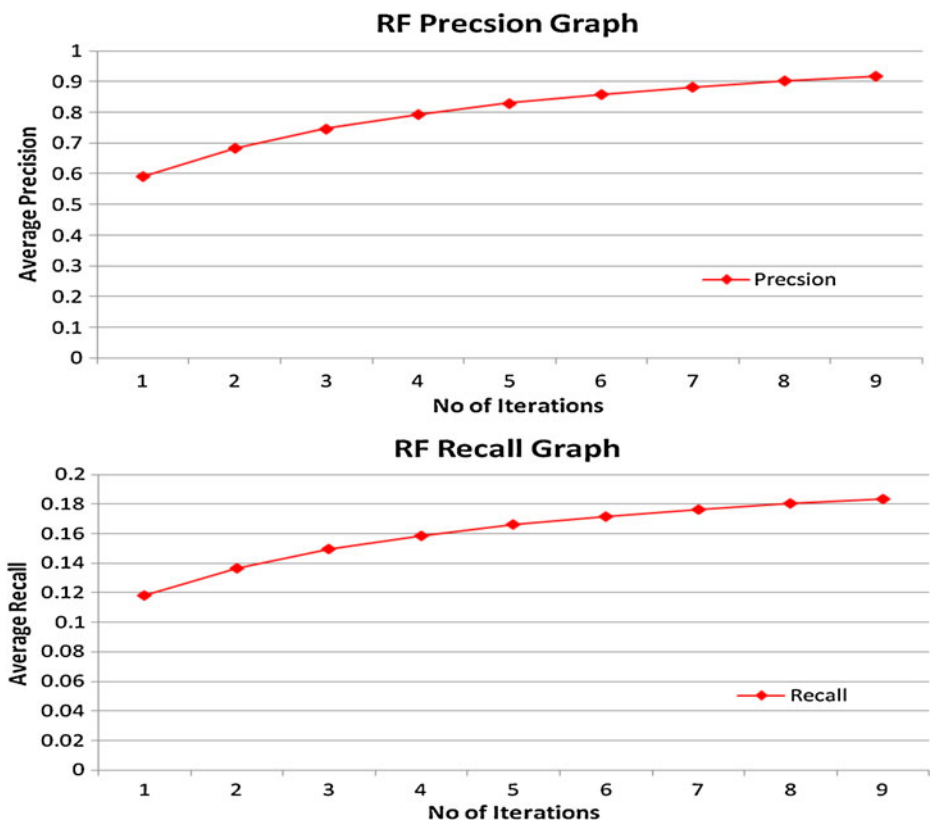

Fig. 13 Precision and Recall graphs for Relevance feedback

Corel datasets, we can say that proposed method is much more precise and effective as compare to other CBIR systems.

4.4 Relevance feedback

Relevance feedback (RF) system we have proposed generates the initial output against a query image on the base of Pearson correlation. For this it returns a set of images to the user which appears close to the query image in terms of distance as described in Section 3. User gives initial feedback by selecting only positive images in returned output, rest of the images in the returned output are considered as negative feedbacks. We train backpropagation neural networks on these two classes of inputs and output is generated. User gives feedback and process continues until user become satisfied from the output.

In our experiments, we randomly selected 300 images and then relevance feedback is automatically performed by the computer. All query relevant images (i.e., images with the same concept as the query) are marked as positive feedback samples and all the other images are marked as negative feedback samples. Method is tested on top 20 returned images. We have used 9 iterations of RF for our experiments; in which the 0th iteration returns the results obtained through Pearson correlation. The performance of the system has been measured as averaged precision, and recall values (Fig. 13). Precision and recall curves are the average values of 300 queries. The precision curve evaluates the effectiveness of a given algorithm and recall curve evaluates the robustness of the algorithm.

5 Conclusion

The paper has introduced an image retrieval system, which is based on the concept of semantic class association through trained neural networks. For an efficient image retrieval system, it is necessary that it must focus on three main things. One is to uniquely represent the thumb impact of repository images which is the image signature; secondly to measure the similarity with other repository images it must have an efficient similarity calculation way; and lastly it must be able to retrieve the image results which are semantically similar to the query image. So in this paper we have focused on these three key issues. Wavelet packets based signature development is introduced which are fused with Eigen Gabor values; this makes a wonderful combination for signature development that it guarantees retrieval in an efficient way. For similarity calculation, Pearson correlation based similarity calculation is introduced; and for efficient retrieval, neural networks based technique is introduced; according to which semantic association occurs in a more systematic way. To avoid the risk of mis-association relevance feedback is also incorporated. The results of the proposed method are compared with several standard retrieval systems.

References

1. Andrysiak T, Chora's M (2005) Texture image retrieval based on hierarchal Gabor filters. Int J Appl Math Comput Sci 15:471–480

2. Babu Rao M, Prabhakara Rao B, Govardhan A (2011) Content based image retrieval system based on dominant color, texture and shape. Int J Eng Sci Technol (IJEST) 4:2887–2896
3. Babu Rao M, Rao BP, Govardhan A (2011) CTDCIRS: content based image retrieval system based on dominant color and texture features. Int J Comput Appl 18:0975–8887
4. Bian W, Tao D (2010) Biased discriminant euclidean embedding for content-based image retrieval. IEEE Trans Image Process 19:545–554
5. Bunte K, Biehl M, Jonkman MF, Petkov N (2011) Learning effective color features for content based image retrieval in dermatology. Elsevier J Pattern Recogn 44:1892–1902
6. Chen Y, Wang JZ, Krovetz R (2005) CLUE: cluster-based retrieval of images by unsupervised learning. IEEE Trans Image Process 14:1187–1201
7. da Silva AT, Falco AX, Magalhes LP (2011) Active learning paradigms for CBIR systems based on optimum-path forest classification. Elsevier J Pattern Recogn 44:2971–2978
8. Faloutsos C, Barber R, Flickner M, Hafner J, Niblack W, Petkovic D, Equitz W (1994) Efficient and effective querying by image content. J Intell Inf Syst 3:231–262
9. Flickner M, Sawhney H, Niblack W, Ashley J, Huang Q, Dom B et al (1995) Query by image and video content: the QBIC system. IEEE Comput 28:23–32
10. Gupta A, Jain R (1997) Visual information retrieval. Commun ACM 40:70–79
11. Kijsipongse E, U-ruekolan S, Ngamphiw C, Tongsima S (2011) Efficient large Pearson correlation matrix computing using hybrid MPI/CUDA. In: Eighth international joint conference on computer science and software engineering (JCSSE)
12. Lama M, Disney T et al (2007) Content based image retrieval for pulmonary computed tomography nodule images. In: SPIE medical imaging conference. San Diego
13. Lai C-C, Chen Y-C (2011) A user-oriented image retrieval system based on interactive genetic algorithm. IEEE Trans Instrum Meas 60:3318–3325
14. Liane A, Fan J (1993) Texture classification by wavelet packets signature. IEEE Trans Pattern Anal Mach Intell 15:1186–1191
15. Lin C-H, Chen R-T, Chan Y-K (2009) A smart content-based image retrieval system based on color and texture feature. Elsevier J Image Vis Comput 27:658–665
16. Quellec G, Lamar M et al (2010) Wavelet optimization for content-based image retrieval in medical databases. Elsevier J Med Image Anal 14:227–241
17. Rao NG, Kumar VV Dr, Karishna VV (2009) Texture based image indexing and retrieval. IJCSNS Int J Comput Sci Netw Secur 9:206–210
18. Seo K-K (2007) Content-based image retrieval by combining genetic algorithm and support vector machine. Lect Notes Comput Sci 37:537–545
19. Smith JR, Chang S-F (1996) VisualSEEK: a fully automated contentbased query system. In: Proc. 4th ACM int. conf. multimedia, date, pp 87–98
20. Stejic Z, Takama Y, Hirota K (2003) Genetic algorithm-based relevance feedback for image retrieval using local similarity patterns. Inf Process Manag 39:1–23
21. Su J-H, Huang W-J, Yu PS, Tseng VS (2011) Efficient relevance feedback for content based image retrieval by minning user navigation patterns. IEEE Trans Knowl Data Eng 23:360–372
22. Tao D, Tang X, Li X, Li X (2006) Asymmetric bagging and random ubspace for support vector machines-based relevance feedback in image retrieval. IEEE Trans Pattern Anal Mach Intell 28:1088–1099
23. Tong S, Chang E (2001) Support vector machine active learning for image retrieval. In: Proc. ACM int. conf. multimedia, pp 107–118
24. Wang JZ, Wiederhold G, Firschein O, Sha XW (1998) Content based image indexing and searching using Daubechies' wavelets. Int J Digit Libr 4:311–328
25. Wang JZ, Li J, Wiederhold G (2001) Simplicity: semantics-sensitive integrated matching for picture libraries. IEEE Trans Pattern Anal Mach Intell 9:947–963
26. Weber M, Crilly PB, Blass WE (1991) Adaptive noise filtering using an error backpropagation neural network. IEEE Trans Instrum Meas 40:820–826
27. Wei B, Tao D (2010) Biased discriminant Euclidean embedding for content-based image retrieval. IEEE Trans Image Process 19:545–554
28. Yildizer E, Balci AM, Hassan M, Alhajj R (2012) Efficient content-based image retrieval using multiple support vector machines ensemble. Elsevier J Exp Syst Appl 39:2385–2396
29. Yildizer E, Balci AM, Hassan M, Alhajj R (2012) Efficient content-based image retrieval using multiple support vector machines ensemble. Exp Syst Appl 39:2385–2396
30. Youssef SM (2012) ICTEDCT-CBIR: Integrating curvelet transform with enhanced dominant colors extraction and texture analysis for efficient content-based image retrieval. Comput Electr Eng 38(5):1358–1376

31. Zhang J (2011) Robust content based image retrieval of multiple example queries. Doctor of Philosphy thesis, School of Computer Science and Software Engg. Univ. of Wollongong. http://ro.uow.edu.au/theses/3222
32. Zhang L, Fuzong L, Zhang B (2002) A CBIR method based on color-spatial feature. In: Proceedings of the IEEE Region 10 conference, TENCON 99
33. Zhang D (2004) Improving image retrieval performance by using both color and texture features. IEEE Conference on Image and Graphics, pp 172–175
34. Zhou XS, Huang TS (2001) Edge based structural features for content based image retrieval. Elsevier J Pattern Recogn Lett 22:457–468

Aun Irtaza   received his MS degree in computer science in 2009 from National University of Computer and Emerging Sciences, NU-FAST Islamabad, Pakistan. Currently he is pursuing his PhD degree in computer science at the Department of Computer Science, National University of Computer and Emerging Sciences, NU-FAST, Islamabad. His research interests include image processing and computational intelligence.

M. Arfan Jaffar   received his BSc degree from Bahauddin Zakariya University Multan, Pakistan in 2000 and got distinction (Gold Medal). He later received MSc degree in computer science in 2003 from Quaid-e-Azam University Islamabad, Pakistan. Then he earned his PhD degree in computer science in 2009 from National University of Computer and Emerging Sciences, NU-FAST, Islamabad, Pakistan. Currently he is working as a research professor at the Department of Mechatronics, GIST, Korea. His research interests include image processing and computational intelligence.

Eisa Aleisa received PhD degree from department of computer science, Lehigh University, Bethlehem, PA, USA in 2000. He is working as a Dean of College of Computer and Information Sciences at Imam University, Riyadh, Saudi Arabia. His research interests include heterogeneous distributed networks, image processing and machine learning techniques.

Tae-Sun Choi received a BS degree in electrical engineering from Seoul National University, Seoul, Korea, in 1976, an MS degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Seoul, Korea, in 1979, and a PhD degree in electrical engineering from the State University of New York at Stony Brook, in 1993. He is currently a professor in the Department of Mechatronics at the Gwangju Institute of Science and Technology, Gwangju, Korea. His research interests include image processing, machine/robot vision and visual communications.