



An effective hybrid framework for content based image retrieval (CBIR)

Umer Ali Khan¹ · Ali Javed¹ · Rehan Ashraf²

Received: 12 March 2020 / Revised: 6 October 2020 / Accepted: 24 November 2020 /
Published online: 12 May 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

In recent years, we have witnessed a massive growth in the generation of images on the cyberspace which demands to develop automated solutions for effective content management. Content-based image retrieval (CBIR) systems have been proposed to reduce the dependency on textual annotations-based image retrieval systems. There exists a variety of features-classifier combinations based CBIR methods to analyze the content of query image for relevant images retrieval. Although, these methods provide better retrieval performance in single-class scenario, however, we experience a significant performance drop in multi-class search environments due to semantics similarity among the images of different classes. CBIR methods based on the hybrid classification model offer better retrieval accuracy, however, we experience a biased classification towards the negative class due to the class imbalance problem when we experience an increase in the number of negative samples due to highly correlated semantic classes. Thus, multiple classifiers based CBIR models become unstable especially in one-against-all classification settings. To address the aforementioned problem, we proposed a CBIR method based on a hybrid features descriptor with the genetic algorithm (GA) and SVM classifier for image retrieval in multi-class scenario. More specifically, we employed the first three color moments, Haar Wavelet, Daubechies Wavelet and Bi-Orthogonal wavelets for features extraction, refine the features using GA and then train the multi-class SVM using one-against-all approach. L_2 Norm is used as a similarity measurement function between the query image and retrieved images against the query image from the image repository. The proposed technique successfully addresses the class imbalance problem in CBIR. Performance of the proposed method is evaluated on four standard datasets i.e. WANG, Oxford Flower, CIFAR-10, and kvasir and compared with 25

✉ Rehan Ashraf
rehan@ntu.edu.pk

Umer Ali Khan
umerdcs@gmail.com

Ali Javed
ali.javed@uettaxila.edu.pk

¹ Department of Software Engineering, University of Engineering & Technology, Taxila, Pakistan

² Department of Computer Science National Textile University, Sheikhpura Road, Faisalabad, Pakistan

different CBIR methods. Experimental results illustrate that our method outperforms the existing state-of-the-art CBIR methods in terms of image retrieval.

Keywords CBIR · Discrete wavelet transform · Support vector machine · Genetic algorithm · L_2 norm · Color moments · Haar wavelet · Daubechies wavelet · Bi-orthogonal wavelet transform

1 Introduction

The exponential growth of digital images in the cyberspace have sparked the research activities in developing effective image content management techniques. The rapid progression of digital capturing devices and social media in the last decade has resulted in significant expansion of the image repositories. To obtain the relevant information from these massive repositories has provided the motivation to research community to explore effective solutions without using any text-based description against every image. This exploration resulted in the solution of content-based image retrieval (CBIR). CBIR is an application of computer vision that solves the problem of searching digital images in image retrieval systems for large image datasets on the basis of image visual contents. Content-based image retrieval applications have been developed for various applications such as surveillance systems [14], Geographical Information Systems [19], Remote Sensing [56], Architectural Design [11], Medical Image Retrieval [4, 8, 39, 42, 64] and Object Recognition [64].

Existing CBIR methods [5, 7, 10, 42] use various visual features i.e. color, texture, shape, etc. to develop the feature repository. The most similar images based on the distance are returned as an output of a CBIR system that represents the semantic response of the system against the query image. Rahman et al. combined Euclidean and Bhattacharyya distances to classify skin lesion with the help of color and texture features [42]. Lonescu et al. applied fuzzy hamming distance to measure the similarity between the query and dataset images [22]. Naik et al. proposed a boosted distance metric that weightage individual features on their discrimination power for digital histopathology to classify the breast tissues [39]. Balzerini et al. proposed a query by example CBIR system using weighted Euclidean distance as a similarity measure to classify among five classes of skin lesion [10]. These visual features based CBIR systems can achieve better image retrieval accuracy, however, performance of the existing CBIR methods degrades in case of multiple classes. Additionally, the computational cost of these methods is very high for large-scale image datasets.

Non-learning methods provide better results in certain conditions, however, the gap between the high-level semantics and low-level feature representations in different images results in degraded performance of image retrieval. Existing methods have also employed various learning-based methods [6, 12, 17, 65] to reduce this semantic gap and improve the performance of image retrieval. Zhu et al. [65] applied unsupervised visual hashing learning methodology with features semantic understanding and enhanced discriminative characteristic of hash codes. Fadaei et al. employed the particle swarm optimization algorithm, and optimally combined the curvelet transformation and color features in HSV space to evaluate parametric calculations [17]. Chaudhuri et al. applied the unsupervised graph-based similarity approach to retrieve region based remote sensing images [12]. Although, these learning-based methods effectively fill the gap between the high-level semantics and low-level feature representations in different images, however, these methods are dependent on a variety of different features and unable to perform well on all kinds of images for any given feature descriptor. Moreover, these learning-based methods are computationally more complex over non-learning based CBIR methods.

In the proposed work, first three-color moments mean, standard deviation and skewness are used to extract color features of the RGB image. Scale invariant and rotation invariant are important properties of color moment methodology. In the proposed work, we used the first three lower-order color moments because maximum information about the color distribution exists in the lower-order color moments. The usage of lower-order color moments also provides an added advantage of processing only the portion of the entire color distribution [59] that helps to reduce the computational cost of features extraction. Additionally, less features are employed for comparison in image retrieval task. These benefits of lower-order color moments help to reduce the time of image retrieval. Since we also aim to improve the accuracy of our feature descriptor, therefore, we employed the texture features in combination of color features for image representation. For this purpose, we used the two-dimensional discrete wavelet transforms i.e. Haar, Daubechies, and Biorthogonal wavelets. Haar wavelet computation is simple, fast, memory efficient and robust against the edge effects problem. Daubechies wavelets belonging to the class of orthogonal wavelets are useful in terms of providing the maximum number of vanishing moments. Taking advantage of the vanishing moments, this wavelet can help to extract more smoothing features. Biorthogonal wavelets are superior over orthogonal wavelets in terms of minimizing false features extraction by constructing the symmetric wavelet functions. Discrete wavelet transforms (DWT) represent an effective multi-scale image representation that is capable of analyzing those image patterns which are difficult to analyze in one resolution. Wavelets are limited in both the time and frequency domains and hence provide an effective alternative to frequency analysis. The main contributions of the proposed work are:

- Optimization of multiple support vector machine architecture using Genetic Algorithm.
- Genetic algorithm is used to choose best suitable features to increase the retrieval performance of proposed system.
- Propose an effective feature representation to reduce the semantic gap between high-level and low-level features.
- Rigorous experimentation was performed against 25 state-of-the-art CBIR methods on four standard datasets comprising of diverse image classes to prove the effectiveness of our method.

2 Literature review

This section provides a critical investigation of existing state-of-the-art CBIR methods. Existing CBIR methods have used various features such as color, shape, texture and spatial layout. Similarly, different interest points-based features descriptors such as local features descriptors (SIFT, SURF, HOG, FAST, Harris) and binary features descriptors (BRISK, FREAK) have been proposed to extract the features for image retrieval [21, 61]. Liu et al. [34] proposed a micro structure descriptor (MSD) that is computed using the edge orientation and color features to retrieve images. However, this approach is unable to exploit the relation between positions of dissimilar objects using global features of an image. Xia et al. [55] proposed a CBIR method for cloud computing-based models to preserve the privacy of user images. KNN was used to extract and encode the visual features and then these features were used to compute the relevance of retrieved images with the query image. Water marking based protocol was used to prevent the illegal copying of retrieved images. However, the drawback of this water marking methodology lies in lack of strength to analyze in the presence of geometric features distortions. Existing CBIR approaches have also

used Bag of visual features model widely and it proved very useful in image classification, image retrieval, automatic image annotation and object recognition task due to its high discriminative power. However, these bag of visual features-based systems has a limitation of ignoring the spatial information [27, 48]. Moreover, representation in BoVW model also lack semantic meanings. Object Bank (OB) model was used with high level image representation to meet the challenges of spatial and semantic information faced in BoVW models. However, this approach suffers from high dimensionality problem [9, 31, 62]. Recent works [46, 54] also demonstrate the efficacy of deep learning practices on scene classification. However, major drawbacks of CNN based deep learning models are the complex training process for parameters tuning, need for massive amount of training data and excessive training time. Therefore, CNN based models are also limited and cannot be considered as an optimal choice for CBIR on different datasets [28, 47].

Existing CBIR approaches have also used transform based methods for image retrieval and classification. Youssef et al. [58] combined the curvelet transform with Region based vector code book sub band clustering (RBSC) to extract dominant color and texture features. However, this method achieves lower image segmentation performance. Rao et al. [44] proposed a hybrid feature vector consisting of color, shape, texture features. Texture features are extracted using gray level co-occurrence matrix (GLCM), shape features are obtained in terms of edges by computing gradient vector and visual features are computed using SIFT and SURF descriptors based on the salient points. However, these visual features exhibited outstanding performance for CBIR. However high dimensionality of SIFT and SURF based feature vectors may suffer with the problem of over fitting for image retrieval purposes when employed with the SVM. Wang et al. [52] proposed a wavelets-based semantics classification method and applied region based matching approach to compare the images. However, this approach has limited utility due to uncertain modeling for texture classification. Chen et al. [13] introduced clustering based unsupervised technique to address the issue of uncertain modeling for texture classification and achieved better results as compared to [52]. However, the performance of this method degrades due to improper clusters segmentation and identification uncertainty.

Lai et al. [29] proposed a CBIR method based on RF and interactive genetic algorithm (IGA) using a hybrid descriptor consisting of texture and color features. However, RF based methods are not popular due to reluctance of users to provide feedback throughout the interactive image retrieval task. Thus, autonomous CBIR based methods that requires minimum user feedback in response to the query image are more demanding these days. Evolutionary computation is also important to improve the retrieval performance of CBIR systems as reported in [24, 29, 45]. The main idea behind GA's is the selection of effective subset of features that can be further used for the optimization of any classifier to improve the image retrieval performance of the proposed system. In the proposed work, we emphasize the ways through which genetic algorithm (GA) can be used to enhance the SVM retrieval performance in an autonomous way without requiring any user involvement in case of multiple semantic classes.

3 Proposed methodology

In the proposed work, Haar, Daubechies and Biorthogonal wavelets transform are used to extract the texture features, whereas, first three lower-order color moments are used to extract the color features. This hybrid feature descriptor is then used to train the support

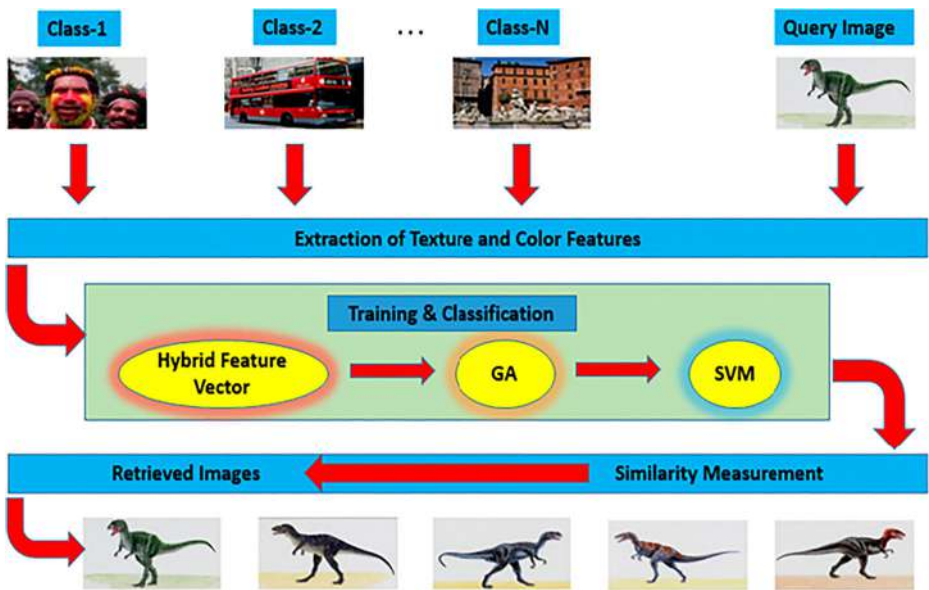


Fig. 1 Image retrieval methodology of proposed system

vector machine with the help of genetic algorithm for image classification. L_2 Norm is used to measure similarity between the dataset images and user query image. The architecture of the proposed system is shown in Fig. 1.

3.1 Extraction of color features

We employed the first three (lower-order) color moments i.e. mean, standard deviation and skewness to extract the color features from the input image in RGB color space. These three color moments are computed for each of the RGB channel to provide the 9 color moments. The details of color features extraction are provided in Algorithm 1.

Algorithm 1 Extraction of color moments of RGB image to from color feature vector.

- 1: Begin
 - 2: *Input* : *RGBImage*; P_{ij} = *Component of inputImage*
 - 3: *Output*: CFV (Color Feature Vector)
 - 4: E_i = *Meanvalue*; σ_i = *Standard deviation*; S_i = *skewness* ▷ For each image from image dataset.
 - 5: While
 - 6: $E_i = \frac{1}{N} \sum_N^{j=1} P_{ij}$ ▷ E_i is computed for all three color channels of RGB image
 - 7: $\sigma_i = \sqrt{\frac{1}{N} \sum_N^{j=1} (P_{ij} - E_i)^2}$ ▷ σ_i is computed for all three color channels of RGB image
 - 8: $S_i = \sqrt[3]{\frac{1}{N} \sum_N^{j=1} (P_{ij} - E_i)^2}$ ▷ S_i is computed for all three color channels of RGB image
 - 9: End While
 - 10: $CFV = E_i \parallel \sigma_i \parallel S_i$
 - 11: End For
 - 12: End.
-

3.2 Extraction of texture features

For textures feature extraction, two-dimensional DWT decomposition is performed on the grayscale images. At each level of decomposition, texture of an image is divided into four channels that retains information about image frequencies in low horizontal and low vertical (LL), low horizontal and high vertical (LH), high horizontal and low vertical (HL), high horizontal and high vertical (HH) channels. Image decomposition process in four channels is shown in Fig. 2.

We employ the DWT that uses low pass and high pass filtering and decomposes the image into four different sub-bands. These sub-bands contain approximated image, horizontal details of an image are contained in horizontal band, whereas, vertical and diagonal details of the image are contained in vertical and diagonal bands respectively [18, 43, 51]. It has been observed in that HH channel contain the noise in an image, which decrease the classification performance of the proposed system. Therefore, we removed the HH channel from the texture features.

First, we compute the Haar transforms by taking the difference between the pair of input values to compute their summation. This procedure is iterative where, sums are paired up to achieve the next scale and obtained the resultant difference and one final summation value [35]. Next, Daubechies wavelet is computed by taking the running averages and sum of differences similar to Haar wavelet. However, Daubechies wavelet is different from Haar wavelets in terms of using overlapping windows during feature extraction. Daubechies wavelet that belongs to orthogonal wavelet transform, use window overlapping to ensure that high frequency coefficients of spectrum could be able to show all of changes in high frequency components. Finally, we compute the biorthogonal wavelets transform. Biorthogonal wavelets also solve the phase distortion problem that exists in the orthogonal wavelets. Biorthogonal wavelets include spline wavelets that eliminate phase distortion issue with the help of Impulse Response filters to extract more smooth texture features [26, 41]. Spline wavelet is order N polynomial that is computed by the convolution of filter function $m_0(Z)$ as follows:

$$m_0 = \left(\frac{1 + Z^{-1}}{2} \right)^N \tag{1}$$

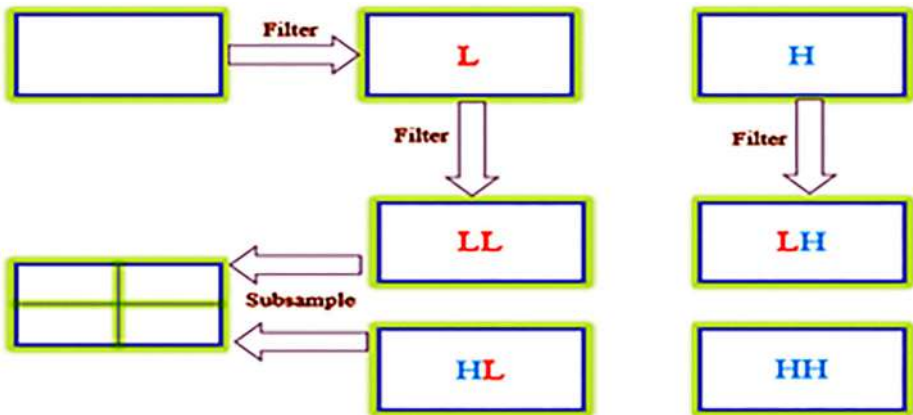


Fig. 2 Decomposition of an image using two-dimensional DWT

Each factor $\left(\frac{1+Z^{-1}}{2}\right)$ is a convolution of box function. We split the polynomial P to build biorthogonal spline (B-Spline) wavelet in the following way:

$$\sim m_0(Z) = \left(\frac{1 + Z^{-1}}{2}\right)^{\sim M} \tag{2}$$

where $\sim m_0 = \sim 2l$ is even or $\sim M = \sim 2l + 1$ is odd, and

$$m_0(x) = \left(\frac{1 + Z^{-1}}{2}\right)^{M l + \sim l - \sigma} \sum_{m=0}^{\sim l - \sigma} (l + \sim l - \sigma + m) \left(\frac{1 - Z^{-1}}{2}\right)^m \tag{3}$$

where $M = 2l$ is even or $M = 2l + 1$ is odd and $\sigma = 0$ if m is odd or $\sigma = 1$ if M is even. The texture features extraction process is presented in Algorithm 2.

Algorithm 2 Extraction of texture features of gray scale image to form texture feature vector.

- 1: Begin
 - 2: *Input* $f(x, y)$: *GrayScaleImage*, HW : *HaarWavelet*, DW : *DaubechieWavelet*, BW : *BiorthogonalWavelet* P_{ij} = *Component of input Image*
 - 3: *Output*: TFV (Texture Feature Vector)
 - 4: DWT2: 2D–Discrete Wavelet Transformation Function
 - 5: **for** each image from image dataset. **do**
 - 6: While
 - 7: $(HH, LL, LH, HL) = Dwt2(HW, \sum_{x=1}^m \sum_{y=1}^n f(x, y))$
 $\triangleright TFV_{Haar} = \frac{1}{N} \sum_N^{j=1} (LL_j + LH_j + HL_j)$
 - 8: $(HH, LL, LH, HL) = Dwt2(DW, \sum_{x=1}^m \sum_{y=1}^n f(x, y))$
 $\triangleright TFV_{Daubechie} = \frac{1}{N} \sum_N^{j=1} (LL_j + LH_j + HL_j)$
 - 9: $(HH, LL, LH, HL) = Dwt2(BW, \sum_{x=1}^m \sum_{y=1}^n f(x, y))$
 $\triangleright TFV_{Biorthogonal} = \frac{1}{N} \sum_N^{j=1} (LL_j + LH_j + HL_j)$
 - 10: End While
 - 11: $TFV = TFV_{Haar} \| TFV_{Daubechie} \| TFV_{Biorthogonal}$
 - 12: End For
 - 13: End.
-

3.3 Hybrid feature vector

We concatenated the color and texture features to develop a hybrid feature vector (HFV) for effective representation of the input images. It is important to mention that color and texture features alone are unable to better classify the images of similar color and texture, and ultimately results in increasing the false positive rate of the system. To address this vulnerability, we proposed a hybrid feature vector consisting of both the color and texture features. The overall dimension of hybrid feature vector is (100×654) for a single semantic class in case of Corel image dataset while the dimension of single feature vector is (1×654) . Dimension of color feature vector is the collective dimension of all the color features. Overall dimension of color feature vector for a single image is (1×264) . Similarly dimension of texture feature vector is the collective dimension of Haar, Daubechies and Biorthogonal feature vector. Each feature vector has the dimension (1×130) and hence the overall

dimension of texture feature vector for a single image is (1×390) . Representation of HFV is given below where n is the total no of images.

$$\begin{aligned}
 [x(1,1), x(1,2), x(1,3), \dots, x(1,264)] & \quad [y(1,1), y(1,2), y(1,3), \dots, y(1,390)] = [z(1,1), z(1,2), z(1,3), \dots, z(1,654)] \\
 [x(2,1), x(2,2), x(2,3), \dots, x(2,264)] & \quad [y(2,1), y(2,2), y(2,3), \dots, y(2,390)] = [z(2,1), z(2,2), z(2,3), \dots, z(2,654)] \\
 \dots\dots\dots & \quad \dots\dots\dots = \dots\dots\dots \\
 [x(n,1), x(n,2), x(n,3), \dots, x(n,264)] & \quad [y(n,1), y(n,2), y(n,3), \dots, y(n,390)] = [z(n,1), z(n,2), z(n,3), \dots, z(n,654)]
 \end{aligned} \tag{4}$$

3.4 Semantic classification and retrieval of images using GA with SVM

In our method, feature placement in the feature vector is not restricted with the position. Any best combination of features helps to increase the retrieval accuracy of the proposed system. This aforementioned scenario supports the use of genetic algorithm (GA). The purpose of GA is to provide best children as input to the SVM for effective training. Crossover operation is used to combine parent genes to produce two children that are evaluated on the basis fitness function $F(x)$ that is computed through L_2 distance between parents (p_1, p_2) and children (c_1, c_2) using (5) to select the children with minimum distance and standard deviation from the parents. The selection of children using the GA is provided in Algorithm 3.

$$F(x) = \text{Min-Standard-Deviation}(\text{Min-Dist} \| \sqrt{\sum_{i=1}^2 (p_i - c_i)^2 \cdot \sum_{j=2}^1 (p_i - c_i)^2} \|) \tag{5}$$

Algorithm 3 Selection of best children as input to support vector machine using GA.

- 1: Begin
- 2: **Input P** : Initial – Population, P_c : Population – Chromosomes
- 3: **CFV**: Color Feature Vector, **TFV**: Texture Feature Vector
- 4: **N**: Number of Genes, **M**: Population Size, σ : Population Size,
- 5: **Output P***: Generated Population
- 6: $P = [(CFV)^j + (TFV)^k]$, ▷ Where : $j \in 1, 2 \dots N$ and $k \in 1, 2 \dots N$
- 7: $P_c = \alpha_i$, ▷ Where : $i \in 1, 2, \dots, M$
- 8: $P^* = \Phi$
- 9: **While** (size of P^*) \neq Size of P.
- 10: **Select**: P_1, P_2 from P.
- 11: **Crossover**: P_1, P_2 to obtain h_1 , and h_2 .
- 12: **Mutate**: h_1, h_2 to obtain c_1 , and c_2 .
- 13: *if* [$L_2dis(P_1, c_1) + L_2dis(P_2, c_2) < L_2dis(P_1, c_2) + L_2dis(P_2, c_1)$]
- 14: *if* [$\sigma(c_1) < \sigma(p_1)$] then $p^* = p^* \cup C_1$ else $p^* = p^* \cup p_1$
- 15: *if* [$\sigma(c_1) < \sigma(p_2)$] then $p^* = p^* \cup C_2$ else $p^* = p^* \cup p_2$
- 16: **else**
- 17: *if* [$\sigma(c_1) < \sigma(p_2)$] then $p^* = p^* \cup C_1$ else $p^* = p^* \cup p_2$
- 18: *if* [$\sigma(c_2) < \sigma(p_1)$] then $p^* = p^* \cup C_2$ else $p^* = p^* \cup p_1$
- 19: **EndWhile**
- 20: $P = P^*$
- 21: $SVM = P$
- 22: **End**.

After obtaining the refined features descriptor through GA, we used this feature vector to train a multi-class SVM for image classification. For this purpose, we applied the one against all (OAA) approach to design a multi-class SVM classifier. An image database is divided by SVM using hyperplane in two set of classes, one class belongs to the images that resemble the query image and other class belongs to the images that do not resemble the query image. In our implementation we used maximum margin classifier principal and quadratic kernel. The quadratic kernel function is applied to map data into high dimensional space of the kernel. In kernel space, the corresponding SVM hyperplane decision function is applied, after specifying the maximum margin classifier. Now suppose input set belong to two classes [49, 57] as:

$$(x_i, y_i)_{i=1}^N \quad y_i = \{+1, -1\} \tag{6}$$

where x_i and y_i belongs to input features and corresponding labels respectively while hyperplane is determined by computing weight vector ‘W’ and bias ‘b’ as following:

$$W^T \cdot x + b = 0 \tag{7}$$

Input set can be divided into two classes as:

$$W^T \cdot x + b \geq +1 \tag{8}$$

$$W^T \cdot x + b \leq -1 \tag{9}$$

For binary classification, kernel version of Wolfe dual problem with Lagrangian multiplier α_i is computed as:

$$W(\alpha) = \sum_{i=1}^m \alpha - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{10}$$

where $\alpha_i \geq 0$ and

$$\sum_{i,j=1}^m \alpha_i y_i = 0 \tag{11}$$

The SVM decision function based on the quadratic kernel after achieving the optimal value of α_i is represented as:

$$F(x) = Sgn[g(x)] \tag{12}$$

$$g(x) = \sum_{i=1}^m \alpha_i y_i K(x, x_i) + b \tag{13}$$

where $g(x)$ is the SVM hyperplane decision function. High results for $g(x)$ mean higher prediction confidence and vice versa. In this way category specific support vector machines are trained for all semantic classes to determine the semantic class of all the images in the dataset.

3.5 Computation of similarity measurement function

To measure the similarity among the feature vector of input query image and the dataset images, we computed the L_2 norm as follows

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \tag{14}$$

4 Experimental work and results

In this section, we presented the results of different experiments designed to evaluate the performance of the proposed method. Moreover, we have also presented the discussion of the results. We used mean precision and mean recall metrics to evaluate the performance of our method over four standard datasets i.e. WANG dataset, Oxford Flower dataset, CIFAR-10 dataset and Kvasir dataset. The details of these datasets are also provided in this section. Performance of the proposed method is also compared with existing state-of-the-art methods to show the effectiveness of our method for image retrieval.

4.1 Dataset

We used four standard datasets i.e. WANG dataset, Oxford Flower dataset, CIFAR-10 dataset and Kvasir dataset for performance evaluation of the proposed CBIR method. Wang dataset consists of 1000 images of 10 semantic classes where each class contain 100 images. It is a subgroup of Corel image dataset consisting of 10908 images. The WANG dataset is publicly available at [66]. The Oxford Flower dataset consists of images belonging to 17 semantic classes, where each class comprise of 80 images. The Oxford Flower dataset is publicly available at [67]. The CIFAR-10 dataset consists of 60,000 images belonging to 10 semantic classes and each class contain 6000 images. The CFIAR-10 dataset is publicly available at [68]. Kvasir dataset [68] is a medical images dataset that consist of 4000 images belonging to 08 classes and each class comprise of 500 images. Figures 3, 4, 5 and 6 show the sample images against each semantic class for all of these three datasets. This dataset consists of anatomical landmarks semantic classes i.e. Z-line, Pylorus and cecum semantic classes. This dataset also contains pathological or endoscopic findings semantic classes i.e. esophagitis, polyps and ulcerative colitis. In addition, this dataset also provide two semantic

Clusters No.	Semantic name	A sample of the images in each cluster						
1	African people village							
2	Beach							
3	Building							
4	Buses							
5	Dinosaurs							
6	Elephants							
7	Flowers							
8	Horses							
9	Mountains and glaciers							
10	Food							

Fig. 3 Sample WANG dataset images in 10 semantic classes

Classes No.	Semantic name	A sample of the images in each class									
1	Airplane										
2	Automobile										
3	Bird										
4	Cat										
5	Deer										
6	Dog										
7	Frog										
8	Horse										
9	Ship										
10	Truck										

Fig. 4 Sample CIFAR-10 dataset images in 10 semantic classes

classes related to removal of polyp i.e. dyed lifted polyps (DLP) and dyed resection margins (DRM). This dataset is publicly available at [68].

4.2 Performance evaluation metrics

We employed the mean precision and mean recall to measure the performance of our CBIR system as these metrics are commonly used by the existing CBIR systems. Precision



Fig. 5 Sample Oxford flower dataset, one image for each semantic class

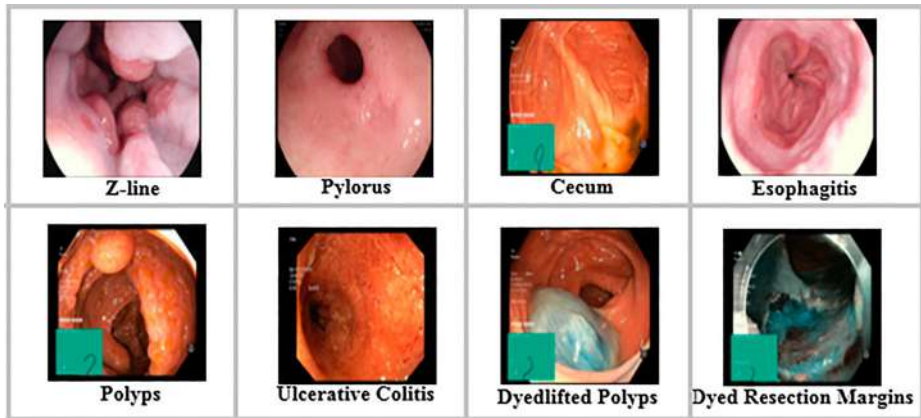


Fig. 6 Sample Kvasir dataset, one image for each semantic class

measures the system performance against the relevant models returned by the system and recall measure the system performance against all the relevant models in the entire dataset [49]. We computed the precision and recall as follows:

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

$$Recall = \frac{TP}{TP + FN} \quad (16)$$

where true positives (TP) represent the class of images of the query image correctly detected by our CBIR method, false positives (FP) are those images returned by the system that do not belong to the class of query image. False negatives (FN) represent those images belong to the class of query image that are not returned by the system. Precision or recall alone is insufficient to assess the performance of any CBIR system, therefore, we used both the precision and recall and measured the performance of the proposed system. Additionally, we also employed the mean precision-recall (PR) curve to compare the performance of the proposed and comparative methods.

4.3 Experimental results of the proposed method

This section presents the results of the proposed CBIR method evaluated on the WANG dataset, Oxford Flower dataset, CIFAR-10 dataset and Kvasir dataset. We performed a four-stage experiment and computed the mean precision and mean recall of our CBIR method. In the first stage of this experiment, we evaluated the performance of our method on WANG dataset and the retrieval results of top 20 images against the user's query image are reported in Table 1. From the results presented in Table 1, we can clearly observe that the proposed CBIR method provides better retrieval performance. More specifically, we achieved the mean precision of 0.905 and mean recall of 0.181 on WANG dataset. It is important to mention that we achieved best results in terms of precision and recall for Elephant and Dinosaur semantic classes, whereas, we achieved the worst results on the Beach and Africa semantic classes among all ten classes. The reason of getting best results on the Elephants and Dinosaurs semantic classes is that both of these classes have distinct features and high

Table 1 Results of mean precision and mean recall value of proposed method on WANG dataset on retrieving top 20 images

Semantic Class	Proposed Method Precision	Proposed Method Recall
Africa	0.85	0.17
Beach	0.80	0.16
Buildings	0.85	0.17
Buses	0.95	0.19
Dinosaurs	1.00	0.20
Elephants	1.00	0.20
Flowers	0.95	0.19
Horses	0.95	0.19
Mountains	0.85	0.17
Food	0.85	0.17
Mean	0.905	0.181

dissimilarity over other semantic classes, whereas, Beach and Africa semantic classes have similar features and less dissimilarity over other semantic classes. Therefore, we achieved low mean precision and mean recall on Beach and Africa classes of WANG dataset.

In the second stage, we computed the performance of the proposed method on Oxford Flower Dataset and results are shown in Table 2. In this stage, we evaluated the performance of our method on different number of images returned by our system. We achieved the best results on class daisy, iris, pansy, crocus, cowslip and worst results on class colts' foot, dandelion, lily valley, tiger lily. More specifically, we achieved mean precision and mean recall of 0.95 and 0.20 for class daisy, and 0.80 and 0.25 for class colts' foot. It can also be seen that best results are achieved in terms of precision and recall at lower no of returned images as compared to the results achieved at higher number of images returned by the proposed system. This demarcation between the results occur due to the fact that as the number of returned images increases, features similarity also increases among different images and hence results in an increase in the number of false positive images.

In the third stage of this experiment, we measured the performance of our system on CIFAR-10 dataset and results are reported in Table 3. Here, we evaluated the performance of the proposed system on retrieving the top 20 percent images against the query image. In this stage of the experiment, demarcation between the results in terms of mean precision and mean recall values have also been observed due to the reasons as described earlier. We achieved the best results on class automobile, bird, ship and worst results on class frog. More specifically, we achieved mean precision and mean recall of 0.979 and 0.196 for class automobile, and 0.791 and 0.158 for class frog.

Finally, in the last stage of our experiment, we evaluated the performance of our system on Kvasir dataset that consists of medical images. The results obtained on Kvasir dataset are also reported on the retrieval of top 20 percent images in terms of mean precision and mean recall as shown in Table 4. We achieved the best results on dyed lifted polyps (DLP), normal cecum and worst results on class dyed resection margins (DRM), normal pylorus. More specifically, we achieved mean precision and mean recall of 0.930 and 0.214 for class dyed

Table 2 Results of mean precision and mean recall value of proposed method on Oxford Flower on retrieving different no of images

No of Returned Images	Proposed method precision								Proposed method recall							
	20	30	40	50	60	70	80	20	30	40	50	60	70	80		
Butter cup	0.85	0.80	0.75	0.70	0.70	0.65	0.65	0.20	0.30	0.40	0.40	0.55	0.60	0.60		
Colts foot	0.80	0.75	0.70	0.65	0.60	0.60	0.65	0.25	0.30	0.50	0.50	0.55	0.55	0.60		
Daffodil	0.85	0.80	0.75	0.70	0.65	0.65	0.6	0.15	0.35	0.40	0.40	0.45	0.50	0.55		
Daisy	0.95	0.90	0.85	0.80	0.8	0.75	0.75	0.20	0.25	0.35	0.50	0.5	0.55	0.60		
Dandelion	0.80	0.75	0.70	0.65	0.65	0.60	0.55	0.20	0.30	0.45	0.45	0.45	0.50	0.60		
Fritillary	0.90	0.85	0.80	0.75	0.70	0.70	0.65	0.20	0.35	0.35	0.35	0.45	0.50	0.55		
Iris	0.95	0.95	0.90	0.85	0.80	0.75	0.70	0.25	0.30	0.40	0.45	0.45	0.50	0.60		
Pansy	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.20	0.35	0.50	0.50	0.55	0.60	0.65		
Sunflower	0.85	0.80	0.75	0.70	0.65	0.65	0.60	0.15	0.35	0.40	0.40	0.45	0.55	0.65		
Windflower	0.85	0.80	0.75	0.70	0.70	0.70	0.65	0.20	0.25	0.40	0.40	0.45	0.45	0.50		
Snowdrop	0.85	0.80	0.75	0.70	0.65	0.60	0.65	0.25	0.30	0.50	0.50	0.55	0.55	0.55		
Lily Valley	0.80	0.75	0.70	0.65	0.60	0.60	0.60	0.25	0.30	0.45	0.45	0.45	0.50	0.50		
Blue Bell	0.85	0.80	0.75	0.70	0.65	0.65	0.60	0.15	0.35	0.50	0.50	0.55	0.55	0.55		
Crocus	0.95	0.90	0.85	0.80	0.80	0.80	0.75	0.20	0.30	0.40	0.50	0.55	0.55	0.60		
Tiger Lily	0.80	0.75	0.70	0.65	0.60	0.60	0.65	0.25	0.30	0.45	0.45	0.5	0.60	0.65		
Tulip	0.90	0.85	0.80	0.75	0.75	0.70	0.70	0.20	0.25	0.35	0.50	0.55	0.60	0.65		
Cowslip	0.95	0.90	0.85	0.8	0.75	0.75	0.70	0.20	0.30	0.40	0.40	0.55	0.55	0.55		
Mean	0.874	0.826	0.776	0.726	0.694	0.674	0.653	0.203	0.306	0.424	0.450	0.503	0.541	0.585		

lifted polyps (DLP), normal caceum, and 0.910 and 0.218 for class dyed resection margins (DRM), normal pylorus. This demarcation of results is achieved due to inter class features similarity. As this similarity of features increase, we find worst results and vice versa.

Table 3 Results of mean precision and mean recall value of proposed method on CIFAR-10 dataset

Semantic class	Proposed Method Precision	Proposed Method Recall
Butter cup	0.916	0.183
Airplane	0.979	0.196
Automobile	0.85	0.75
Bird	0.979	0.196
Cat	0.958	0.192
Deer	0.937	0.187
Dog	0.875	0.175
Frog	0.791	0.158
Horse	0.916	0.183
Ship	0.979	0.196
Truck	0.833	0.166
Mean	0.916	0.183

Table 4 Results of mean precision and mean recall value of proposed method on Kvasir dataset

Semantic class	Proposed Method Precision	Proposed Method Recall
Dyed Lifted Polyps (DLP)	0.930	0.214
dyed resection margins (DRM)	0.910	0.218
Esophagitis	0.900	0.220
Normal Caceum	0.930	0.214
Normal Pylorus	0.910	0.218
Normal Z-Line	0.900	0.220
Polyps	0.920	0.216
Ulcerative Colitis	0.900	0.220
Mean	0.913	0.218

4.4 Performance comparison

We designed an experiment to compare the performance of the proposed CBIR method against the existing state-of-the-art CBIR methods [2, 3, 6, 7, 13, 15, 16, 20, 23, 25, 30, 32, 33, 36–38, 40, 44, 50, 53, 57, 58, 60, 63]. For fair performance comparison, we selected these methods as they evaluated their methods on Wang, Oxford Flower, and CIFAR-10 datasets and used the same evaluation metrics as we adopted for performance evaluation. In the work of [30], a hybrid feature descriptor comprising of DWT entropy and peak oriented octal pattern derived majority voting (POPMV) was proposed to efficiently extract the relevant color images. Methods presented in [6, 7] used the color and texture features to extract feature vector and to classify and retrieve images using the support vector machine (SVM). In [16], color SIFT, and edge-oriented histogram features were employed to generate the code book with coded word distribution to classify and retrieve the images.

In the work of [50, 60], local binary patterns, SIFT and curvelet transformation were employed for features extraction and used the K-nearest neighbor and most similar highest priority principal for classification and retrieval of images. In [57, 58], daubechies wavelet transformation and dynamic dominant color approach were used for features extraction and employed multiple SVMs and motif co-occurrence matrix to classify and retrieve the images. In [15, 32, 44], a hybrid feature vector consisting of 3D color histogram, Gabor filter, color co-occurrence matrix, and LUV color space were used with the Genetic algorithm and graph theoretic clustering algorithm for the classification and retrieval of images.

In [13, 20, 25, 36], motif co-occurrence matrix, pattern string of energy distribution, gradient sub-band vector, and SIFT features consisting of salient objects of image were used to extract the features and employed variation among the motif co-occurrence matrix of two dissimilar images, adaptive searching and region matching methods, Bag of Visual Words (BoVW) model and SVM for image classification and retrieval. The methods in [3, 37, 53] used the edge-oriented pixels, SIFT dense features, and late fusion of Fast Retina Key point features and employed the weighted code word distribution, radial basis function artificial neural network, and SVM for image retrieval. The methods in [2, 33, 38, 40, 63] used the

Table 5 Comparison of mean precision value of proposed method with other methods on retrieving top 20 images

Semantic class	Africa	Beach	Buildings	Buses	Dinosaurs	Elephants	Flowers	Horses	Mountains	Food	Mean
Proposed method	0.850	0.850	0.850	0.950	1.000	1.00	0.950	0.950	0.850	0.850	0.910
Latha and Raj [30]	0.700	0.910	0.870	0.920	1.000	0.830	0.810	0.800	0.760	0.920	0.852
Ashraf et al [6]	0.800	0.750	0.750	0.900	0.100	0.900	0.800	0.900	0.700	0.800	0.830
Ashraf et al [7]	0.650	0.700	0.750	0.950	1.000	0.800	0.950	0.900	0.750	0.750	0.820
ElAlami [16]	0.726	0.593	0.587	0.891	0.772	0.993	0.702	0.928	0.856	0.562	0.761
Tian et al [50]	0.750	0.380	0.540	0.970	0.990	0.660	0.920	0.870	0.590	0.620	0.729
Yu et al [60]	0.550	0.470	0.560	0.910	0.940	0.490	0.850	0.520	0.370	0.550	0.621
Youssef [58]	0.640	0.640	0.700	0.920	0.990	0.780	0.950	0.950	0.740	0.810	0.812
Yildizer et al [57]	0.500	0.700	0.200	0.800	0.900	1.00	0.600	0.800	0.500	0.600	0.661
Rao and Rao [44]	0.560	0.530	0.610	0.890	0.980	0.570	0.890	0.780	0.510	0.690	0.701
ElAlami [15]	0.703	0.561	0.571	0.876	0.741	0.987	0.675	0.914	0.834	0.536	0.739
Lin et al [32]	0.683	0.540	0.562	0.888	0.733	0.992	0.658	0.891	0.803	0.522	0.727
Chen et al [13]	0.500	0.450	0.350	0.650	0.950	0.300	0.750	0.700	0.300	0.600	.555
Jhanwar et al [25]	0.453	0.398	0.374	0.741	0.369	0.915	0.304	0.852	0.568	0.293	0.527
Huang and Dai [20]	0.424	0.446	0.411	0.852	0.427	0.587	0.426	0.898	0.589	0.268	0.533
Mehmood et al [36]	0.730	0.740	0.800	0.950	0.970	0.870	0.850	0.860	0.820	0.780	0.837
Irtaza et al [23]	0.650	0.600	0.620	0.850	0.930	0.650	0.940	0.770	0.730	0.810	0.755
Poursistani et al [40]	0.700	0.440	0.700	0.760	1.000	0.630	0.920	0.940	0.560	0.740	0.739
Wang et al [53]	0.640	0.540	0.530	0.940	0.980	0.780	0.710	0.930	0.420	0.50	0.697
Mehmood et al [37]	0.730	0.740	0.770	0.940	0.970	0.900	0.910	0.930	0.810	0.810	0.851
Ali et al [3]	0.630	0.600	0.680	0.920	1.000	0.720	0.910	0.800	0.590	0.580	0.743
Zeng et al [63]	0.720	0.650	0.70	0.890	1.00	0.700	0.940	0.910	0.720	0.780	0.801
Ali et al [2]	0.690	0.540	0.630	0.890	0.980	0.480	0.920	0.890	0.470	0.700	0.719
Montazer and Giveki [38]	0.580	0.480	0.530	0.950	0.980	0.640	0.850	0.800	0.540	0.630	0.698
Lin et al [33]	0.570	0.580	0.430	0.930	0.980	0.580	0.830	0.680	0.460	0.530	0.657

Table 6 Comparison of mean recall value of proposed method with other methods on retrieving top 20 images

Semantic class	Africa	Beach	Buildings	Buses	Dinosaurs	Elephants	Flowers	Horses	Mountains	Food	Mean
Proposed Method	0.170	0.170	0.170	0.190	0.200	0.200	0.190	0.190	0.170	0.170	0.182
Latha and Raj [30]	0.070	0.091	0.087	0.092	0.100	0.083	0.081	0.080	0.076	0.092	0.085
Ashraf et al [6]	0.160	0.150	0.150	0.180	0.200	0.180	0.160	0.180	0.140	0.160	0.166
Ashraf et al [7]	0.130	0.140	0.150	0.190	0.200	0.160	0.190	0.180	0.150	0.150	0.164
EAlami [16]	0.161	0.103	0.141	0.126	0.148	0.109	0.163	0.129	0.144	0.136	0.161
Tian et al [50]	0.150	0.080	0.110	0.190	0.130	0.130	0.180	0.170	0.120	0.130	0.146
Yu et al [60]	0.110	0.090	0.110	0.180	0.100	0.100	0.170	0.100	0.080	0.110	0.124
Youssef [58]	0.130	0.130	0.140	0.180	0.200	0.160	0.190	0.190	0.150	0.160	0.163
Yildizer et al [57]	0.100	0.140	0.040	0.140	0.180	0.120	0.170	0.160	0.100	0.120	0.130
Rao and Rao [44]	0.150	0.130	0.150	0.110	0.090	0.150	0.110	0.130	0.120	0.130	0.146
EAlami [15]	0.153	0.118	0.132	0.116	0.138	0.098	0.156	0.118	0.139	0.128	0.152
Lin et al [32]	0.141	0.112	0.144	0.121	0.132	0.101	0.149	0.112	0.134	0.113	0.146
Chen et al [13]	0.100	0.070	0.090	0.130	0.190	0.060	0.150	0.140	0.06	0.120	0.111
Jhanwar et al [25]	0.115	0.121	0.127	0.092	0.129	0.072	0.132	0.087	0.102	0.135	0.111
Huang and Dai [20]	0.126	0.113	0.132	0.099	0.122	0.104	0.119	0.093	0.103	0.152	0.116
Mehmood et al [36]	0.146	0.149	0.160	0.191	0.195	0.175	0.170	0.172	0.164	0.157	0.167
Irtaza et al [23]	0.130	0.120	0.124	0.17	0.186	0.130	0.188	0.154	0.146	0.162	0.151
Poursistami et al [40]	0.140	0.189	0.141	0.152	0.200	0.127	0.184	0.189	0.112	0.149	0.150
Wang et al [53]	0.128	0.108	0.106	0.188	0.196	0.156	0.142	0.186	0.140	0.100	0.145
Mehmood et al [37]	0.147	0.148	0.154	0.189	0.194	0.180	0.182	0.187	0.162	0.162	0.170
Ali et al [3]	0.127	0.122	0.136	0.185	0.200	0.145	0.183	0.160	0.119	0.117	0.149
Zeng et al [63]	0.145	0.130	0.121	0.178	0.200	0.141	0.189	0.183	0.144	0.157	0.158
Ali et al [2]	0.139	0.108	0.127	0.179	0.197	0.176	0.184	0.178	0.146	0.141	0.157
Montazer et al [38]	0.117	0.979	0.107	0.191	0.196	0.128	0.171	0.160	0.108	0.126	0.128
Lin et al [33]	0.114	0.116	0.860	0.186	0.196	0.116	0.166	0.136	0.120	0.106	0.111

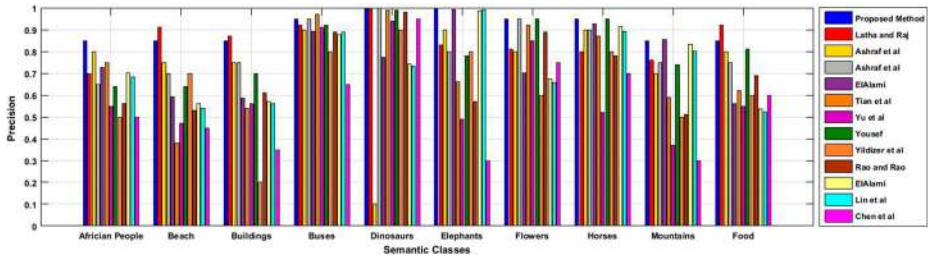


Fig. 7 Class wise comparison of precision results of proposed method with other methods

HSV and YCbCr color space and texture features to train the artificial neural network, SVM, and sequential forward selection strategies for image classification and retrieval.

4.4.1 Comparative analysis on WANG images data set

In this experiment, we performed the comparative analysis of the proposed and existing CBIR methods on WANG Image Data Set. WANG dataset consists of 1000 images and 10 semantic categories. Some images of all semantic classes such as Africa, Beach, Monuments, Buses, Food, Dinosaurs, Elephants, Flowers, Horses, and Mountains are shown in Fig. 3. We choose these semantic classes for fair performance comparison as these semantic classes are also used by [2, 3, 6, 7, 13, 15, 16, 20, 23, 25, 30, 32, 33, 36–38, 40, 44, 50, 53, 57, 58, 60, 63] to report the performance of their systems.

The results of proposed and comparative methods in terms of mean precision and mean recall are reported in Tables 5 and 6 respectively. We compared the performance of the proposed method against 25 other CBIR methods to provide a detailed comparison over a diverse range of CBIR methods. We also provided a bar chart analysis of the proposed method and these 25 comparative methods [55–79] in terms of precision values as shown in Fig. 7 and Fig. 8 in terms of recall values as shown in Figs. 9 and 10. From

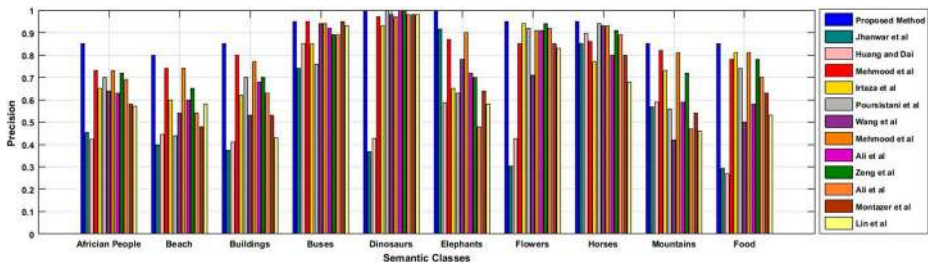


Fig. 8 Class wise comparison of precision of the proposed and comparative methods

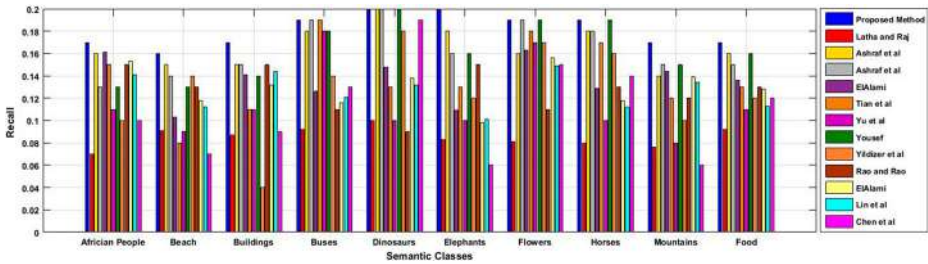


Fig. 9 Class wise comparison of recall results of proposed method with other methods

the results provided in Figs. 7, 8, 9 and 10, we can observe that the proposed CBIR model provides superior detection performance over other systems in terms of precision and recall. From the results presented in Table 5 and Table 6, we can clearly observe that the proposed CBIR method provides better retrieval performance. More specifically, we achieved the mean precision of 0.905 and mean recall of 0.181 using WANG dataset. It is important to mention that we achieved best results in terms of precision and recall for Elephant and Dinosaur semantic classes, whereas, we achieved the worst results on the Beach and Africa semantic classes among all ten classes. The reason of getting best results on the Elephants and Dinosaurs semantic classes is that both of these classes have distinct features and high inter class features dissimilarity as compare to other semantic classes, whereas, Beach and Africa semantic classes have similar inter class. Therefore, we achieved low mean precision and mean recall on Beach and Africa classes of WANG dataset.

Finally, we provided the Precision-Recall (PR) curve/graph as shown in Fig. 11 where we have provided a comparative analysis of the proposed and existing methods [2, 3, 6, 7, 13, 15, 16, 20, 23, 25, 30, 32, 33, 36–38, 40, 44, 50, 53, 57, 58, 60, 63] in terms of mean precision and mean recall. From the results obtained on the Wang dataset, we can conclude that the proposed CBIR method performs best among all the comparative methods and can reliably be used for image retrieval.

Table 7 Results of mean precision and mean recall value of proposed method on Oxford Flower on retrieving different no of images

No of Returned Images	Proposed method		[16]		[15]		[32]		[25]		[20]	
	P	R	P	R	P	R	P	R	P	R	P	R
20	0.874	0.203	0.821	0.173	0.793	0.161	0.764	0.153	0.603	0.124	0.621	0.129
30	0.826	0.306	0.804	0.261	0.765	0.241	0.701	0.228	0.491	0.159	0.502	0.163
40	0.776	0.424	0.767	0.324	0.742	0.304	0.682	0.298	0.484	0.193	0.491	0.204
50	0.726	0.450	0.706	0.413	0.687	0.386	0.668	0.367	0.479	0.235	0.488	0.246
60	0.694	0.503	0.687	0.452	0.656	0.418	0.643	0.399	0.458	0.246	0.462	0.289
70	0.674	0.541	0.665	0.489	0.632	0.435	0.616	0.422	0.436	0.263	0.453	0.328
80	0.653	0.585	0.642	0.542	0.605	0.476	0.597	0.456	0.427	0.303	0.442	0.345

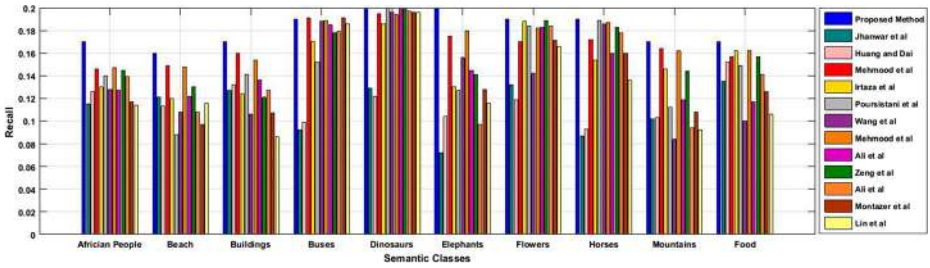


Fig. 10 Class wise comparison of recall of the proposed and comparative methods

4.4.2 Comparative analysis on Oxford flower images data set

In this experiment, we evaluated the performance of proposed and comparative techniques [15, 16, 20, 25, 32] on the Oxford flower dataset and results are reported in Table 7. This image dataset consists of 17 semantic classes and 80 images per class. To investigate the effect of different number of retrieved images, we have designed this experiment where we computed the mean precision and mean recall of all techniques on different number of returned images i.e. 20, 30, 40, 50, 60, 70, and 80 and results are shown in Table 7. It can be clearly observed that best results are achieved in terms of precision and recall at lower no of returned images as compared to the results achieved at higher number of images returned by the proposed system. This difference between the results occur due to the fact that as the number of returned images increases, features similarity also increases among different images and hence results in an increase in the number of false positive images. From the results presented in Table 7, we can argue that the proposed CBIR model provides the best mean precision and mean recall over all comparative techniques for all cases of the returned images. Similarly, Figures 12 and 13 also show that the proposed system provides superior detection performance over comparative CBIR systems in terms of mean precision and mean recall on different number of returned images.

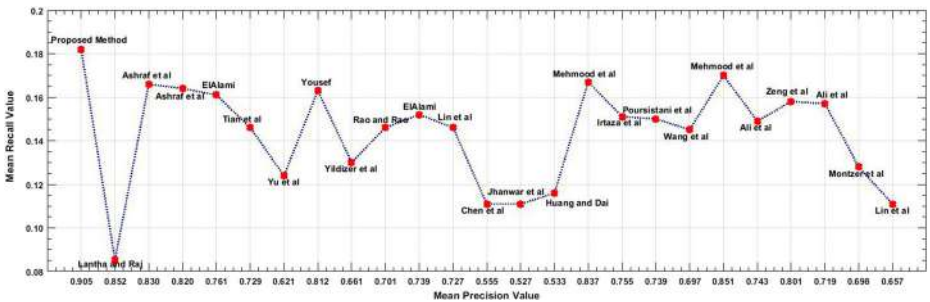


Fig. 11 Comparison of mean precision and mean recall value of proposed method with other methods

Table 8 Comparison of mean precision and mean recall value of proposed method with other methods

Semantic class	Proposed method		[15]		[32]		[25]		[20]	
	P	R	P	R	P	R	P	R	P	R
Airplane	0.916	0.183	0.762	0.178	0.732	0.167	0.521	0.129	0.568	0.146
Automobile	0.979	0.196	0.705	0.235	0.681	0.224	0.432	0.138	0.395	0.127
Bird	0.979	0.196	0.816	0.210	0.789	0.196	0.365	0.153	0.464	0.158
Cat	0.958	0.192	0.921	0.137	0.910	0.136	0.861	0.115	0.826	0.118
Deer	0.937	0.187	0.918	0.115	0.872	0.110	0.821	0.084	0.782	0.112
Dog	0.875	0.175	0.854	0.169	0.756	0.163	0.435	0.151	0.482	0.136
Frog	0.791	0.158	0.768	0.134	0.743	0.128	0.692	0.107	0.719	0.121
Horse	0.916	0.183	0.893	0.153	0.842	0.148	0.692	0.128	0.736	0.125
Ship	0.979	0.196	0.754	0.258	0.682	0.231	0.425	0.162	0.568	0.182
Truck	0.833	0.166	0.826	0.162	0.764	0.153	0.642	0.141	0.689	0.147
Mean	0.916	0.183	0.821	0.175	0.777	0.165	0.589	0.130	0.622	0.137

4.4.3 Comparative analysis on CIFAR-10 images data set

To further investigate the performance of the proposed method in terms of diversity over other comparative methods for CBIR, we designed this experiment and computed the performance on CIFAR-10 dataset consisting of 60,000 images. CIFAR-10 data set includes 10 different semantic classes as shown in the Fig. 5. CIFAR-10 dataset is further divided into training dataset and test dataset. Training data set contains 50,000 images with 5,000 images in each training class whereas, 10,000 images in the test data. The results of this experiment in terms of mean precision and mean recall are presented in Table 8 and Fig. 14 which shows the superiority of our method over comparative CBIR methods.

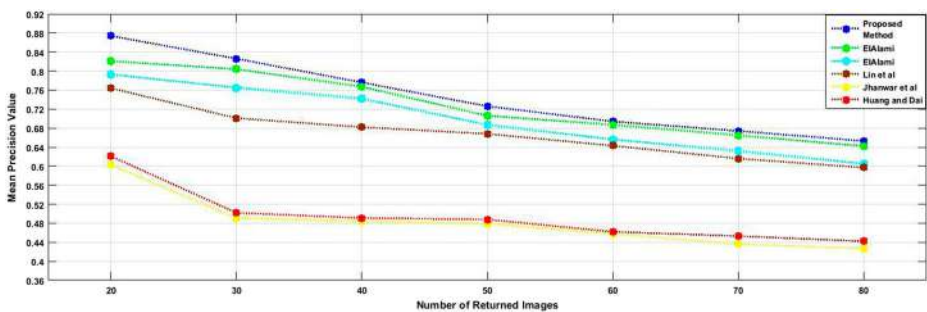


Fig. 12 Comparison of mean precision value of proposed method and other methods with different number of returned images

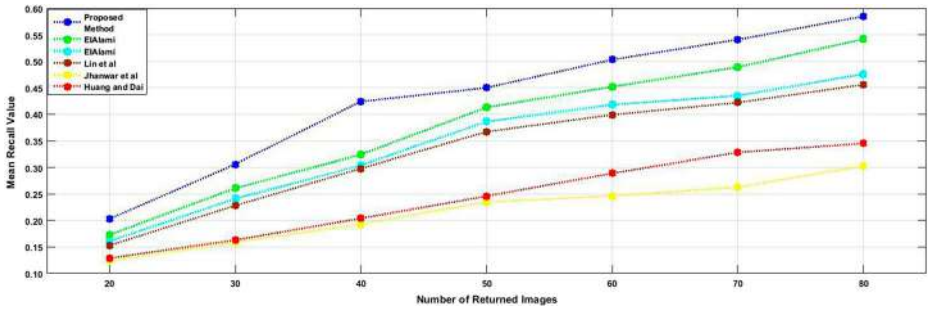


Fig. 13 Comparison of mean Recall value of proposed method and other methods with different number of returned images

4.4.4 Comparative analysis on Kvasir medical images data set

To further investigate the robustness of our method on the more diverse dataset, we selected one dataset of medical images i.e. Kvasir dataset containing a collection of 4000 medical images that belong to 08 semantic classes. Each semantic class contains 500 images. For this experiment, we have selected the RFRM method [1] for comparative analysis with the proposed method. Since this method [1] have evaluated the performance against eight base coefficients (i.e. Bray, Canberra, Chebyshev, City Block, Cosine, Euclidean, Index of Association, and Jaccard), therefore, we have also provided a comparative analysis of the proposed CBIR method against these coefficients and the RFRM method. The results of this experiment are reported in Table 9 that clearly indicate the superiority of the proposed method for image retrieval over these comparative techniques. More specifically, we have achieved the mean precision of 0.91 and mean recall of 0.218, whereas, RFRM performs second best and achieved the mean precision and mean recall of 0.86 and 0.191. We also provided a bar chart analysis of the proposed and comparative methods in terms of mean precision and mean recall in Fig. 15. From the results, we can clearly observe that the proposed method achieves best performance over all comparative techniques especially on

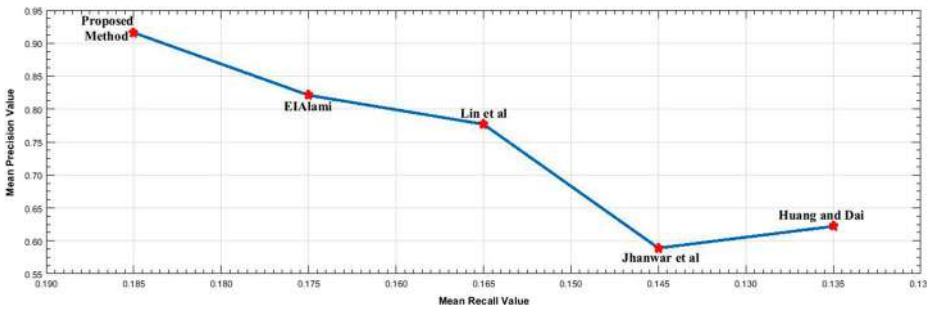


Fig. 14 Comparison of mean precision and mean recall value using CIFAR-10 dataset

Table 9 Comparing mean precision value with different methods images using Kyasir medical images dataset

Semantic class	Proposed method		RFRM		Bray		Canberra		Chebyshev		City block		Cosine	
	P	R	P	R	P	R	P	R	P	R	P	R	P	R
DLP	0.93	0.214	0.90	0.191	0.54	0.138	0.48	0.137	0.46	0.135	0.54	0.139	0.45	0.136
DRM	0.91	0.218	0.90	0.190	0.48	0.142	0.49	0.142	0.40	0.13	0.49	0.141	0.48	0.142
Esophagitis	0.90	0.220	0.85	0.180	0.47	0.138	0.49	0.139	0.41	0.135	0.49	0.138	0.47	0.140
N-Caecum	0.93	0.214	0.85	0.191	0.68	0.142	0.63	0.142	0.74	0.138	0.67	0.141	0.75	0.142
N-Pylorus	0.91	0.218	0.85	0.200	0.76	0.142	0.76	0.141	0.77	0.14	0.76	0.142	0.74	0.138
NZ-Line	0.90	0.220	0.90	0.180	0.57	0.138	0.56	0.138	0.46	0.13	0.57	0.139	0.56	0.140
Polyps	0.92	0.216	0.85	0.200	0.53	0.140	0.57	0.141	0.52	0.14	0.54	0.140	0.46	0.140
U-Colitis	0.90	0.220	0.80	0.190	0.51	0.140	0.51	0.140	0.44	0.135	0.51	0.140	0.48	0.140
Mean	0.913	0.218	0.862	0.191	0.567	0.140	0.561	0.140	0.525	0.135	0.571	0.140	0.541	0.140

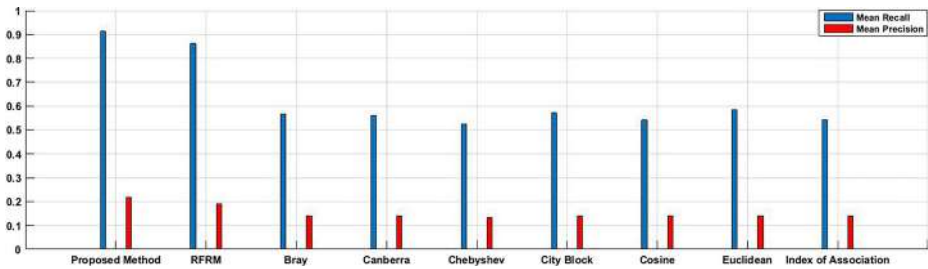


Fig. 15 Comparison of mean precision and mean recall value using Kvasir medical images dataset

the eight coefficients where our method achieves more than 40% mean precision and mean recall values.

From the results of this experiment, where we compared the performance of our CBIR method on four diverse datasets, we can argue that the proposed method can reliably be used to retrieve the images of multiple datasets containing a diverse range of images.

5 Conclusion

This paper has presented an effective CBIR system capable of semantically retrieving the correct images with high retrieval performance. For this purpose, we proposed a hybrid features descriptor comprising of color and texture features. We employed the first three color moment i.e. mean, standard deviation and skewness to extract the color features, whereas, used the Haar Wavelet, Daubechies Wavelet and Bi-Orthogonal Wavelet Transform to extract the texture features. We employed the SVM with the Genetic Algorithm for the classification of semantically similar images. L_2 Norm has been used as a similarity measurement function to measure similarity among the query image and the images returned by the proposed system. We evaluated the performance of our method on four standard image datasets (i.e. WANG Dataset, Oxford Flower Dataset, CIFAR-10 Dataset, and Kvasir medical images dataset) of different genre. Experimental results on these datasets signify the effectiveness of the proposed method in terms of image retrieval. Furthermore, we compared the performance of the proposed CBIR method with 25 different CBIR methods on these datasets. The proposed method provides superior retrieval performance over the comparative approaches and can reliably be used to retrieve the images of more diverse image classes.

References

1. Ahmed A (2020) Implementing relevance feedback for content-based medical image retrieval. *IEEE Access* 8:79969–79976
2. Ali N, Bajwa K, Sablatnig R, Chatzichristofis SA, Iqbal Z, Rashid M, Habib HA (2016) A novel image retrieval based on visual words integration of sift and surf. *PLoS ONE* 11(6):e0157428
3. Ali N, Mazhar DA, Iqbal Z, Ashraf R, Ahmed J, Khan FZ (2017). arXiv:170308492
4. Ashraf R, Ahmed M, Ahmad U, Habib MA, Jabbar S, Naseer K (2020) Mdcbir-mf: multimedia data for content-based image retrieval by using multiple features. *Multimed Tools Appl* 79(13):8553–8579
5. Ashraf R, Ahmed M, Jabbar S, Khalid S, Ahmad A, Din S, Jeon G (2018) Content based image retrieval by using color descriptor and discrete wavelet transform. *J Med Syst* 42(3):44

6. Ashraf R, Bajwa K, Mahmood T (2016) Content-based image retrieval by exploring bandletized regions through support vector machines. *J Inf Sci Eng* 32(2):245–269
7. Ashraf R, Bashir K, Irtaza A, Mahmood MT (2015) Content based image retrieval using embedded neural networks with bandletized regions. *Entropy* 17(6):3552–3580
8. Ashraf R, Habib MA, Akram M, Latif MA, Malik MSA, Awais M, Dar SH, Mahmood T, Yasir M, Abbas Z (2020) Deep convolution neural network for big data medical image classification. *IEEE Access* 8:105659–105670
9. Ashraf R, Mahmood T, Irtaza A, Bajwa K (2014) A novel approach for the gender classification through trained neural networks. *J Basic Appl Sci Res* 4:136–144
10. Ballerini L, Li X, Fisher RB, Rees J (2009) A query-by-example content-based image retrieval system of non-melanoma skin lesions. In: *MICCAI International workshop on medical content-based retrieval for clinical decision support*. Springer, pp 31–38
11. Bharadi VA, Meena M (2015) Novel architecture for cbir saas on azure cloud. In: *International conference on information processing (ICIP)*. IEEE, pp 366–371
12. Chaudhuri B, Demir B, Bruzzone L, Chaudhuri S (2016) Region-based retrieval of remote sensing images using an unsupervised graph-theoretic approach. *IEEE Geosci Remote Sens Lett* 13(7):987–991
13. Chen Y, Wang JZ, Krovetz R (2005) Clue: cluster-based retrieval of images by unsupervised learning. *IEEE Trans Image Process* 14(8):1187–1201
14. Chen X, Zhang C (2006) An interactive semantic video mining and retrieval platform—application in transportation surveillance video for incident detection. In: *Sixth international conference on data mining (ICDM'06)*. IEEE, pp 129–138
15. ElAlami ME (2011) A novel image retrieval model based on the most relevant features. *Knowl-Based Syst* 24(1):23–32
16. ElAlami ME (2014) A new matching strategy for content based image retrieval system. *Appl Soft Comput* 14:407–418
17. Fadaei S, Amirfattahi R, Ahmadzadeh MR (2016) New content-based image retrieval system based on optimised integration of dcd, wavelet and curvelet features. *IET Image Process* 11(2):89–98
18. Gupta E, Kushwah RS (2015) Combination of global and local features using dwt with svm for cbir. In: *2015 4th international conference on reliability, Infocom technologies and optimization (ICRITO)(Trends and future directions)*. IEEE, pp 1–6
19. Hafiane A, Chaudhuri S, Seetharaman G, Zavidovique B (2006) Region-based cbir in gis with local space filling curves to spatial representation. *Pattern Recogn Lett* 27(4):259–267
20. Huang PW, Dai S (2003) Image retrieval by texture similarity. *Pattern Recog* 36(3):665–679
21. Iakovidou C, Anagnostopoulos N, Lux M, Christodoulou K, Boutalis Y, Chatzichristofis SA (2019) Composite description based on salient contours and color information for cbir tasks. *IEEE Trans Image Process* 28(6):3115–3129
22. Ionescu M, Ralescu A (2004) Fuzzy hamming distance in a content-based image retrieval system. In: *IEEE international conference on fuzzy systems (IEEE Cat. No. 04CH37542)*. IEEE, vol 3, pp 1721–1726
23. Irtaza A, Jaffar MA, Aleisa E, Choi TS (2014) Embedding neural networks for semantic association in content based image retrieval. *Multimed Tools Appl* 72(2):1911–1931
24. Irtaza A, Jaffar MA, Muhammad MS (2015) Content based image retrieval in a web 3.0 environment. *Multimed Tools Appl* 74(14):5055–5072
25. Jhanwar N, Chaudhuri S, Seetharaman G, Zavidovique B (2004) Content based image retrieval using motif cooccurrence matrix. *Image Vis Comput* 22(14):1211–1220
26. Kaur S, Dadhwal HS (2015) Biorthogonal wavelet transform using bilateral filter and adaptive histogram equalization. *Int J Intell Syst Appl* 7(3):37
27. Khan R, Barat C, Muselet D, Ducottet C (2012) Spatial orientations of visual word pairs to improve bag-of-visual-words model. *British Machine Vision Conference 2012, Sep 2012, United Kingdom*. pp. 89.1–89.11, 10.5244/C.26.89. ujm-00738708
28. Khan R, Barat C, Muselet D, Ducottet C (2015) Spatial histograms of soft pairwise similar patches to improve the bag-of-visual-words model. *Comput Vis Image Underst* 132:102–112
29. Lai CC, Chen YC (2011) A user-oriented image retrieval system based on interactive genetic algorithm. *IEEE Trans Instrum Meas* 60(10):3318–3325
30. Latha D, Raj YJV (2019) Hybrid cbir method using statistical, dwt-entropy and popmv-based feature sets. *IET Image Process* 13(12):2031–2044
31. Li LJ, Su H, Lim Y, Fei-Fei L (2014) Object bank: An object-level image representation for high-level visual recognition. *Int J Comp Vision* 107(1):20–39
32. Lin CH, Chen RT, Chan YK (2009) A smart content-based image retrieval system based on color and texture feature. *Image Vis Comput* 27(6):658–665

33. Lin CH, Huang DC, Chan YK, Chen KH, Chang YJ (2011) Fast color-spatial feature based image retrieval methods. *Expert Syst Appl* 38(9):11412–11420
34. Liu GH, Li ZY, Zhang L, Xu Y (2011) Image retrieval based on micro-structure descriptor. *Pattern Recogn* 44(9):2123–2133
35. Mahmoud MI, Dessouky MI, Deyab S, Elfouly FH (2007) Comparison between haar and daubechies wavelet transform on fpga technology. *World Academy of Science. Eng Technol* 26:68–72
36. Mehmood Z, Anwar SM, Ali N, Habib HA, Rashid M (2016) A novel image retrieval based on a combination of local and global histograms of visual words. *Math Probl Eng* 2016 August:1–12
37. Mehmood Z, Anwar SM, Altaf M (2018) A novel image retrieval based on rectangular spatial histograms of visual words. *Kuwait J Sci* 1:45
38. Montazer GA, Giveki D (2015) An improved radial basis function neural network for object image retrieval. *Neurocomputing* 168:221–233
39. Naik J, Doyle S, Basavanahally A, Ganesan S, Feldman MD, Madabhushi A (2009) A boosted distance metric: application to content based image retrieval and classification of digitized histopathology. In: *Medical imaging 2009: computer-aided diagnosis*, vol. 7260. International society for optics and photonics, pp 72603F
40. Poursistani P, Nezamabadi-pour H, Moghadam RA, Saeed M (2013) Image indexing and retrieval in jpeg compressed domain based on vector quantization. *Math Comput Model* 57(5-6):1005–1017
41. Prakash O, Srivastava R, Khare A (2013) Biorthogonal wavelet transform based image fusion using absolute maximum fusion rule. In: *2013 IEEE conference on information & communication technologies*. IEEE, pp 577–582
42. Rahman MM, Desai BC, Bhattacharya P (2006) Image retrieval-based decision support system for dermatoscopic images. In: *IEEE symposium on computer-based medical systems (CBMS'06)*. IEEE, pp 285–290
43. Rangkuti AH, Hakiem N, Bahaweres RB, Harjoko A, Putro AE (2013) Analysis of image similarity with cbr concept using wavelet transform and threshold algorithm. In: *2013 IEEE symposium on computers & informatics (ISCI)*. IEEE, pp 122–127
44. Rao MB, Rao BP, Govardhan A (2011) Ctdcirs: content based image retrieval system based on dominant color and texture features. *Int Jomput Appl* 18(6):40–46
45. Saadatmand-Tarzan M, Moghaddam HA (2007) A novel evolutionary approach for optimizing content-based image indexing algorithms. *IEEE Trans Syst Man Cybernetics B Cybernet* 37(1):139–153
46. Scott GJ, England MR, Starns WA, Marcum RA, Davis CH (2017) Training deep convolutional neural networks for land-cover classification of high-resolution imagery. *IEEE Geosci Remote Sens Lett* 14(4):549–553
47. Scott GJ, Marcum RA, Davis CH, Niviv TW (2017) Fusion of deep convolutional neural networks for land cover classification of high-resolution imagery. *IEEE Geosci Remote Sens Lett* 14(9):1638–1642
48. Song Y, McLoughlin IV, Dai LR (2014) Local coding based matching kernel method for image classification. *PLoS ONE* 9(8):e103575
49. Stejić Z., Takama Y, Hirota K (2003) Genetic algorithm-based relevance feedback for image retrieval using local similarity patterns. *Inf Process Manag* 39(1):1–23
50. Tian X, Jiao L, Liu X, Zhang X (2014) Feature integration of eodh and color-sift: Application to image retrieval based on codebook. *Sig Process Image Commun* 29(4):530–545
51. Upadhyaya N, Dixit M (2016) A Review: Relating Low Level Features to High Level Semantics in CBIR. *Int J Adv Res Comput Sci Softw Eng*, 9(3):433–444
52. Wang JZ, Li J, Wiederhold G (2001) Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Trans Pattern Anal Mach Intell* 23(9):947–963
53. Wang C, Zhang B, Qin Z, Xiong J (2013) Spatial weighting for bag-of-features based image retrieval. In: *International symposium on integrated uncertainty in knowledge modelling and decision making*. Springer, pp 91–100
54. Wei W, Wang Y (2019) Color image retrieval based on quaternion and deep features. *IEEE Access* 7:126430–126438
55. Xia Z, Wang X, Zhang L, Qin Z, Sun X, Ren K (2016) A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. *IEEE Trans Inf Forensic Sec* 11(11):2594–2608
56. Ye F, Dong M, Luo W, Chen X, Min W (2019) A new re-ranking method based on convolutional neural network and two image-to-class distances for remote sensing image retrieval. *IEEE Access* 7:141498–141507
57. Yildizer E, Balci AM, Hassan M, Alhadj R (2012) Efficient content-based image retrieval using multiple support vector machines ensemble. *Expert Syst Appl* 39(3):2385–2396
58. Youssef SM (2012) Ictedct-cbir: Integrating curvelet transform with enhanced dominant colors extraction and texture analysis for efficient content-based image retrieval. *Comp Electr Eng* 38(5):1358–1376

59. Yu H, Li M, Zhang HJ, Feng J (2002) Color texture moments for content-based image retrieval, In: Proceedings international conference on image processing. IEEE, vol. 3, pp 929–932
60. Yu J, Qin Z, Wan T, Zhang X (2013) Feature integration analysis of bag-of-features model for image retrieval. *Neurocomputing* 120:355–364
61. Zafar B, Ashraf R, Ali N, Iqbal MK, Sajid M, Dar SH, Ratyal NI (2018) A novel discriminating and relative global spatial image representation with applications in cbir. *Appl Sci* 8(11):2242
62. Zang M, Wen D, Liu T, Zou H, Liu C (2018) A pooled object bank descriptor for image scene classification. *Expert Syst Appl* 94:250–264
63. Zeng S, Huang R, Wang H, Kang Z (2016) Image retrieval using spatiograms of colors quantized by gaussian mixture models. *Neurocomputing* 171:673–684
64. Zhang C, Wen G, Lin Z, Yao N, Shang Z, Zhong C (2016) An effective bag-of-visual-word scheme for object recognition. In: 2016 9th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI). IEEE, pp 417–421
65. Zhu L, Shen J, Xie L, Cheng Z (2016) Unsupervised visual hashing with semantic assistant for content-based image retrieval. *IEEE Trans Knowl Data Eng* 29(2):472–486
66. [Online]. Available: <http://wang.ist.psu.edu/docs/related/>
67. [Online]. Available: <https://www.robots.ox.ac.uk/vgg/data/flowers/17/index.html>
68. [Online]. Available: <https://www.cs.toronto.edu/kriz/cifar.html>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.