



# Content based image retrieval based on weighted fusion of texture and color features derived from modified local binary patterns and local neighborhood difference patterns

Nasim Kayhan<sup>1</sup> · Shervan Fekri-Ershad<sup>2,3</sup> 

Received: 8 October 2020 / Revised: 7 January 2021 / Accepted: 5 July 2021 /  
Published online: 6 August 2021  
© Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

Today, large amount of data are stored in image format. Content based image retrieval from bulk databases has become an interesting research topic in last decade. Most of the recent approaches use joint of texture and color information. In most cases, the color and texture features are concatenated together and equal importance is given to each one. The human visual system, usually pays more attention to the textural properties of objects to recognize. In this paper a new approach is proposed for content based image retrieval based on weighted combination of color and texture features. Firstly, to achieve discriminant features, texture features are extracted using modified local binary patterns (MLBP) and local neighborhood differences patterns (LNDP) and filtered gray level co-occurrence matrix (GLCM). Also, quantization color histogram is used to extract color features. Next, the similarity matching is performed based on canberra distance in color and texture features separately. Finally, a weighted decision is performed to retrieve most similar database images to the user query. The performance of the proposed approach is evaluated on Corel 1 K and Corel 10k datasets. Results show that proposed approach provide better performance than state-of-the-art methods in terms of precision and recall rate.

**Keywords** Content based image retrieval · Feature extraction · Texture analysis · Local binary patterns · Local neighborhood difference patterns · Weighted combination

---

✉ Shervan Fekri-Ershad  
fekriershad@pco.iaun.ac.ir

<sup>1</sup> Department of Computer Engineering, Faculty of Engineering, Shahid Ashrafi Esfahani University, Isfahan, Iran

<sup>2</sup> Faculty of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran

<sup>3</sup> Big Data Research Center, Najafabad Branch, Islamic Azad University, Najafabad, Iran

## 1 Introduction

The increasing development of computer and communication technologies has turned the category of information storage and retrieval, especially those related to images, into one of the most active domains in the development of multimedia systems. Image retrieval refers to searching an image metadatabase for images most similar to the query image and extracting them from there. It has been more than a decade that the need for content-based image retrieval in different domains such as data mining, medical imaging, education, crime prevention, etc. has been increasing. Image retrieval methods are divided into two general groups: namely, text-based image retrieval (TBIR) and content-based image retrieval (CBIR). In TBIR, methods are used to add metadata such as titles, keywords, or descriptive texts to images so that retrieval can be performed on these data. Due to problems such as the costly process of assigning texts to images, the subjectivity of labels and the choice of them being based on different tastes, the difficulty of defining image content in text format, and lack of texts for some types of images such as web pages, TBIR has become a relatively inefficient method. Attempts to solve these problems led to the emergence of CBIR systems. These systems use visual content features in an image, such as objects, color, texture, background, and so on. The method presented in this paper can be classified as a CBIR method. There are several challenges to CBIR systems, which need to be taken into consideration. The search time and the accuracy and precision of the extracted results are of great importance in these systems. Other challenges such as noise, rotation, and scale changes affect the efficacy of these systems as well.

Previous studies have shown that features such as texture, shape, and color alone cannot provide a good description of the content and objects within an image. Therefore, researchers have been encouraged, in recent years, to use hybrid approaches to describe the features of an image instead of using only one feature, such as color, shape, or texture. In hybrid methods presented so far, features extracted from images have usually been considered to have identical values in image retrieval, and weighted combinations of information and separation of feature values have not been used. Therefore, this paper presents a method, which uses a combination of local texture information and total color information in a weighted form to describe the contents of an image. The method proposed in this paper consists of two parts. The first part is allocated to texture feature extraction and color feature extraction, and the second part to weighted combinations of extracted features as well as the matching of the most similar ones. In this regard, a combination of the operators  $MLBP_{P,R}$  and  $LNDP$ , a Gaussian filter, and a co-occurrence matrix have been used at the texture feature extraction stage. In relation to color feature extraction, the input color image is divided into three separate channels; namely, H, S, and V, and a color histogram is extracted from each channel. After extracting the feature vectors of texture and color, they are merged together and the feature vector is obtained. And finally, the most efficient similarity measure with respect to the weight of each group of extracted features is used at the similarity matching stage. The performance of the proposed approach is evaluated on Corel-1k and Corel-10k datasets in terms of precision and recall. The results show that the proposed method has increased the precision and recall rates compared with many methods presented in this domain. The main contributions of the proposed approach are as follows:

- a) To reduce the effect of semantic gap, the combination of color and texture features has been used to bring the visual features closer to the semantic features. Combination color and texture features have a lot of impact on the performance of image retrieval systems. These descriptors extract low-level features that reduce computational complexity.

- b) Another main contribution of this paper is the weighting analysis of color and texture similarities in a CBIR system. It is the main novelty of this work that is proved based on experimental results. These weights allowed us to discriminate between the features in such a way that large weights are assigned to the most relevant feature entry. These feature relevance weights are designed to reduce the semantic gap between the extracted visual features and the user's high-level semantics. We have done a detailed study of the performance of different combination of weights to color and texture similarities on a large database of images. The retrieved result is dependent on the weight given to each of the similarities. It can be seen that the combination of measuring the similarity of color and texture with weight has a higher precision than the combination without weight. Adequate explanation is given in Section 4.2.1. The evaluation of the article's innovation is given in Section 5.2.

## 2 Related works

A group of studies in this domain is trying to improve the image retrieval system by working on feature extraction techniques and the other with the aid of classification algorithms. As mentioned previously, the method presented in this paper is based on the weighted combination of texture and color features of the image. Therefore, methods that have worked on the improvement of feature extraction operators, will be reviewed. In this section, this group of methods is divided into four groups based on the type of the extracted features:

### 2.1 Retrieval methods based on the analysis of color information

Color information plays a very important role in the mechanism of human visual perception. Moreover, color is invariant relative to the size of the image and the direction of objects in it. For these reasons, the color feature is often used in image retrieval to describe the contents of an image. Several methods have been presented for image retrieval based on color information. The simplest way to extract color information is to generate a color channel histogram.

The General Color Histogram (GCH) is used to display the color information of images, and the similarity between two images is determined by the distance between their color histograms. This method is sensitive to intensity variations, color distortions, and cropping. Local color histograms (LCH) divide images into blocks and obtain a histogram for each block separately. Thus, an image is represented by a combination of these histograms. To compare two images, each block in one image is compared to a similar block in another image [1].

In [14] Guang-Hai Liu et. Proposed a new method for content-based image retrieval using the color difference histogram (CDH). The unique characteristic of CDHs is that they count the perceptually uniform color difference between two points under different backgrounds with regard to colors and edge orientations in  $L^*a^*b^*$  color space. The proposed algorithm consists of two types of special histograms that calculate background colors and orientations in parallel. The first histogram shows a uniform perceptual color difference between the edge orientations of neighboring pixels. The second histogram shows the uniform perceptual color difference between the color indices of neighboring pixels and the edge orientation information.

Jaehyun An et al. in [3] proposed a content-based color image retrieval system based on color features from the salient regions and their spatial relationship. The proposed method first extracts the salient regions by a color contrast method, and finds several dominant colors for each region. Then, the spatial distribution of each dominant color is described as a binary map.

In [20] Hamed Qazanfari et al. Used the color difference histogram (CDH) in the HSV color space. CDH includes the perceptually color difference between two neighboring pixels with regard to colors and edge orientations. Among the extracted features, efficient features are selected using entropy and correlation criteria. This method consists of three steps, feature extraction, feature selection, and similarity measurement.

In [18] Metty Mustikasari et al. Have proposed a technique to retrieve images based on color feature using local histogram. In [18], the image is first converted from RGB color space to HSV color space. The image is then divided into nine sub blocks of equal size. The color of each sub-block is extracted by quantifying the HSV color space into histogram. While the process of color bin code is done, 3D local histogram is computed then the 3D histogram is converted to 1D histogram.

In [24] a color image retrieval method based on the primitives of colour moments is proposed. First, an image is divided into several blocks. Then, the colour moments of all blocks are extracted and clustered into several classes. The mean moments of each class are considered as a primitive of the image. All primitives are used as features. For color images, the YIQ color model is used in this article.

In [12] a new feature descriptor; namely, a color volume histogram, is proposed for image representation and content-based image retrieval. It converts a color image from RGB color space to HSV color space and then uniformly quantizes it into 72 bins of color cues and 32 bins of edge cues. Finally, color volumes are used to represent the image content.

## 2.2 Retrieval methods based on the analysis of texture

The texture of an image introduces information about the structural arrangement of surfaces and objects in the image. The texture of an image describes how the light intensity is distributed throughout the image. The texture information of an image plays an important role in defining the contents of the image.

In [6] Rahima Boukerma et al. Used a combination of several texture descriptors to retrieve. This system consists of three main stages including indexing, search and optimization. Extraction of texture features is performed by local binary pattern (LBP) operators, local ternary patterns (LTP) and modified local binary patterns ( $MLBP_{P,R}$ ). For the similarity step, the query image histogram is compared with the database images based on the Manhattan distance criterion. In the optimization phase, optimal weight vectors for local patterns are obtained. To do this, the Differential Evolution algorithm is used, which is first applied to an optimization database and then tested on a test database.

In [21] Provides three new texture descriptors, LotP (Local Octa Pattern), LHdP (Local Hexadeca Pattern), and DELBP (Direction Encoded Local Binary Pattern) for Content-Based Image Retrieval (CBIR). LotP and LHdP encode the relationship between center pixel and its neighbours based on the pixels' direction obtained by considering the horizontal, vertical and diagonal pixels for derivative calculations. In DELBP, direction of a referenced pixel is determined by considering every neighbouring pixel for derivative calculations. The histogram calculated from binary patterns is considered as the features vector of these patterns. The results in [21] show that LHdP is better than LBP, LTrP, LotP and DELBP in terms of precision and recall.

In [27] Prashant Srivastava et al. have proposed an image retrieval technique based on moments of wavelet transform. Discrete wavelet transform (DWT) coefficients of grayscale

images are computed which are then normalized using z-score normalization. Then Geometric moments of these normalized coefficients are computed to construct feature vectors. The proposed method consists of four steps, Computation of DWT coefficients of a grayscale image, Conversion of resulting DWT coefficients into z-score normalized coefficients. Computation of geometric moments of normalized coefficients. Similarity measurement.

### 2.3 Retrieval based on hybrid approaches

In many images, use of texture or color information alone will not be effective. To overcome this problem, a combination of features is used. In this section, we will review articles that have used hybrid approaches in image retrieval. Techniques for the extraction of various features used in these systems have their own advantages and disadvantages, and when these features are selected and combined properly, the system becomes more efficient.

In [5] Ayan Kumar Bhunia et al. have proposed a novel feature descriptors combining color and texture information collectively. In proposed color descriptor component, the inter-channel relationship between Hue (H) and Saturation (S) channels in the HSV color space has been explored which was not done earlier. Channel H is quantized into a number of bin and voting is done with saturation values and vice versa. The texture component descriptor, called the Diagonally Symmetric Local Binary Co-occurrence Pattern, considers the relationship between the pixels symmetric about both the diagonals of a  $(3 \times 3)$  window. The proposed descriptor for image retrieval has been tested in five databases. The proposed method provided satisfactory results from the experiments. Feature vector length and image retrieval rate are also competitive with most approaches.

Anu Bala et al. in [4] a novel feature descriptor, local texton XOR patterns (LTxXORP) have proposed for contentbased image retrieval. The proposed method collects the texton XOR pattern which gives the structure of the query image or database image. First, the RGB (red, green, blue) color image is converted into HSV (hue, saturation and value) color space. In this study, 7 different Texton shapes are considered to produce a Texton image, which are labeled from zero to seven. The image in the color channel  $v$  is divided into blocks  $(2 \times 2)$  that do not overlap. Each block that contains the desired Texton pattern is encoded with the corresponding label, and the Texton image is computed for the  $v$  color channel. Then Texton XOR Pattern operation is performed on the Texton image. Then, exclusive OR (XOR) operation is performed on the texton image between the center pixel and its surrounding neighbors. Finally, the feature vector is constructed based on the LTxXORPs and HSV histograms. The results after investigation show a significant improvement as compared to the state-of-the-art features for image retrieval.

In [22] a new feature descriptor, named correlated primary visual texton histogram features (CPV-THF), for image retrieval is proposed. CPV-THF integrates the visual content and semantic information of the image by finding correlations among the colour, texture orientation, intensity, and local spatial structure information of an image. The colour, texture orientation, and intensity feature histograms that are proposed in CPV-THF are represented by correlated attributes of the co-occurrence matrix. Experimental results indicate that the proposed descriptor outperforms various state-of-the-art descriptors, such as GLCM, MTH, MSD, MS-LSP, SED, STH, and MTSD.

In [16] hybrid features which combines three types of feature descriptors, including spatial, frequency, CEDD (Color and Edge Directivity Descriptors) and BSIF (Binarized Statistical Image Features) features are used to develop efficient CBIR algorithm. spatial domain features

including color auto-correlogram, color moments, HSV histogram features, and frequency domain features like moments using SWT, features using Gabor wavelet transform are used. Feature extraction using HSV histogram includes color space conversion, color quantization and histogram computation. Feature extraction using BSIF includes conversion of input RGB to gray image, patch selection from gray image, subtract mean value from all components, patch whitening process followed by estimation of ICA components. CEDD feature extraction process consists of HSV color two stage fuzzy linking system. The color moments feature extraction process includes conversion of RGB into individual components and involves computation of mean and standard deviation of each component. These stored feature vectors are then compared with the feature vectors of query image. In [16], two additional features BSIF and CEDD which improves the precision of the proposed algorithm.

In [23] have introduced a new variant of multi-trend structure descriptor (MTSD) called (Multi-direction and location distribution of pixels in trend structure (MDLDPTS)) for content-based image retrieval. In this method, If the Image is in gray-scale discrete Haar wavelet is applied upto level 3. If the Image is in RGB color space, it converted into HSV color space and is Divide image into H, S and V color channels. To extract color features, Color information is uniformly quantized. Then The image is divided into the number of non-overlapping blocks and the number of quantized colors is calculated. The Sobel operator is applied to extract the edge from the intensity image and the orientation of edges are computed then the edge orientations are uniformly quantized. The edge extracted image Divide into number of non-overlapping blocks. To extract texture features, Intensity information in V channel is uniformly quantized. Computed feature vectors are combined then data dimensionality reduction is performed. In Method [23] Due to the effective characterization of local level structures and its spatial arrangements, better performance has been achieved to retrieval natural, textural and medical images.

## 2.4 Retrieval based on deep Learning approaches

Recent years have witnessed some important advances of new techniques in machine learning. One important breakthrough technique is known as “deep learning”. Deep learning algorithms are the subset of machine learning algorithms. These algorithms have a very deep architecture with multiple layers of transformations. They work like the human brain as they train the computers according to human understandings [8]. Deep learning networks uses the neural networks and the deep term means the number of hidden layers used in the network. There are various applications in which these networks are used such as automated driving, image classification in large datasets, medical research and many more. In the recent literature, these algorithms have been used in the CBIR system for the feature extraction. This section gives the review on various previous works mainly on deep learning systems.

The authors in the article [28] used a combination of deep Convolutional neural networks and hash functions to retrieve images. In this system, the deep network and the hash function are trained simultaneously, and the feature vector extracted from each long image is converted to a much shorter code string 0 and 1 by the hash function. Therefore, the most important feature of this article has been the reduction of computational load. The highest precision obtained in this article is on the Oxford database with an average retrieval precision of 0.72. Therefore, although the computational load has decreased and the speed has increased, it has not achieved the desired precision for retrieval.

In [15] Liu et al. used a combination of features of two models of deep convolutional networks. Firstly, the improved network architecture LeNet-L is obtained by improving

convolutional neural network LeNet-5. Then, fusing two different deep convolutional features which are extracted by LeNet-5 and AlexNet. Finally, after the fusion, the similar image is obtained through comparing the similarity between the image being retrieved and the image in database by distance function.

Alrahhah et al. in [2] presented WNAHVF algorithm that jointly uses AlexNet and two visual features to retrieve and classify general and diagnostic images. WNAHVF a combined Weighted and Normalized AlexNet with Handcrafted Visual Features (BOF and LNP) for extract feature from the image and use those vectors for image retrieval and classification. WNAHVF tested method on two general datasets and one medical dataset.

In [25] described a novel and an efficient technique for Content-based image retrieval (CBIR) system which is focused on the formation of a hybrid feature vector (HFV). This HFV is formed utilizing the independent feature vectors of three visual attributes of an image, namely texture, shape and color which are extracted by using Gray level co-occurrence matrix (GLCM), region props procedure employing varied parameters and color moment respectively. These hybrid features are applied to an Extreme learning machine (ELM) deputing as a classifier which is a feed-forward neural network having one hidden layer. After that, to retrieve the higher level semantic attributes of an image, Relevance feedback is used in the form of some iterations based on the user's feedback. The average precision obtained respectively on four datasets.

Deep leaning-based appoache may provide high accuracy in some cases, but some limitation may increase these real usability in comparsion with handcrafted features.

- a) Deep Learning requires massive data sets to train on, and these should be inclusive/unbiased, and of good quality. There can also be times where they must wait for new data to be generated.
- b) This methods needs enough time to let the algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy.
- c) It also needs massive resources to function. This can mean additional requirements of computer power for you.
- d) Another major challenge is the ability to accurately interpret results generated by the algorithms. You must also carefully choose the algorithms for your purpose.

Due to the mentioned disadvantages, handcrafted feature are used in the proposed approach. Handcrafted features require less data for training than deep learning methods. They also do not require strong hardware to extract the features and image retrieval process.

## 3 Fundamentals

### 3.1 Modified local binary pattern (MLBP)

In [19], Ojala et al. proposed an improved version of the local binary pattern abbreviated to  $MLBP_{P,R}$ , which is resistant to changes in grayscale, and which recognizes the local spatial structure in the image with encoded local textures. The main contribution of this work to recognition is that some specific local binary patterns, called “uniform,” constitute the fundamental features of the local image's texture. The term “uniform” refers to the uniform appearance of a local binary pattern; that is to say, there are a limited number of transitions or



discontinuities in the circular presentation of this pattern. The uniformity degree of the number of mutations occurring in the light intensity of points present in a neighborhood is represented by the symbol “U.” For example, the binary numbers 00000000 and 11,111,111 have no mutations, and the binary number 11,010,110 shows 6 mutations. Thus, the improved operator LBP is insensitive to rotation. The homogeneity degree of a local neighborhood in  $MLBP_{P,R}$  is calculated according to Eq. (1).

$$U(LBP) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=0}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (1)$$

Most “uniform” binary patterns match primitive microfeatures such as edges, corners, and spots. Hence, they can be considered as feature detectors. The proposed texture operator allows for detecting “uniform” local binary patterns at circular neighborhoods of any quantization of the angular space and at any spatial resolution. The operator  $MLBP_{P,R}$  is derived based on a symmetric neighbor set of P members on a circle of radius R, denoted by  $LBP_{P,R}^{riu2}$ . The parameter P controls the quantization of the angular space, while R determines the spatial resolution of the operator. To analyze the contents of an image, the threshold of uniformity must be defined in a way that a large part of the image’s neighborhoods is detected as homogeneous patterns and labeled. Therefore, it has been shown in [19] that the uniformity threshold is considered equal to P/4. Therefore, patterns with a homogeneity less than or equal to P/4 are defined as uniform patterns and those with a homogeneity greater than P/4 as non-uniform patterns. According to Eq. (2), the labels 0 to P are assigned to homogeneous neighborhoods, and the label P + 1 to non-homogeneous neighborhoods. After applying this operator to texture images and assigning different labels, the probability of meeting a particular label can be approximated as the ratio of the number of points having that label to the total number of points.

$$MLBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if } U(LBP) \leq 2 \\ P + 1 & \text{otherwise} \end{cases} \quad (2)$$

For better analysis, the results of radii can be combined together and/or different radii can be considered for a neighborhood.

### 3.2 Local neighborhood difference pattern (LNDP)

In [29], Manisha Verma et al. proposed a new feature descriptor called the “Local neighborhood difference pattern (LNDP).” The proposed local feature descriptor is a method complementary to LBP. Because it extracts the relationships between neighboring pixels by comparing them to each other interactively, whereas LBP calculates the relationships of neighboring pixels with the central pixel. In LNDP, local features are extracted based on the differences of neighboring pixels from each other, and a binary pattern is generated to demonstrate each neighborhood, in the manner that a block is considered for each pixel to calculate the pattern. Each neighboring pixel in the block is compared with two adjacent and appropriate pixels. These two adjacent pixels are the closest neighbors of that pixel in the



rotation mode. The differences between these two adjacent pixels and each neighboring pixel are obtained and a binary number is determined. Similarly, the corresponding binary code for each neighborhood in the image is extracted, and finally, a histogram is created. For neighboring pixels at  $(n = 1, 2, 3, \dots, 8)$  of each central pixel  $I_c$ , LNDP can be calculated using the following method:

$$k_1^n = I_8 - I_n, \quad k_2^n = I_{n+1} - I_n, \quad \text{for } n = 1 \tag{3}$$

$$k_1^n = I_{n-1} - I_n, \quad k_2^n = I_{n+1} - I_n, \quad \forall n = 2, 3, \dots, 7 \tag{4}$$

$$k_1^n = I_{n-1} - I_n, \quad k_2^n = I_1 - I_n, \quad \text{for } n = 8 \tag{5}$$

The differences between each neighboring pixel and the other two neighboring pixels in  $k_1^n$  and  $k_2^n$  have been calculated. Based on these two differences, a binary number is assigned to each neighboring pixel (Eq. 6).

$$F_3(k_1^n, k_2^n) = \begin{cases} 1, & \text{if } k_1^n \geq 0 \text{ and } k_2^n \geq 0 \\ 1, & \text{if } k_1^n < 0 \text{ and } k_2^n < 0 \\ 0, & \text{if } k_1^n \geq 0 \text{ and } k_2^n < 0 \\ 0, & \text{if } k_1^n < 0 \text{ and } k_2^n \geq 0 \end{cases} \tag{6}$$

For the central pixel  $I_c$ , LNDP can be calculated using the binary values above, as follows:

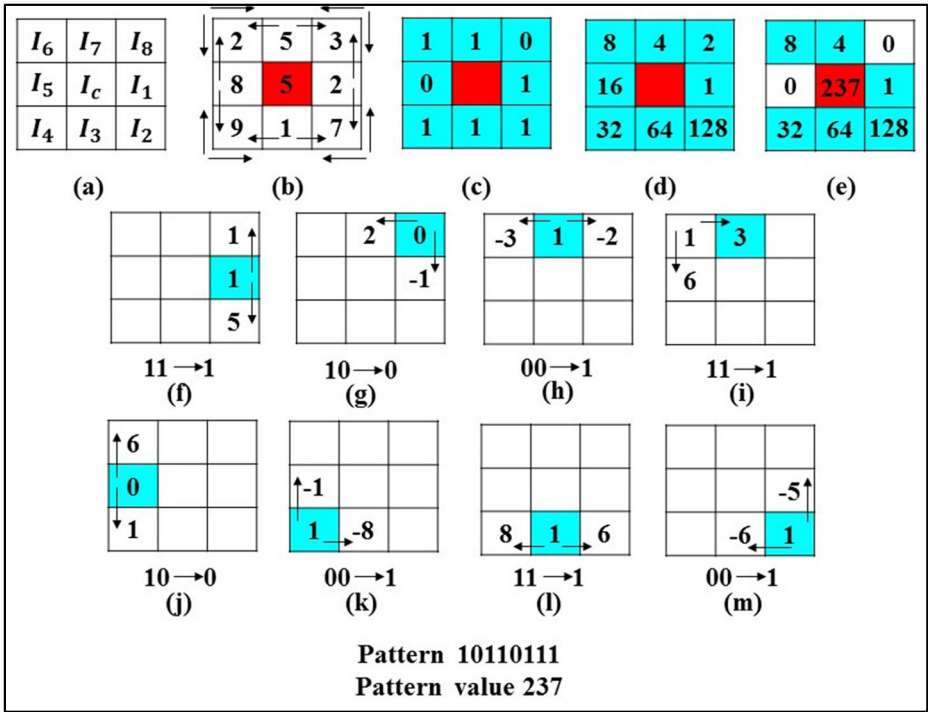
$$LNDP(I_c) = \sum_{n=1}^8 2^{n-1} \times F_3(k_1^n, k_2^n) \tag{7}$$

Eventually, the histogram of the LNDP values in the image is considered as a feature vector extracted from the image. Figure 1 shows an example of an LNDP calculation. Windows (a) and (b) show the numbering of pixels and the neighbor intensity. Windows (f) to (m) present the pattern calculation. For example, in Window (f), the pixel  $I_1$  is taken into consideration and the differences between  $I_1$  on the one hand and  $I_2$  and  $I_8$  on the other are calculated through Eq. (3), which are equal to 1 and 5, respectively. Since both differences are positive, the value of Pattern 1 is assigned to  $I_1$  using Eq. (6). Similarly, pattern values for other pixels are calculated and presented in Window (c). The pattern values are multiplied with weights shown in Window (d) and LNDP is obtained by adding the pattern values in Window (e).

### 3.3 Gaussian filter

Filters are one of the most appropriate tools that can be used to extract the features of an image. This paper uses a Gaussian filter to extract the behavior present in the texture of images. The coefficients used in the Gaussian filter are obtained through Eq. (8).

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{8}$$



**Fig. 1** Local neighborhood difference pattern calculation (a) pixel presentation. (b) a window example (f-m) pattern calculation for each neighboring pixel. (c) binary values assigned to each neighboring pixel (d) weights. (e) weights multiplied by LNDP pattern and sum up to pattern value [29]

Where; sigma represents the standard deviation of the function. The smoothing degree mostly depends on the size of the kernel and the value of sigma. In the formula above, x and y represent the width and height of the kernel, respectively. Use of a Gaussian filter allows for the examination of the texture information of an image in different conditions in terms of luminance, noise, and contrast, thus increasing the final accuracy of image recognition. It also allows for extracting the desired changes from the image by changing the standard deviation. A Gaussian filter generally acts as an intermediate band in terms of frequency behavior. This paper uses Gaussian filters with standard deviations of 1, 2, 3, and 4 on the image. It was found out during computations that considering  $\sigma$  with values greater than 4 increases the computational load.

### 3.4 Gray level Co-occurrence matrix (GLCM)

The concept of gray level co-occurrence matrix (GLCM) has been presented by Haralick et al., in which they have examined 24 features to analyze the image texture [11]. A GLCM is used to detect the occurrence frequency of pixel pairs at a specified distance and in a specific direction in an image. This matrix expresses the probability that a pixel as a reference pixel with an intensity of  $i$  occurs with a definite relationship with a pixel with an intensity of  $j$ , which is called the adjacent pixel. Therefore, each element of the matrix, which is called the element (i, j), represents the number of the occurrences of a pair of pixels with values of  $i$  and  $j$ .

which are at a relative distance of  $d$  from each other. The spatial relationship between two adjacent pixels can be determined in different ways and at different angles. From a mathematical perspective, for Image  $I$  with dimensions of  $K \times K$ , the elements of a co-occurrence matrix with dimensions of  $G \times G$  with a displacement vector of  $d$  are defined as follows:

$$M_{i,j} = \sum_{x=1}^k \sum_{y=1}^k \begin{cases} 1, & \text{if } I(x,y) = i \text{ and } I(x+d, y+d) = j \\ 0, & \text{else} \end{cases} \tag{9}$$

In Eq. (9),  $M_{i,j}$  represents the number of times two pixels with values of  $i$  and  $j$  occur at a distance of  $d$  from each other and at an arbitrary angle of  $\theta$ . Figure 2 shows an example of how to calculate a GLCM. Following statistical properties [11] are extracted in this paper from co-occurrence matrices (Table 1).

In the equations above,  $P(i, j)$  is a normalized co-occurrence matrix, the quantity  $\mu_i$  a mean calculated along the rows of the co-occurrence matrix ( $G$ ),  $\mu_j$  a mean calculated along the columns, and  $i$  and  $j$  represent the rows and columns, respectively.

Contrast measures the quantity of local changes in an image, which indicates the sensitivity of textures to changes made to intensity. If difference occurs in grayscale continuously, the texture will roughen and the contrast will increase too. If contrast has a small value, the texture will become acute. Correlation measures the degree to which a pixel relates to its neighborhood. If an image has horizontal textures, the correlation in a zero-degree direction will usually be larger than that in other directions. Variance measures contrast in an image (i.e., it measures the texture heterogeneity.) Inverse difference moment measures the texture homogeneity. Entropy measures image disorders and abnormalities, and reaches its highest value when all elements in the matrix  $P$  become equal. When the image is not homogeneous in terms of texture, the elements of the GLCM are very small, thus resulting in a high degree of entropy. Entropy is inversely related to energy. All features considered in this paper are related to distance and angle. The said 14 features have been extracted for each co-occurrence matrix that has been created in different directions.

### 3.5 The HSV color space

The HSV color space consists of three components: hue (H), saturation (S), and value (V). The component hue represents the pure color and does not change with brightness changes. Hue

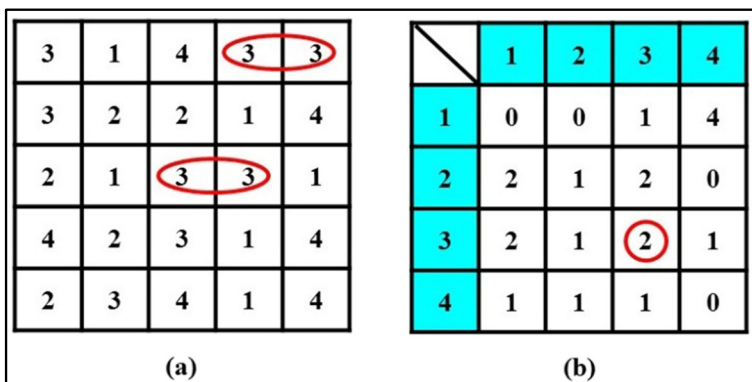


Fig. 2 : Example showing the gray level co-occurrence pattern matrix calculation in (b) for matrix (a)

**Table 1** Properties extracted from the GLCM matrix

No.	Texture feature	Expression
1	Angular Second Moment (ASM)	$\sum_{i=1}^N \sum_{j=1}^N \{p(i,j)\}^2$
2	Contrast	$\sum_{i=1}^N \sum_{j=1}^N (i-j)^2 p(i,j)$
3	Correlation	$\sum_{i=1}^N \sum_{j=1}^N \left(\frac{i-\mu_x}{\sigma_x}\right) \left(\frac{j-\mu_y}{\sigma_y}\right) p(i,j)$
4	Sum of squares : variance	$\sum_{i=1}^N \sum_{j=1}^N (i-\mu)^2 p(i,j)$
5	Inverse difference moment (IDM)	$\sum_{i=1}^N \sum_{j=1}^N \frac{p(i,j)}{1+ i-j }$
6	Sum average	$\sum_{k=2}^{2N} k p_{x+y}(k)$
7	Sum variance	$\sum_{k=2}^{2N} (k - \mu_{x+y})^2 p_{x+y}(k)$
8	Sum entropy	$-\sum_{k=2}^{2N} p_{x+y}(k) \log p_{x+y}(k)$
9	Entropy	$-\sum_{i=1}^N \sum_{j=1}^N p(i,j)^2$
10	Difference variance	$\sum_{k=0}^{N-1} (k - \mu_{x-y})^2 p_{x-y}(k)$
11	Difference entropy	$-\sum_{k=0}^{N-1} p_{x-y}(k) \log p_{x-y}(k)$
12	Information measure of correlation 1	$\frac{HXY - HXY1}{\max(HX, HY)}$
13	Information measure of correlation 2	$\sqrt{1 - \exp[-2(HXY2 - HXY)]}$
14	Maximal Correlation Coefficient	$\sqrt{\lambda_2(Q(i,j))}$

varies from 0 to 360 degrees and is characterized by red at zero degrees, green at 120 degrees, blue at 240 degrees, and so on. The component saturation represents the color intensity. Saturation indicates to what extent a pure color has been diluted by light or a white color. Its degrees are defined between Numbers 0 to 100. Number 0 is actually colorless and 100 indicates the color intensity and clearness (between these two, shades of gray occur, which fall between black and white). The component value indicates the degree of a color's brightness or its value and it is defined between Numbers 0 and 100. Number 0 is always black, and depending on the color saturation, 100 can be white. Choosing a color space compatible with the color descriptor operator is of great importance in extracting color information with higher resolution. From among the drawbacks of using the RGB color model is that the color information in these three channels is highly correlated and that it is difficult to determine a specific color in the RGB model. The HSV color space is quite similar to the way humans perceive colors, and due to its nature, the color differences are regular.

### 3.6 Color histogram

The color space used to extract the features plays a very important role in CBIR systems. Such a color space must be close to the human visual system. From among all color spaces, the HSV and \*L\*a\*b color spaces are uniform in perceptual terms, and the HSV space is more similar to the parameters of human perception. A color histogram represents the distribution of colors in

an image. A color histogram only focuses on the ratio of the number of color types regardless of the spatial location of the colors. Histogram values show the statistical distribution of colors and the main tonnage of an image. From the definition above, we can easily count the number of pixels for every 256 scales in all three HSV channels and plot them in three separate bar graphs. When we calculate different color pixels in an image, if the color space is large, we can first divide the color space into a certain number of small intervals, each of which is called a bin. This process is called color quantization. Then, by counting the number of pixels in each bin, we obtain the color histogram of the image.

### 4 The proposed method for content-based image retrieval

The content of an image is detectable by visual components such as texture, color, and shape in the human visual system. New studies show that approaches that only rely on a single component, such as color, texture, or shape, are not effective in retrieving images from databases with very different images. Therefore, in order to prepare a CBIR system, texture and color information must be extracted simultaneously. A method based on a combination of texture and color information in relation to CBIR has been proposed in this paper as well.

So far, diverse methods have been presented for image retrieval based on combinations of texture and color information. In all these methods, the system attributes equal importance to these two groups of information at the similarity measurement stage. Whereas in the human visual system, the texture usually plays a more influential role in recognizing objects and concepts, and color information as a complement makes decision-making more accurate. Therefore, first, an improved operator for local texture feature extraction with high-resolution power is presented in this paper. And then, an algorithm will be presented for the weighted measurement of similarity based on the combination of texture and color information. The general framework of the proposed method is shown in Fig. 3.

The method proposed in this paper consists of two main stages: feature extraction and similarity matching. The texture and color information of the image is extracted in the form of features with high resolution at the feature extraction stage, and a distance measure, which is the most compatible with the extracted features, is presented at the similarity matching stage.

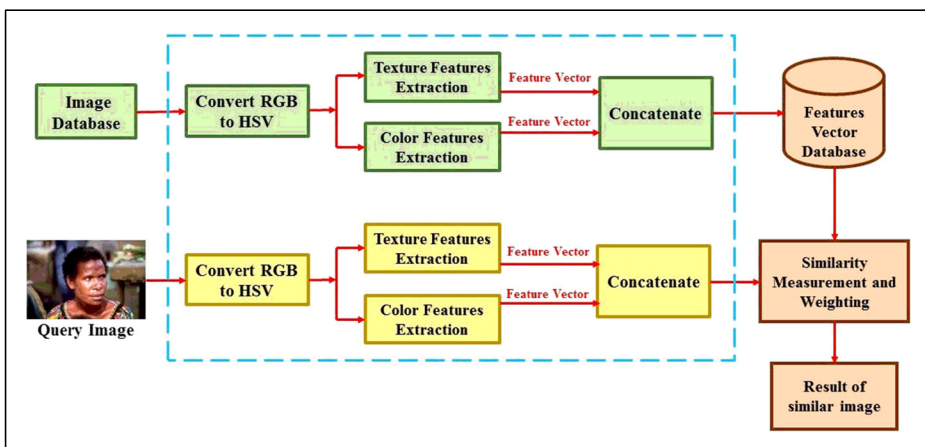


Fig. 3 Block diagram of the proposed system

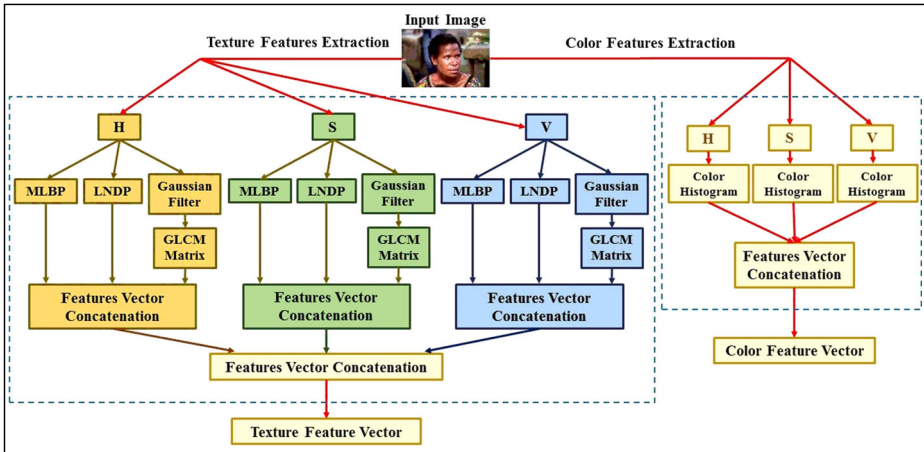


Fig. 4 Block diagram extracting texture and color features of the proposed method

Figure 4 shows the framework of the texture feature extraction and color feature extraction of the proposed method.

## 4.1 Feature extraction

### 4.1.1 Texture feature extraction

Diverse methods have so far been presented for the texture feature extraction of images. From among the existing methods, local binary patterns have been known as one of the most efficient and widely used local texture descriptors. Local binary patterns, in their original and improved forms, have attracted the attention of many researchers in this field due to their simplicity in the implementation and extraction of resolution features.

In this regard, a combination of the operators  $MLBP_{P,R}$  and LNDP has been used in this paper. The operator LNDP is a method complementary to  $MLBP_{P,R}$ . Because it extracts the relationships between neighboring pixels by comparing them to each other interactively, whereas  $MLBP_{P,R}$  calculates the relationships of neighboring pixels with the central pixel.  $MLBP_{P,R}$  and LNDP are combined because they complement each other based on the extraction of local features. Gaussian filters with different standard deviations (with values from 1 to 4) have also been used. Use of a Gaussian filter makes it possible to examine the texture information of an image in different conditions in terms of brightness, noise, and contrast, thus increasing the final precision of image detection. After applying the Gaussian filter, a GLCM is applied to the output images at different angles. A GLCM is a very popular statistical method for calculating features, which is used to detect the occurrence frequency of pixel pairs at a specified distance and in a specific direction in an image. In the feature extraction from co-occurrence matrices, usually 14 statistical features can be calculated. The aforesaid 14 features have been extracted for each co-occurrence matrix that has been created in different directions. And eventually, the feature vectors obtained from each stage are merged together and the texture feature vector is obtained.

According to the explanation given above, in this section, the image is first converted from RGB color space into HSV color space and divided into three separate channels: H, S, and V. In relation to the texture feature extraction; the following operations are performed on each channel separately:

- i) The operator  $MLBP_{P,R}$  is applied and its feature vector is extracted.
- ii) The operator LNDP is applied and its feature vector is extracted.
- iii) Gaussian filters with different standard deviations are used and a GLCM is applied to the output images at different angles.
- iv) The statistical features are extracted for each co-occurrence matrix that has been created in different directions and the feature vectors obtained from them are merged together.
- v) Eventually, the feature vectors obtained from Stages (i), (ii), and are merged together and the texture feature vector is obtained.

#### 4.1.2 Color feature extraction

As mentioned above, the human visual system first perceives the object or concept inside the image based on information about the texture or at least puts it within limited options, then makes its final decision based on color information. In this section, the image is first converted from RGB color space into HSV color space and divided into three separate channels: H, S, and V. In relation to the color feature extraction; the following operations are performed on each channel separately:

- i) A color histogram is extracted from each channel and then we use the values between the features of each color channel as a set of effective features.
- ii) The histograms extracted from the three channels, which are considered as feature vectors, are merged together, and ultimately, the color feature vector is obtained.

The advantage of this group of features is that they give us an overview of the status of the colors present in the image, which can be a very good measure for perceiving the image and comparing it with other images in order to find a similar image.

#### 4.2 Similarity matching

In a content-based image retrieval system, similarity measurement is of equal importance for retrieving and classifying images besides measuring color and texture features. The distance between the feature vector of the query image and each image from the database in the feature space is obtained through similarity measurement. Achieving high accuracy in CBIR is possible through using a distance measure compatible with the extracted features. Empirical experiments have shown that the Canberra distance measure provides the highest accuracy for measuring the similarity between the query image and the images present in the database based on the method presented in this paper.

$$\text{Canberra distance measure : } D(Q, DB) = \sum_{i=1}^L \frac{|f_{DB_i} - f_{Q_i}|}{|f_{DB_i}| + |f_{Q_i}|} \quad (10)$$

for which:



- $f_Q=(f_{Q_1},f_{Q_2},\dots,f_{Q_L})$  represents the feature vector of the query image.
- $f_{DB}=(f_{DB_1},f_{DB_2},\dots,f_{DB_L})$ , where  $i = 1, 2, \dots, L$  represents the feature vector of each of the images present in the database. And  $L$  represents the number of dimensions the feature vector has.

So, finally two distances can be computed between image query (Q) and database sample (DB) with acronym  $D_t(Q, DB)$  and  $D_c(Q, DB)$ . Where,  $D_t$  is distance between the feature vector of the query image and database sample image based on texture information. And,  $D_c$  is the distance between the feature vector of the query image and database sample image based on color information.

#### 4.2.1 Weighted fusion

As mentioned previously, the human visual system does not attribute equal importance to texture information and color information in detecting the content of images. Therefore, a simple connection of the color and texture feature vectors cannot provide much accuracy for an image retrieval system. On the other hand, operators that extract texture and color information in a combined form and in the form of a group of features cannot show this function of the human visual system properly. Therefore, an innovative algorithm of a weighted combination of texture and color features is proposed in this paper. In this method, first, the similarity between the query image and the database sample image is measured separately from two perspectives: texture and color description. Then, in the end, the final similarity between these two images will be a weighted combination of texture and color similarities. How to calculate a weighted similarity is shown in the following equation.

$$D_{final}(Q, DB) = W_t \times D_t(Q, DB) + W_c \times D_c(Q, DB) \quad \text{where } W_t + W_c = 1 \quad (11)$$

Where,  $w_t$  is the weight assigned to the distance obtained based on texture information and  $w_c$  is the weight assigned to the distance obtained based on color information.

The equation proposed for weighting texture and color similarities is a parametric equation, whose input parameters are adjustable. In this paper, based on the experiments performed, an attempt has been made to determine an optimal combination of weights based on the training of a part of images inside the database in order to provide the highest retrieval accuracy.

## 5 Experimental results

In this section, the performance of the proposed method is evaluated on two well-known databases in this domain and based on common criteria in this domain.

### 5.1 Datasets

The most common natural scene datasets, such as Corel-1k and Corel-10k, have been selected to test the mean retrieval accuracy of the proposed method. Most state-of-the-art methods use the Corel family datasets for performance validation, which is due to the variety of contents present in natural scene images.



Fig. 5 Examples of corel-1k dataset

### 5.1.1 Corel-1k dataset

The CD1k dataset is a combination of 10 semantically organized categories of natural scene images (such as beaches, buildings, buses, flowers, horses, elephants, foods, mountains, dinosaurs, and Africans). All images on CD1k are of the same size; i.e.  $(384 \times 256)$  or  $(256 \times 384)$ . Each category contains 100 images. Therefore, there are a total number of 1,000 images on CD1k. Examples of images of this database are presented in Fig. 5.

### 5.1.2 Corel-10k dataset

The second dataset used for performance evaluation is the CD10k dataset, which includes the category of CD5k dataset and 50 other categories of natural scene images. There are 100 categories on CD10k, each containing 100 images with dimensions of  $(192 \times 128)$  or  $(128 \times 192)$ . There are a total number of 10,000 images on CD10k. Examples of images of this database are presented in Fig. 6.

## 5.2 Evaluation parameters

The most common CBIR evaluation parameters are precision and recall. precision is defined as the ratio of the number of retrieved relevant images to the total number of retrieved images. Mathematically, it can be defined through Eq. (12).

$$Precision = \frac{N_R}{N_T} \quad (12)$$

Where;  $N_R$  represents the number of retrieved relevant images and  $N_T$  the total number of retrieved images for each query image. Recall is defined as the ratio of the number of retrieved relevant images to the total number of images in each category of a database.

$$Recall = \frac{N_R}{N_C} \quad (13)$$

Where;  $N_C$  represents the number of images in each category of a database.

### 5.3 Specifications of the system used in the experiments

All experiments were performed on a laptop with a 2.5GHz Intel Core i5 processor and 4GB RAM, with Windows 8 running on it. The software used in the proposed method is MATLAB R2017.



Fig. 6 Examples of corel-10k dataset

**Table 2** Evaluation the precision of the proposed method based on different radii of the operator on the corel-10k dataset

$w_t$	$w_c$	Descriptor $MLBP_{P,R}$						
		P=8 R=1	P=16 R=2	P=24 R=3	8,1+16,2	8,1+24,3	16,2+24,3	8,1+16,2+24,3
Auto	Auto	0.54366	0.54423	0.54502	0.5454	0.54601	0.54664	0.54744
0.1	0.9	0.47587	0.47614	0.47635	0.47666	0.47695	0.47734	0.4778
0.2	0.8	0.49674	0.49757	0.49784	0.49823	0.49872	0.49921	0.5003
0.3	0.7	0.51454	0.51522	0.51575	0.51649	0.517	0.51754	0.51879
0.4	0.6	0.53015	0.53155	0.53231	0.53227	0.53336	0.53388	0.53457
0.5	0.5	0.54366	0.54423	0.54502	0.5454	0.54601	0.54664	0.54744
0.6	0.4	0.55422	0.55539	0.55605	0.55568	0.55627	0.5571	0.55727
0.7	0.3	0.55983	0.5611	0.56201	0.56127	0.56227	<b>0.56312</b>	0.56291
0.8	0.2	0.5596	0.56096	0.56172	0.56008	0.56147	0.56188	0.56093
0.9	0.1	0.54289	0.54431	0.54582	0.54358	0.54485	0.54549	0.54415

## 5.4 Performance evaluation

The method proposed in this paper uses a combination of texture and color information. Therefore, during the feature extraction process, some of the input parameters of the operation must be adjusted so that the highest retrieval accuracy can be achieved. As explained before, the operator  $MLBP_{P,R}$  is used at the texture feature extraction stage of the proposed method. Therefore, different radii must be adjusted to produce the feature vector of  $MLBP_{P,R}$ . The performance of the proposed method has been evaluated based on three radii and their combination based on different weighting parameters and without weighting parameters on Corel-10k dataset. As shown in Table 2, the first row of the table shows the performance of the proposed method without consideration of the weighting parameter, which provides the highest accuracy based on a combination of three radii; i.e., R = 1, R = 2, and R = 3. When considering the weighting parameter, the results show that the feature vector of  $MLBP_{P,R}$  provides the highest performance based on a combination of two radii: R = 2 and R = 3, a weight of 0.7 for texture features, and a weight of 0.3 for color features.

**Table 3** Evaluation the recall of the proposed method based on different radii of the operator on the corel-10k dataset

$w_t$	$w_c$	Descriptor $MLBP_{P,R}$						
		P=8 R=1	P=16 R=2	P=24 R=3	8,1+16,2	8,1+24,3	16,2+24,3	8,1+16,2+24,3
Auto	Auto	0.054366	0.054423	0.054502	0.05454	0.054601	0.054664	0.054744
0.1	0.9	0.047587	0.047614	0.047635	0.047666	0.047695	0.047734	0.04778
0.2	0.8	0.049674	0.049757	0.049784	0.049823	0.049872	0.049921	0.05003
0.3	0.7	0.051454	0.051522	0.051575	0.051649	0.0517	0.051754	0.051879
0.4	0.6	0.053015	0.053155	0.053231	0.053227	0.053336	0.053388	0.053457
0.5	0.5	0.054366	0.054423	0.054502	0.05454	0.054601	0.054664	0.054744
0.6	0.4	0.055422	0.055539	0.055605	0.055568	0.055627	0.05571	0.055727
0.7	0.3	0.055983	0.05611	0.056201	0.056127	0.056227	<b>0.056312</b>	0.056291
0.8	0.2	0.05596	0.056096	0.056172	0.056008	0.056147	0.056188	0.056093
0.9	0.1	0.054289	0.054431	0.054582	0.054358	0.054485	0.054549	0.054415

**Table 4** Evaluation the precision of the proposed method based on different radii of the operator on the corel-10k dataset without considering the weighting parameter

Number of Retrieved	P=8 R=1	P=16 R=2	P=24 R=3	8,1+16,2	8,1+24,3	16,2+24,3	8,1+16,2+24,3
10	0.54366	0.54423	0.54502	0.5454	0.54601	0.54664	<b>0.54744</b>
20	0.44642	0.44706	0.44752	0.44769	0.44846	0.44902	0.44953
30	0.39386	0.39461	0.39519	0.39514	0.39596	0.39655	0.39692
40	0.35842	0.35905	0.35963	0.35953	0.36022	0.36066	0.36117
50	0.33101	0.3318	0.33209	0.33232	0.3327	0.33321	0.33367

It is worth mentioning that in all experiments carried out according to Tables 2 and 3, the color histogram has a quantization rate of  $H = 1$  (256 bins), and the number of images retrieved in this evaluation is  $N = 10$ . In all of the following Tables, the best result is indicated in the Bold manner.

In Table 4, the performance of the proposed method is evaluated on Corel-10k dataset based on the different radii of the operator  $MLBP_{P,R}$  and their combination without consideration of the weighting parameters. As can be seen, the best precision is obtained based on a combination of three radii  $R = 1$ ,  $R = 2$ , and  $R = 3$ , but this precision is lower than that obtained through the method, in which the weighting parameter was used. The results suggest the effectiveness of using the weighting parameter in the improvement of the precision of an image retrieval system. A color histogram with a quantization rate of  $H = 1$  (256 bins) was considered in this experiment too. Also, the performance of the proposed approach is evaluated in terms of recall rate (Table 5).

The color histogram is used at the color feature extraction stage of the proposed method. The quantization rate of a histogram is a parameter that affects the total number of features and the final accuracy of a CBIR system. Thus, it must be adjusted in this experiment. According to Table 6, the precision of the proposed method was evaluated based on five different quantization rates of the histogram; namely,  $H = 1, 2, 4, 8, 16$ . The number of images retrieved in this evaluation is  $N = 10$ . The results suggest that if the image histogram in each color channel contains 256 bins, the highest precision will be shown. In this experiment, the operator  $MLBP_{P,R}$  is considered based on a combination of two radii:  $R = 2$  and  $R = 3$ . Also, the performance is evaluated in terms of recall rate (Table 7).

Another important parameter in improving the precision of an image retrieval system is the distance measure or similarity measurement. In the proposed method, seven distance measures have been used and evaluated on Corel-10k dataset. As can be seen in Table 8, based on a

**Table 5** Evaluation the recall of the proposed method based on different radii of the operator on the corel-10k dataset without considering the weighting parameter

Number of Retrieved	P=8 R=1	P=16 R=2	P=24 R=3	8,1+16,2	8,1+24,3	16,2+24,3	8,1+16,2+24,3
10	0.054366	0.054423	0.054502	0.05454	0.054601	0.054664	0.054744
20	0.089284	0.089412	0.089504	0.089537	0.089691	0.089804	0.089906
30	0.11816	0.11838	0.11856	0.11854	0.11879	0.11897	0.11908
40	0.14337	0.14362	0.14385	0.14381	0.14409	0.14426	0.14447
50	0.16551	0.1659	0.16604	0.16616	0.16635	0.1666	0.16683

**Table 6** Evaluation the precision of the proposed method based on quantization of color histogram on corel-10k dataset

$w_r$	$w_c$	H=1 Bin=256	H=2 Bin=128	H=4 Bin=64	H=8 Bin=32	H=16 Bin=16
0.1	0.9	0.47734	0.48528	0.49966	0.51998	0.53971
0.2	0.8	0.49921	0.51293	0.53024	0.54374	0.55081
0.3	0.7	0.51754	0.53305	0.54523	0.55051	0.54573
0.4	0.6	0.53388	0.54583	0.55252	0.54992	0.53782
0.5	0.5	0.54664	0.55468	0.55383	0.54346	0.5294
0.6	0.4	0.5571	0.55868	0.54979	0.53571	0.52254
0.7	0.3	<b>0.56312</b>	0.55674	0.54128	0.52648	0.51571
0.8	0.2	0.56188	0.5468	0.5299	0.51764	0.51015
0.9	0.1	0.54549	0.52835	0.51623	0.50907	0.50579

weighting parameter of 0.7 for texture features and 0.3 for color features, the Canberra distance measure shows the highest precision. The number of images retrieved in this experiment is  $N = 10$ . The operator  $MLBP_{P,R}$  with a combination of two radii:  $R = 2$  and  $R = 3$ , and a color histogram with a quantization rate of  $H = 1$  (256 bins) have been taken into consideration (Table 9).

Since this paper has used different combinations of weights to evaluate the system, we are going to present an example of the effect of a weighting parameter in improving the precision of a content-based image retrieval system. When taking into account the query image in Fig. 7(a) from Corel-10k image dataset and without using a weighting parameter, it can be seen that out of the 10 images retrieved, only two images are from the query image class and the other ones are not from the query image class but they are most similar to the query image in terms of color features (Fig. 7(b)).

Whereas, when taking a weighting parameter and correction of weights into account, as can be seen in Fig. 7(c), with a greater weight assigned to texture features, out of 10 images retrieved, eight are from the query image class. This experiment indicates the effective role of the texture in the detection of objects. Hence, by considering an innovative algorithm, weighted combinations of texture and color features, and assigning a greater weight to texture features in this paper, the precision of the system has significantly improved.

**Table 7** Evaluation the Recall of the proposed method based on quantization of color histogram on corel-10k dataset

$w_r$	$w_c$	H=1 Bin=256	H=2 Bin=128	H=4 Bin=64	H=8 Bin=32	H=16 Bin=16
0.1	0.9	0.047734	0.048528	0.049966	0.051998	0.053971
0.2	0.8	0.049921	0.051293	0.053024	0.054374	0.055081
0.3	0.7	0.051754	0.053305	0.054523	0.055051	0.054573
0.4	0.6	0.053388	0.054583	0.055252	0.054992	0.053782
0.5	0.5	0.054664	0.055468	0.055383	0.054346	0.05294
0.6	0.4	0.05571	0.055868	0.054979	0.053571	0.052254
0.7	0.3	0.056312	0.055674	0.054128	0.052648	0.051571
0.8	0.2	0.056188	0.05468	0.05299	0.051764	0.051015
0.9	0.1	0.054549	0.052835	0.051623	0.050907	0.050579

**Table 8** Evaluation the precision of the proposed method based on the similarity measurement criterion on the corel-10k database

$w_r$	$w_c$	Euclidean	Chebyshev	Soergel	Chi-Square	Manhatan (L1)	Canberra	Extended Canberra
0.1	0.9	0.3735	0.29498	0.42257	0.43409	0.42253	0.47734	0.44523
0.2	0.8	0.37456	0.29884	0.41863	0.43399	0.42273	0.49921	0.44624
0.3	0.7	0.3754	0.30262	0.41112	0.43395	0.42287	0.51754	0.44732
0.4	0.6	0.37658	0.30728	0.39813	0.43395	0.42276	0.53388	0.44844
0.5	0.5	0.37724	0.312	0.38116	0.434	0.42269	0.54664	0.4499
0.6	0.4	0.37769	0.31688	0.35969	0.43413	0.42291	0.5571	0.45153
0.7	0.3	0.37636	0.32201	0.3325	0.43393	0.42314	<b>0.56312</b>	0.45163
0.8	0.2	0.37027	0.32236	0.29823	0.43384	0.4218	0.56188	0.44828
0.9	0.1	0.34816	0.31387	0.25844	0.43144	0.41458	0.54549	0.43189

Figure 8 shows the validation of the performance of using a weighting parameter in the improved precision of an image retrieval system on Corel-10k image dataset. The vertical axis of this graph shows the number of retrieved images without using a weighting parameter and the horizontal axis the precision of the retrieval system. The green part of the bar graphs shows the number of images whose retrieval precision has been increased using a weighting parameter, the red part the number of images whose precision has been reduced using a weighting parameter, and the blue part the number of images whose precision has not changed in the method, where a weighting parameter was used, compared with that in the method, where no weighting parameter was used.

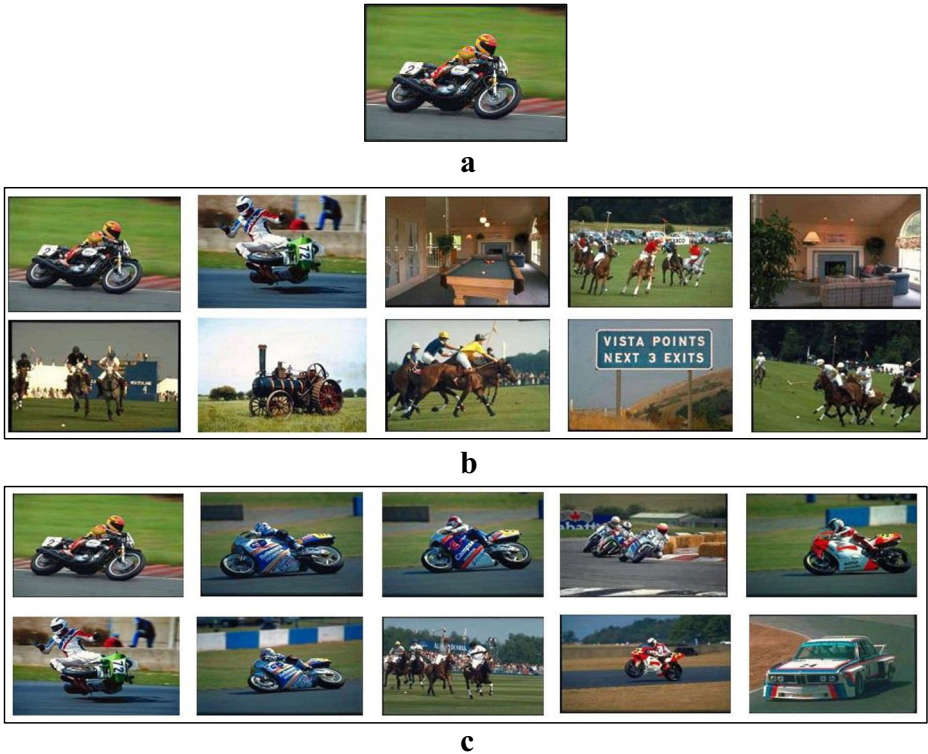
For example, a bar graph with a precision of 80 % shows the total number of images retrieved without using a weighting parameter, in which 251 images had a precision higher than 80 % using a weighting parameter, 199 had a precision lower than 80 % using a weighting parameter, and the precision of 311 images has not changed. As can be seen in the green sections of bars in the graph, the number of images whose retrieval precision has been increased using a weighting parameter is higher than those whose precision has been reduced, which is why the precision of the image retrieval system has improved.

### 5.5 Comparison with state-of-the-art methods

So far, many different methods have been proposed for CBIR. At this stage, the proposed CBIR system is compared with several efficient methods in this domain, and the precision and

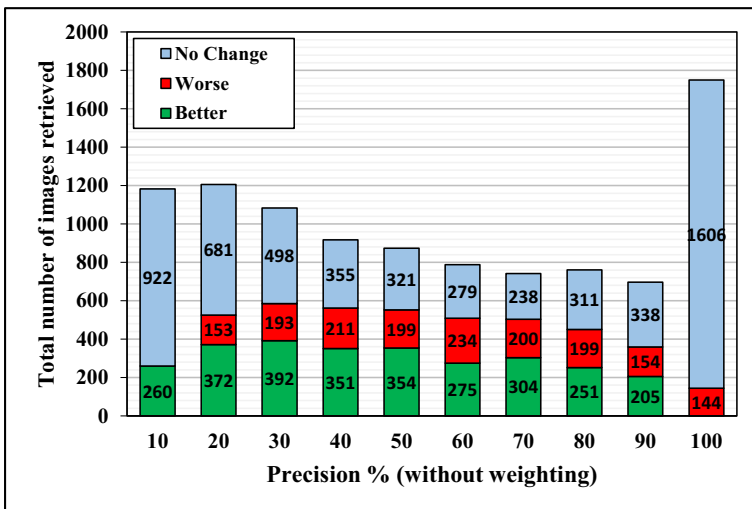
**Table 9** Evaluation the Recall of the proposed method based on the similarity measurement criterion on the corel-10k database

$w_r$	$w_c$	Euclidean	Chebyshev	Soergel	Chi-Square	Manhatan (L1)	Canberra	Extended Canberra
0.1	0.9	0.03735	0.029498	0.042257	0.043409	0.042253	0.047734	0.044523
0.2	0.8	0.037456	0.029884	0.041863	0.043399	0.042273	0.049921	0.044624
0.3	0.7	0.03754	0.030262	0.041112	0.043395	0.042287	0.051754	0.044732
0.4	0.6	0.037658	0.030728	0.039813	0.043395	0.042276	0.053388	0.044844
0.5	0.5	0.037724	0.0312	0.038116	0.0434	0.042269	0.054664	0.04499
0.6	0.4	0.037769	0.031688	0.035969	0.043413	0.042291	0.05571	0.045153
0.7	0.3	0.037636	0.032201	0.03325	0.043393	0.042314	0.056312	0.045163
0.8	0.2	0.037027	0.032236	0.029823	0.043384	0.04218	0.056188	0.044828
0.9	0.1	0.034816	0.031387	0.025844	0.043144	0.041458	0.054549	0.043189



**Fig. 7** An example of image retrieval by the proposed algorithm – (a) : query image – (b) : The retrieved images correspond to the query image (a) without considering the weighting criterion – (c) : The retrieved images correspond to the query image (a) with considering the weighting criterion

recall rates for the two databases Corel-1k and Corel-10k are evaluated. In [20] authors Used the color difference histogram (CDH) in the HSV color space. In [12] a new feature descriptor; namely, a color volume histogram, is proposed for image representation and content-based



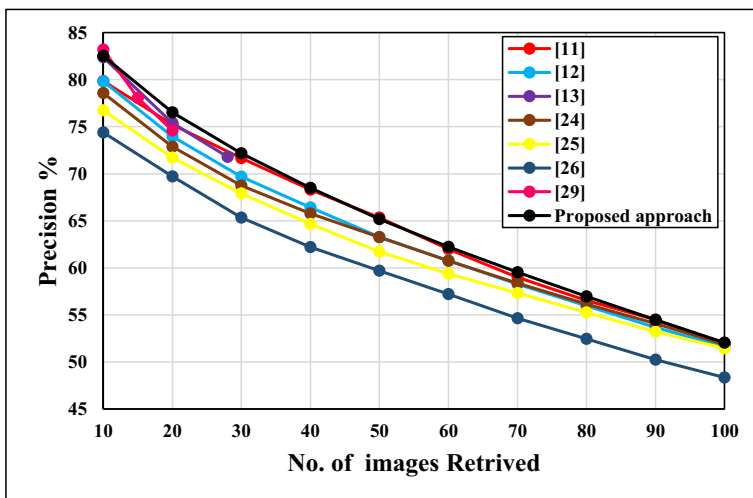
**Fig. 8** Validation of the performance of using the weighting criterion in the proposed CBIR system



**Table 10** Comparison of precision and recall rates of different CBIR systems on corel-1k dataset

Dataset	Performance	Methods									
		[5]	[4]	[22]	[30]	[26]	[10]	[17]	[31]	[9]	Proposed approach
Corel-1k	Precision (%)	79.86	79.83	82.4	78.58	71.60	74.39	75.38	83.20	69.7	82.52
	Recall (%)	7.98	7.98	8.24	7.85	7.16	7.43	7.53	8.32	6.97	8.25

image retrieval. In [5] presented a novel approach towards content based image retrieval by exture descriptor is named Diagonally Symmetric Local Binary Co-occurrence and color descriptor focuses on capturing the inter-channel relationship between the H and S channels of the HSV color space by quantizing the H channel into bins and voting with Saturation value and replicating the process for the S channel. In [4], a combination of local texton XOR patterns and color histograms has been used in CBIR. In [22] a new feature descriptor, named correlated primary visual texton histogram features (CPV-THF), for image retrieval is proposed. In [29], a CBIR method has been presented using a combination of a local binary pattern (LBP) and a local neighborhood difference pattern (LNDP). In [29], both features: LBP and LNDP, have been combined together so that they can extract the largest amount of information using local intensity differences. In [30], first, a local extrema pattern (LEP) was applied to determine the local information of the image. Then a co-occurrence matrix was used to obtain the co-occurrence of the pixels of the LEP map. And finally, a combination of the feature vector of a local extrema co-occurrence pattern (LECoP) and a color histogram was used for CBIR. In [26], a combination of uniform local binary patterns for color images (LBPC), uniform local binary patterns of the hue (LBPH), and a color histogram has been used. In [10], a local neighborhood-based robust color occurrence descriptor (LCOD) has been used, which extracted color occurrences and displayed them in a binary form. In [17], Local Tetra patterns (LTrPs) have been used, which encrypted the relationship between the referenced pixel and its neighbors based on paths calculated using first-order derivatives in vertical and horizontal directions. In [9] presented a novel multi-feature fusion image retrieval method.



**Fig. 9** Precision curve with number of images retrieved for Corel-1 K database

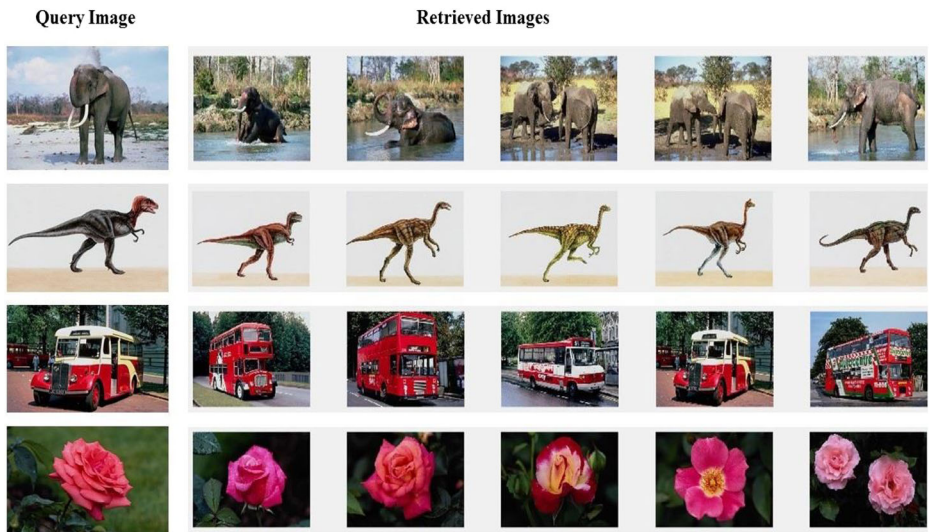


Fig. 10 Query image and retrieved images from Corel-1 K dataset

In the feature extraction process, the number of loop iterations of the traditional NMI feature is difficult to determine. Therefore, this paper proposed an improved PCNN to extract NMI feature. In [31] proposes a lightweight and efficient image retrieval approach by combining region and orientation correlation descriptors (CROCD). The region color correlation pattern and orientation color correlation pattern are extracted by the region descriptor and the orientation descriptor, respectively. In [7] a novel and highly simple but efficient representation based on the multi-integration features model, called the MIFH, is proposed for CBIR using the feature integration theory. In Sections 5.5.1 and 5.5.2, the performance of the proposed method is compared with methods introduced on Corel-1k and Corel-10k image datasets, and the results are presented.

### 5.5.1 Comparison on Corel-1k dataset

As shown in Table 10, on Corel-1k dataset, the precision and recall rates of the proposed method are higher than those of other methods. The number of images retrieved in this experiment is N = 10.

In [31], although the precision of the proposed method is higher than our method, but in the number of retrievals higher than 10 images, our method performs much better. According to the diagram in Fig. 9.

Table 11 Comparison of precision and recall rates of different CBIR systems on corel-10k dataset

Dataset	Performance	Methods											
		[20]	[12]	[5]	[4]	[22]	[29]	[30]	[26]	[10]	[7]	[31]	Proposed approach
Corel-10k	Precision (%)	50.67	48.58	53.72	51.05	54.86	42.75	52.50	50.6	37.02	52.96	44.51	56.31
	Recall (%)	5.06	4.85	5.37	5.10	5.48	4.27	5.25	5.06	3.7	5.29	4.45	5.63

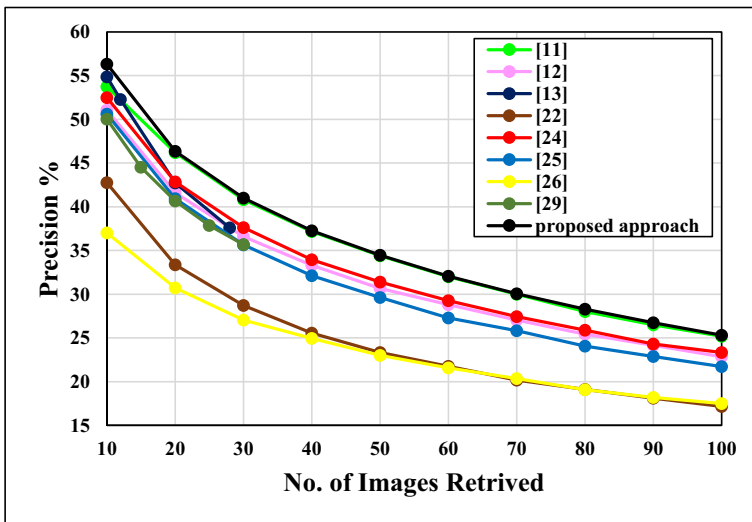


Fig. 11 Precision curve with number of images retrieved for Corel-10 K database

Figure 9 shows precision graph with different numbers of retrieved images for this dataset. For the experiment, first, we retrieved 10 images and then increased the number of retrieved images to 100 simultaneously in steps of 10 images (Fig. 10).

### 5.5.2 Comparison on Corel-10k dataset

As shown in Table 11, on Corel-10k dataset, the precision and recall rates of the proposed method are higher than those of other methods. The number of images retrieved in this experiment is  $N = 10$ .

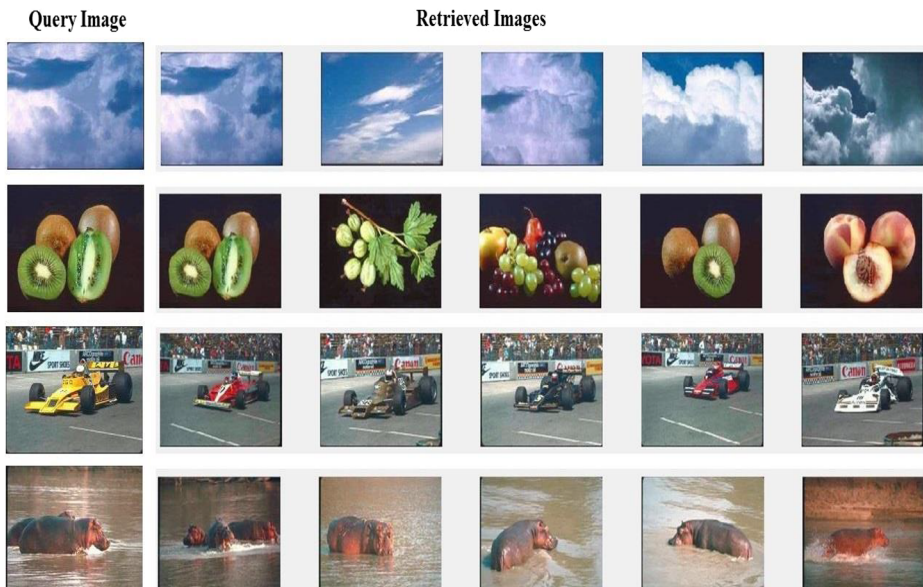


Fig. 12 Query image and retrieved images from Corel-10 K dataset

**Table 12** Comparison of precision and recall rates of different CBIR systems with deep learning on corel-1k

Dataset	Performance	Methods			
		ELM [8]	LetNet [15]	MatConvNet [13]	Proposed approach
Corel-1k	Precision (%)	90	79.62	79	82.52
	Recall (%)	9.00	7.96	7.9	8.25

Figure 11 show precision graph with different numbers of retrieved images for this dataset.

In Fig. 12, the query image is at the beginning of the row and the remaining images show images retrieved for each query image on Corel-10k dataset.

### 5.5.3 Comparison with deep learning methods

In this section, the proposed method is compared with image retrieval methods based on deep learning on Corel-1k and Corel-10k datasets. The number of images retrieved in this experiment is  $N = 10$ .

All the References mentioned in Tables 12 and 13 include the disadvantages of deep learning methods such as the need for high data volume for training and strong hardware. They also have a high computational load and low data processing speed.

For example in [8], One of the problems ELM Classifier is that you can train them really fast, But you pay for it by having very slow evaluation. For most applications, evaluation speed is more important than training speed. In [15] (AL representation) the authors use a combination of features of two models of deep convolutional networks. In this system, the Alexnet model and the improved LeNet model are train separately on the database and used a combination of features of both models for retrieval. The retrieval precision obtained by this model on the Corel database is 0.948. Although this system is more precision than our method, it requires more processing time due to the use of two convolutional networks together. Also, this accuracy is reported only on one dataset. In [32] ResNet-152 model, in addition to not achieving high accuracy compared to other deep learning methods on the Corel-10k dataset, also has a high dimension of the feature vector.

In References [15] (LetNet-L (F6)), [13], [32] (VGGNet-16 model) and [32] (GoogLeNet model) show that the precision and recall rate of our method is higher than these methods. This is while our method is done using a lower volume of data and without the need for strong hardware and classifier.

## 6 Conclusions

The main objective of this paper was to present a content-based image retrieval (CBIR) system. In this regard, an approach was presented, which included two stages. At the feature

**Table 13** Comparison of precision and recall rates of different CBIR systems with deep learning on corel-10k

Dataset	Performance	Methods				
		ELM [8]	ResNet-152 [32]	VGGNet-16 [32]	GoogLeNet [32]	Proposed approach
Corel-10k	Precision (%)	72	61.7	55.2	51.7	56.31
	Recall (%)	7.20	6.17	5.52	5.17	5.63

extraction stage, texture features were obtained from the combination of the operators  $MLBP_{P,R}$  and LNDP, Gaussian features, co-occurrence matrices, and statistical features, and color features were obtained from a color histogram as a color information extractor. In this regard, the input color image was divided into three separate channels; namely, H, S, and V, and a color histogram was extracted from each channel. The most efficient similarity/distance measure with respect to the weight of the extracted features was used at the similarity matching stage. Considering examinations done, the best weighting parameter was assigned to each of the intervals obtained based on the components: texture and color. Through multiplying these parameters by the intervals and then adding the values obtained, the final distance between the query image and the images in the database was obtained. In the last step, the extracted data were classified after the calculation of the final distance, and from among them, N images with higher similarity to the query image (with less distance) were selected and displayed. The most important advantage of the proposed system is the use of weighting parameters, which increases the accuracy of a CBIR system in retrieving images similar to the query image by assigning greater weights to features that are more effective. Rotation invariance and greyscale invariance are among the advantages of the proposed CBIR system, which is due to the use of co-occurrence matrices and general statistical features. Low noise sensitivity is another advantage of using global features and different levels of a Gaussian filter. The results indicate the high performance of the proposed CBIR system compared with those of some efficient methods in this field. The proposed approach is a general CBIR system, which can be used for a variety of images. Therefore, general usability can be considered as another advantage of the proposed approach. The reduced number of extracted features may reduce retrieval time. Therefore, one of the future tasks to be taken into consideration is providing a preprocessing stage to remove categories that are completely away from a query image. Since the weighting parameters in this paper were adjusted manually, developing a global genetic algorithm to optimize the weighting parameters may be considered as another future task.

## References

1. Afifi AJ, Ashour WM (2012) Image retrieval based on content using color feature. *ISRN Comput Graph* 2012:1–11. <https://doi.org/10.5402/2012/248285>
2. Alrahhal M, Supreethi K (2020) Multimedia image retrieval system by combining CNN with handcraft features in three different similarity measures. *Int J Comput Vis Image Process (IJCVIP)* 10(1):1–23
3. An J, Lee SH, Cho NI (2014) Content-based image retrieval using color features of salient regions. In: 2014 IEEE International Conference on Image Processing (ICIP). IEEE, New York, pp 3042–3046
4. Bala A, Kaur T (2016) Local texton XOR patterns: A new feature descriptor for content-based image retrieval. *Eng Sci Technol Int J* 19(1):101–112. <https://doi.org/10.1016/j.jestch.2015.06.008>
5. Bhunia AK, Bhattacharyya A, Banerjee P, Roy PP, Murala S (2019) A novel feature descriptor for image retrieval by combining modified color histogram and diagonally symmetric co-occurrence texture pattern. *Pattern Anal Applic* :1–21
6. Boukerma R, Bougueroua S, Boucheham BA (2019) Local patterns weighting approach for optimizing content-based image retrieval using a differential evolution algorithm. In: 2019 International Conference on Theoretical and Applicative Aspects of Computer Science (ICTAACS), pp 1–8
7. Chu K, Liu G-H (2020) Image retrieval based on a multi-integration features model. *Math Probl Eng* 2020: 1–10. <https://doi.org/10.1155/2020/1461459>
8. Dhingra S, Bansal P (2018) A competent and novel approach of designing an intelligent image retrieval system. *ICST Trans Scalable Inf Syst*. <https://doi.org/10.4108/eai.13-7-2018.160538>
9. Du A, Wang L, Qin J (2019) Image retrieval based on colour and improved NMI texture features. *Automatika* 60(4):491–499. <https://doi.org/10.1080/00051144.2019.1645977>

10. Dubey SR, Singh SK, Kumar Singh R (2015) Local neighbourhood-based robust colour occurrence descriptor for colour image retrieval. *IET Image Proc* 9(7):578–586. <https://doi.org/10.1049/iet-ipr.2014.0769>
11. Haralick RM, Shanmugam K, Dinstein IH (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern* (6):610–621
12. Hua J-Z, Liu G-H, Song S-X (2019) Content-Based Image Retrieval Using Color Volume Histograms. *Int J Pattern Recognit Artif Intell* 33(11). <https://doi.org/10.1142/s021800141940010x>
13. Huang W-q, Wu Q (2017) Image retrieval algorithm based on convolutional neural network. In: *Current Trends in Computer Science and Mechanical Automation*, vol 1. Sciendo Migration, Warsaw, pp 304–314
14. Liu G-H, Yang J-Y (2013) Content-based image retrieval using color difference histogram. *Pattern Recogn* 46(1):188–198. <https://doi.org/10.1016/j.patcog.2012.06.001>
15. Liu H, Li B, Lv X, Huang Y (2017) Image retrieval using fused deep convolutional features. *Prog Comput Sci* 107:749–754. <https://doi.org/10.1016/j.procs.2017.03.159>
16. Mistry Y, Ingole DT, Ingole MD (2018) Content based image retrieval using hybrid features and various distance metric. *J Electr Syst Inf Technol* 5(3):874–888. <https://doi.org/10.1016/j.jesit.2016.12.009>
17. Murala S, Maheshwari RP, Balasubramanian R (2012) Local tetra patterns: a new feature descriptor for content-based image retrieval. *IEEE Trans Image Process* 21(5):2874–2886. <https://doi.org/10.1109/TIP.2012.2188809>
18. Mustikasari M, Madenda S, Prasetyo E, Kerami D, Harmanto S (2014) Content based image retrieval using local color histogram. *Int J Eng Res* 3(8):507–511
19. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24(7):971–987
20. Qazanfari H, Hassanpour H, Qazanfari K (2019) Content-based image retrieval using HSV color space features. *Int J Comput Inf Eng* 13(10):537–545
21. Raju USN, Suresh Kumar K, Haran P, Boppana RS, Kumar N (2019) Content-based image retrieval using local texture features in distributed environment. *Int J Wavelets Multiresolut Inf Process* 18(01). <https://doi.org/10.1142/s0219691319410017>
22. Raza A, Dawood H, Dawood H, Shabbir S, Mehboob R, Banjar A (2018) Correlated primary visual Texton histogram features for content base image retrieval. *IEEE Access* 6:46595–46616. <https://doi.org/10.1109/access.2018.2866091>
23. Sathiamoorthy S, Natarajan M (2020) An efficient content based image retrieval using enhanced multi-trend structure descriptor. *SN Appl Sci* 2(2). <https://doi.org/10.1007/s42452-020-1941-y>
24. Shih J-L, Chen L-H (2002) Color image retrieval based on primitives of color moments. In: *IEE Proceedings- Vision, Image and Signal Processing* 149(6):88–94. <https://doi.org/10.1049/ip-vis:20020614>
25. Shikha B, Gitanjali P, Kumar P (2020) An extreme learning machine-relevance feedback framework for enhancing the accuracy of a hybrid image retrieval system. *Int J Interactive Multimed Artif Intell* 6(2). <https://doi.org/10.9781/ijimai.2020.01.002>
26. Singh C, Walia E, Kaur KP (2018) Color texture description with novel local binary patterns for effective image retrieval. *Pattern Recogn* 76:50–68. <https://doi.org/10.1016/j.patcog.2017.10.021>
27. Srivastava P, Prakash O, Khare A (2014) Content-based image retrieval using moments of wavelet transform. In: *The 2014 International Conference on Control, Automation and Information Sciences (ICCAIS 2014)*. IEEE, New Year, pp 159–164
28. Varga D, Szirányi T (2016) Fast content-based image retrieval using convolutional neural network and hash function. In: *2016 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, New York, pp 002636–002640
29. Verma M, Raman B (2017) Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval. *Multimed Tools Appl* 77(10):11843–11866. <https://doi.org/10.1007/s11042-017-4834-3>
30. Verma M, Raman B, Murala S (2015) Local extrema co-occurrence pattern for color and texture image retrieval. *Neurocomputing* 165:255–269. <https://doi.org/10.1016/j.neucom.2015.03.015>
31. Xie G, Huang Z, Guo B, Zheng Y, Yan Y (2020) Image retrieval based on the combination of region and orientation correlation descriptors. *J Sens* 2020:1–15. <https://doi.org/10.1155/2020/6068759>
32. Zhang F, Zhong, B-J (2016) Image retrieval based on fused CNN features. *DEStech Transactions on Computer Science and Engineering (aics)*