IET Image Processing

Research Article

IET Journals
The Institution of Engineering and Technology

Texture and colour region separation based image retrieval using probability annular histogram and weighted similarity matching scheme

ISSN 1751-9659
Received on 10th December 2018
Revised 6th November 2019
Accepted on 9th December 2019
E-First on 27th April 2020
doi: 10.1049/iet-ipr.2018.6619
www.ietdl.org

Jitesh Pradhan[1] ✉, *Sumit Kumar*[1], *Arup Kumar Pal*[1], *Haider Banka*[1]

[1]*Department of Computer Science and Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand- 826004, India*

✉ *E-mail: jitpradhan02@gmail.com*

Abstract: Content-based image retrieval (CBIR) uses primitive image features for retrieval of similar images from a dataset. Generally, researchers extract these visual features from the whole image. Therefore, the extracted features contain overlapped information of texture, colour, and shape features, and it is a critical challenge in the field of CBIR. This problem can be overcome by extracting the colour features from the colour as well as shape and texture features from the intensity dominant part only. In this study, the authors have proposed an iterative algorithm to separate colour and texture dominant part of the image into two different images. Here, a combination of edge maps and gradients has been used to achieve separate colour and texture images. Further, scale-invariant feature transform and 2D dual-tree complex wavelet transform has been realised to extract unique shape and texture features from the texture image. Simultaneously, a probability-based semantic centred annular histogram has been suggested to extract unique colour features from the colour image. Finally, a novel weighted distance-based feature comparison scheme has been proposed for similarity matching and retrieval. All the image retrieval experiments have been carried out on seven standard datasets and demonstrated significant improvements over other state-of-arts CBIR systems

1 Introduction

1.1 Background

In today's world, an exponential growth has been seen on the use of multimedia devices in the daily life. As a result, multimedia data have been increased drastically on the Internet. These all multimedia data is flooding over the digital repositories. So, handling of these gigantic web repositories is a serious issue since enormous daily life applications use these digital web repositories for multimedia data retrieval. A significantly large portion of these infinitely growing multimedia data has been captured by digital images. Thus, retrieval of similar images from these huge digital repositories is an even more challenging task. In particular, this problem is known as image retrieval problem. Traditionally, contemporary researchers have used tag-based image retrieval (TBIR) [1] systems to solve the above problem. In TBIR systems, a specific keyword (i.e. tag) has been assigned to each database image which was a very monotonous and tedious process. Further, these assigned keywords have been used for image retrieval process. A TBIR system needs human involvement throughout the retrieval process which is a major flaw of this system. At the same time, a single keyword is not sufficient to describe the overall meaning of the image which is another critical flaw of TBIR system. These problems have been solved using the content-based image retrieval system (CBIR) [2]. A CBIR system uses original image visual features for retrieval of the similar image from the digital repositories. It employs different feature extraction approaches to capture the actual image features in form of a feature vector. These primitive image visual features are colour, texture, and shape features of the image. CBIR uses automatic feature extraction and features comparison methodology for image retrieval which ultimately overcomes the limitations of the TBIR systems.

1.2 Motivation and contributions

Every image contains overlapped colour, texture, and shape features and most of the CBIR methods extract all these features directly from the whole image. As a result, the extracted visual features also contain overlapped colour, texture, and shape feature information which degrades the feature quality and retrieval performance. Texture features must be extracted from the texture dominated portion of the image so as colour and shape feature also. Inspired by these problems, we have proposed an iterative process to separate the colour and texture dominated regions of the image into two different images. Generally, different images can have different proportion of the colour and texture feature. Simultaneously, texture and colour features also depend on the human visual perception which makes it very difficult to separate the colour and texture regions. Hence, we have suggested a novel approach for colour and texture region separation so the user can separate colour and texture regions based on their visual perception. Further, we have used both of these resultant images for feature extraction process. All natural images show huge diversity in colour distribution in local image regions. In such cases, use of normal colour histogram will not be efficient to extract suitable colour information since it only captures the overall intensity distribution. Thus, we have divided the image into annular regions and computed the histograms to capture the local diversity of the colour information. Further, the probability-based histogram has been computed and this approach evenly distributes the colour features among all bins. At the same time, texture dominated regions of an image possesses extreme heterogeneous local pixel sub-regions. So, the structural and textural features of the image abruptly changes in the different local sub-regions. To address this problem, an efficient structural and multi-directional texture feature extraction scheme is direly needed. Hence, we have used scale-invariant feature transform (SIFT) and two-dimensional dual-tree complex wavelet transform (2D DT-CWT) to capture the local structural and multi-directional texture features. The SIFT will locate and extract the most salient and unique structural behaviour of the image. Parallelly, 2D DT-CWT will analyse the textural patterns in six different directions.

Another problem associated with the CBIR is the straightforward similarity measurement procedure. Most of the researchers combine all the extracted primitive visual features into

a single one-dimensional feature vector and find the similarity between different images. But, every image has different proportion of colour, texture, and shape features, so we cannot give equal importance to all three visual features. In this paper, we have also proposed a weighted similarity matching scheme which assigns the different weights to the different image features based on their quality of information as a result it overcomes the problems of straightforward similarity matching process. The major contributions of the authors have been listed below:

- A novel colour and texture region separation scheme based on the edge maps and image gradients has been proposed in this paper.
- Colour features have been extracted from the colour region only. So, shape and textural features have been extracted from the intensity dominated region of the image.
- A novel probability-based semantic centred annular colour histogram has been proposed for efficient colour feature analysis.
- SIFT and 2D DT-CWT schemes have been employed for unique structural and multi-directional texture feature analysis.
- A new weighted similarity measurement scheme has been used for feature matching process.
- Significant retrieval experiments have been performed on different natural, object, and texture image datasets to validate the preciseness and robustness of the proposed CBIR system.

1.3 Related works

Among all three primitive visual features, colour features are the most frequently used by researchers for CBIR applications since it is rotational, translation, and scale invariant visual feature. Conventionally, researchers have used only statistical colour features in CBIR process which is colour histogram [3]. Colour histogram simply gives the statistical analysis of the intensity distribution without considering the spatial positions. Huang *et al*. [4] have proposed colour auto-correlograms to solve the above problem. Colour auto-correlograms deal with the spatial position of colour also since it gives the probability of finding different colour pairs at a fixed distance. Joint histograms, spatiograms, colour difference histograms (CDH), scalable colour descriptor and colour structure descriptor [5, 6] are other colour feature extraction processes which are popular among the researchers.

Texture feature is another most commonly used image visual feature in different CBIR applications. Textures are the different directional patterns present on the surface of the image. So, common directional analysis has been carried out in order to extract the texture features of an image. Discrete wavelet transform (DWT) [5] is the simplest way to analyse the textural patterns in three different directions (i.e. vertical, horizontal, and diagonal). Edge histogram descriptor and grey-level co-occurrence matrix [7] are the two commonly used static texture feature extraction process. Block variation of local correlation coefficients (BVLC) and block difference of inverse probabilities (BDIP) [8] are two popular spatial texture feature extraction schemes. Both of these schemes deal with the smoothness of local textural patterns and variations in the local brightness. Wang *et al*. [9] have used colour co-occurrence matrix to retrieve coloured texture feature of images. This approach gives the colour information along with the texture correlation features. However, this approach does not consider the directional as well as structural texture features. Liu *et al*. [10] have combined the merits of colour co-occurrence matrix and colour histograms together to propose a novel multi-texton histogram (MTH) for image retrieval applications. This scheme also combines the texture feature along with the colour features in the final feature vector. But, this scheme also does not consider directional as well as structural texture features. At a recent time, tetrolet transform [11] has been used by the contemporary researchers to capture the local structural and geometrical features of the image. Tetrolet transform analyses 117 different possible local geometrical structural patterns of the texture.

Apart from the colour and texture features, shape features [12] of the image are also used in different CBIR application. Canny, Sobel, and Fuzzy difference based edge detection schemes [13] are most commonly used to detect the boundaries of the image. Generally, a shape feature alone does not perform well in the image retrieval process. So, usually, researchers use shape features in combination with the other primitive features [14] for better retrieval performance. Chun *et al*. [8] have used BVLC and BDIP for multi-resolution texture analysis and colour auto-correlograms for colour feature analysis. Further, they have combined the texture and colour features for the retrieval process. Next, Liu *et al*. [15] have constructed a micro-structure descriptor (MSD) based new CBIR scheme from edge orientation and quantised HSV colour information. Later, Liu and Yang. [16] have also proposed a new CDH scheme based CBIR approach which gives the combined colour and edge feature information. Next, Zeng *et al*. [6] have proposed the colour spatiograms (CS) in combination with Gaussian mixture model for colour quantisation (GMM+CQ) for feature extraction to improve the retrieval efficiency. Later, Varish *et al*. [17] have proposed a two-level hierarchical CBIR scheme which extracts colour and texture features in different levels of hierarchy. They have used non-uniform bin-based HSV colour histograms and 2D dual tree complex wavelet transform (NCH +CWF) for colour and texture feature extraction. Later, Pradhan *et al*. [18] have employed principal texture direction based approach to rearrange the image and constructed colour edge features (PTD-IR+CEF) to perform CBIR applications. Recently, Pradhan *et al*. [19] have proposed a three-level hierarchical CBIR system which uses texture, colour, and shape features in different level of hierarchy. They have used tetrolet transform, joint edge histogram, and colour channel correlation histogram (TT+JEH+CCCH) for texture, shape, and colour feature extraction.

Nowadays, researches are also using different structural and spatial correlation visual features in different types of CBIR applications. Such as, Wang and Wang [20] have introduced a structure elements descriptor (SED) to address the local and structural features of the image. The authors have used SED to compute a structure elements histogram (SEH). Further, they have quantised the HSV colour plane of the image for colour feature extraction. Though, SEH reflects the texture as well as colour features all together. Later, Wang and Wang [21] have proposed a multi-factors correlation scheme for image retrieval which also integrates the benefits of the block truncation coding scheme (MFC-BTC). Initially, they have generated the bitmap and mean colour component images from the RGB images. Here, they have employed the block truncation coding to generate the colour component image. Next, they have extracted the image features using three correlations for final CBIR applications. These two discussed schemes address only the local and structural features of the image which is the main concern of these schemes. Recently, Unar *et al*. [22] have analysed the textual and visual characteristics for textual image retrieval. In their work, the authors have identified the salient regions of the image to locate and recognise the textual regions of the image. Though, they have extracted and created separate feature vectors of textual and visual information. Finally, they have adopted a Kernel method for feature vector fusion and CBIR application. But, this scheme is specifically designed for only textual image retrievals.

All the above discussed primitive visual features are considered as the low-level image features. Other than these low-level image features researchers have also used some high-level image features which directly portrays the overall semantic meaning of the image. Semantic templates, object detection, classification and bag-of-visual-words [23] based approaches are the most common high-level semantic based CBIR schemes. All these image semantic based CBIR schemes bear higher time and space complexity, which are the major limitations of these types of systems. Recently, deep learning and convolution neural network based CBIR systems [24] are gaining popularity among the researchers. But, these types of CBIR systems take very long time in the learning process and the dimensions of the feature set used for image classification is also enormously higher than the straightforward CBIR process. Recently, Meng *et al*. [25] have used a regional convolution mapping feature scheme along with convolution neural network (CNN) based approach for feature extraction and image classification in CBIR applications. These CNN based schemes work in two stages, initially these schemes perform image

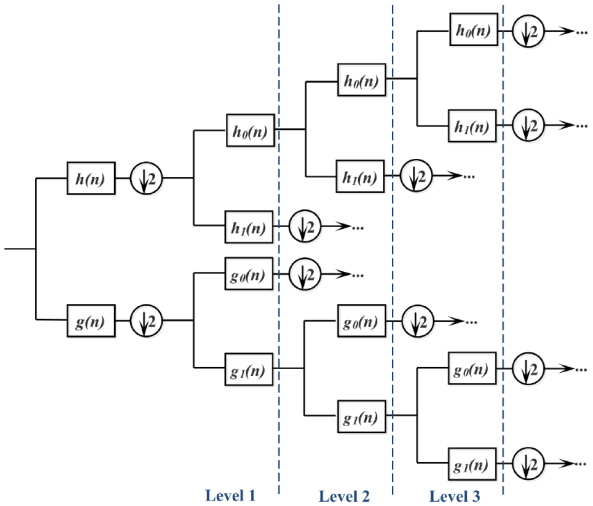

Fig. 1 *Low-pass and high-pass filter set arrangement in two DWT branches of 2D DT-CWT for three levels of decompositions*

classification task. Next, relevant images have been retrieved from the identified image class only. So, the performance of these schemes solely depends on the classification accuracy. These schemes give 100% precision in retrieval for true classification. At the same time, these schemes do not retrieve any relevant images for miss classification which is the major problem. Apart from these, it also requires high amount of annotated training images and database image annotation is another major issue in itself. Whereas our proposed scheme works well for all kinds of datasets and it does not depend on the image dataset size. Our proposed scheme also works well for the unannotated image datasets. At the same time, it also gives significant amount of relevant images for every retrieval outcome. To show the acceptability and robustness of the proposed scheme, we have also compared the retrieval performance of the proposed scheme with well-known VGG-16 CNN [26] based CBIR scheme.

Apart from all the primitive visual features, nowadays researches are also considering different image coding and compression schemes in image retrieval applications [27–29]. Such as, Liu and Wechsler [27] have introduced enhanced fisher linear discriminant and probabilistic reasoning model to retrieve face images from a large image dataset. Later, Wang and Chen. [29] have used the fast fractal encoding approach for image feature extraction and CBIR applications. In 2018, Wang *et al*. [30] have introduced a quaternion polar harmonic Fourier moments for object recognition and image reconstruction. This scheme can effectively extract invariant image features which will improve the retrieval efficiency. Recently, Wang *et al*. [31] have used ternary radial harmonic Fourier moments (TRHFM) for image zero-watermarking. Their approach can deal with the symmetric and asymmetric attacks. Though, the TRHFM technique will be also effective for secure content based image retrial applications. This study shows that other imaging tools will be also useful for improving the retrieval efficiency as well as the image security.

2 Methodologies

In this paper, the authors have used a combination of SIFT and multi-resolution 2D DT-CWT for structural and textural feature analysis. So, in this section, we have briefly presented the SIFT and 2D DT-CWT schemes of structural and textural feature analysis.

2.1 Scale-invariant feature transform

Lowe [32] have introduced the SIFT for object recognition and matching. The SIFT features are rotation and scale invariant hence it is widely used for shape feature based object detection and matching. Lowe has translated the given image into a local coordinate system which is scale, rotation, and translation invariant. The SIFT features have been constructed trough four different stages. Initially, scale-space extrema detection has been performed by searching stable features in different scale images. Next, key-point localisation has been employed by detecting the minima and maxima of difference in a 3×3 local neighbourhood blocks. Further, a local gradient direction histogram has been constructed for different scale images and assigned a canonical orientation at peak of the smoothed histogram. Finally, key-point descriptor has been generated based on the local image gradients of different scale image with different rotations. The detailed steps of key point extraction and descriptor representation can be found in [33]. The feature locality, extensibility, efficiency, distinctiveness, and quantity are the major advantages of the SIFT features.

2.2 Dual-tree complex wavelet transforms

Selesnick *et al*. [34] have proposed a new texture analysis scheme to overcome the limitations of the conventional DWT [5]. This new scheme performs textural analysis in six different directions which are $\pm 15°$, $\pm 45°$, and $\pm 75°$. This new scheme uses two different real DWT to form a tree-like structure and called as dual-tree complex wavelet transforms (DT-CWT) [17, 34]. DT-CWT uses set of different low- and high-pass filters for the two real DWTs. The first DWT is considered as the real part whereas the second DWT is considered as the complex part of the transformation. In every level of decomposition, it produces six directional coefficients and two approximations coefficients.

Let $f(r, c)$ is an input image and 2D DT-CWT uses a complex function along with the six complex functions to decompose the input image into low-pass and high-pass directional sub-bands. Let h and g are the high-pass and low-pass filter sets for both real DWTs. Here $h_0(n)$ and $g_0(n)$ are the low-pass filter sets whereas $h_1(n)$ and $g_1(n)$ are the high-pass filter sets for the two DWTs. Based on these assumptions, the 2D DT-CWT function has been defined as follows:

$$f(r, c) = f(r) \times f(c) \tag{1}$$

$$f(x) = f_h(x) + j \times f_g(x) : x = r \text{ or } c \tag{2}$$

where $f_h(x)$ is the real part and $f_g(x)$ is the imaginary part of complex wavelet transform. So, the function for the 2D DT-CWT has been defined as follows:

$$f(r, c) = \{f_h(r) + j \times f_g(r)\}\{f_h(c) + j \times f_g(c)\} \tag{3}$$

$$\begin{aligned} f(r, c) \quad = \quad & \{f_h(r) f_h(c) - f_g(r) f_g(c)\} + j \\ & \times \{f_h(r) f_g(c) + f_g(r) f_h(c)\} \end{aligned} \tag{4}$$

where $\text{Real}(f(r, c)) = \{f_h(r) f_h(c) - f_g(r) f_g(c)\}$ and $\text{Imaginary}(f(r, c)) = \{f_h(r) f_g(c) + f_g(r) f_h(c)\}$. Fig. 1 demonstrates the different low-pass and high-pass filter set arrangements in two DWT tree branches of the DT-CWT.

In Fig. 1, we can see that the DT-CWT uses two sets of low-pass and high-pass filters. Here, any complex arithmetics are not required to implement these filters since all these filters are real [34]. In DT-CWT, the complex wavelet part is the Hilbert transform of the real wavelet part. The Hilbert transform is a phase transformation technique as a result the complex wavelet part extracts different directional texture features as compared to the real wavelet part.

3 Proposed CBIR scheme

In this paper, we have proposed a novel CBIR which uses all three primitive image visual features for image retrieval. Initially, we have employed the proposed iterative colour and texture region separation approach to discriminate the input image into colour and texture dominated salient components. Next, we have used a novel probability semantic centred annular histogram to extract the positional invariant colour features. This approach uses a cumulative probability based technique to reduce the overall colour feature vector length without affecting the colour feature quality.

Input: An RGB colored image I_j.
Output: Texture region image TF_j and color region image Cl_j.

1: Select an input image I_j.
2: Set a Counter variable $Ctr = 0$.
3: **For** $Ctr = 1 \rightarrow n$
4: Compute the gray-scale image G_j and saturation component S_j to calculate the intensity and saturation gradient image IG_j and SG_j as follows:

$$\frac{\partial}{\partial x}Q_j(x,y)) = \frac{1}{2}\left(Q_j(x+1,y) - Q_j(x-1,y)\right) \quad (5)$$

$$\frac{\partial}{\partial y}Q_j(x,y)) = \frac{1}{2}\left(Q_j(x,y+1) - Q_j(x,y-1)\right) \quad (6)$$

$$IQ_j(x,y) = \frac{\partial}{\partial x}Q_j(x,y)) + \frac{\partial}{\partial y}Q_j(x,y)) \quad (7)$$

where, $Q_j(x,y) = G_j(x,y)/S_j(x,y)$ and $IQ_j(x,y) = IG_j(x,y)/IS_j(x,y)$
5: Compute the combined gradient map $CG_j(x,y) = IG_j(x,y) + IS_j(x,y)$
6: Calculate the binary gradient map BG_j from the combined gradient map CG_j using improved Otsu thresholding method [34].
7: Compute the binary Sobel and fuzzy edge map Es_j and Ef_j of input image I_j.
8: Calculate the detailed edge map E_j as follows:

$$E_j(x,y) = \begin{cases} 1, & \text{If } Es_j == 1 \ \ OR \ \ Ef_j == 1 \\ 0, & \text{If } Es_j == 0 \ \ AND \ \ Ef_j == 0 \end{cases} \quad (8)$$

9: Calculate the binary texture image BT_j as follows:

$$BT_j(x,y) = \begin{cases} 1, & \text{If } E_j == 1 \ \ OR \ \ BG_j == 1 \\ 0, & \text{If } E_j == 0 \ \ AND \ \ BG_j == 0 \end{cases} \quad (9)$$

10: Calculate the original texture region T_j and color region Cl_j of the image as follows:

$$T_j^C(x,y) = \begin{cases} I_j^C(x,y), & \text{If } BT_j(x,y) == 1 \\ 0, & Otherwise \end{cases} \quad (10)$$

$$Cl_j^C(x,y) = \begin{cases} I_j^C(x,y), & \text{If } BT_j(x,y) == 0 \\ 0, & Otherwise \end{cases} \quad (11)$$

where, $C \in \{red, green, blue\}$ color components of the image I_j.
11: **if** $Ctr >= n$ **then**
| $Set : I_j = Cl_j$ and $TF_j = TF_j + T_j$
end
12: **End**

Fig. 2 *Algorithm 1: texture and colour separation algorithm*

Simultaneously, we have used SIFT and 2D DT-CWT approaches on extracted texture image to capture the unique structural properties and multi-directional textural features. Finally, we have used a weighted similarity scheme which assigns different weights to the different image features based on their importance. So, our proposed CBIR system works on three different stages which are (i) pre-processing: image texture and colour separation, (ii) feature extraction, and (iii) similarity matching and image retrieval. The proposed CBIR scheme has been explained in detail in the following subsections.

3.1 Pre-processing: image texture and colour separation

This is the first stage of the proposed CBIR system where we have decomposed the input image I_j into its colour dominated region and texture dominated regions. First a query image I_j has been selected. Next, the query image I_j has been converted into a grey-scale image G_j to calculate the intensity gradient image IG_j.

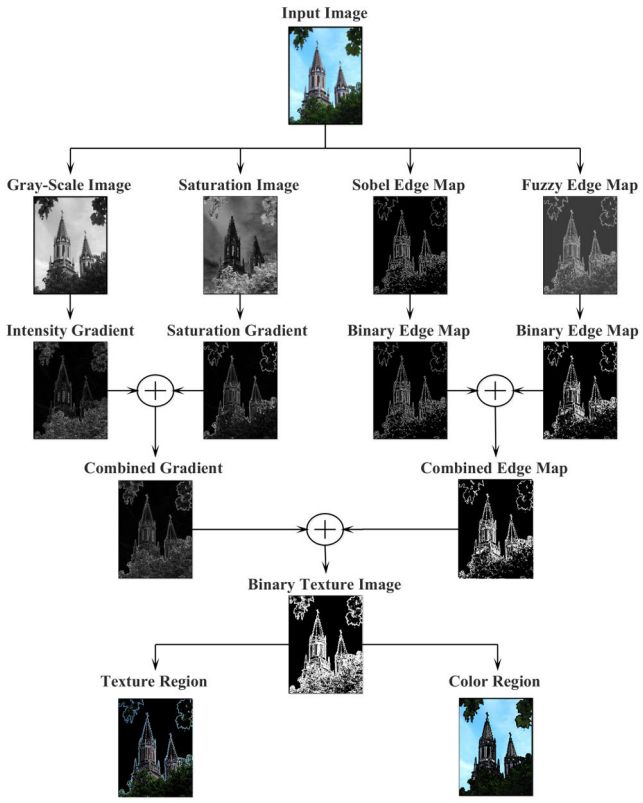

Fig. 3 *Schematic flow diagram of texture and colour region generation process for first iteration*

At the same time, we have converted the *RGB* query image into *HSV* image and extracted the saturation components S_j of the image to compute the saturation gradient SG_j of the input image. Simultaneously, we have extracted the Sobel operator [13] and fuzzy difference [13] operator based edge maps of the input image I_j. The combination of the intensity gradient and saturation gradient will capture all the possible intensity variations present in the image which will ultimately describe the textural property of the image. Parallelly, use of sobel and fuzzy difference based edge maps will locate all possible geometrical structures of the image. So, the combination of these two will extract the detailed textural and structural portion of the image. Algorithm 1 (see Fig. 2 [35]) explains the image texture and colour separation process in detail. Here, Fig. 3 shows the schematic flow diagram of texture and colour region generation process for the first iteration. Simultaneously, Algorithm 1 (Fig. 2) shows the detailed steps of the proposed texture and colour region separation scheme.

3.2 Probability annular colour histogram (colour feature extraction)

In this proposed novel colour feature extraction scheme, initially we have extracted the semantic region (object region) of the image I_j using technique proposed by Pradhan *et al*. [36]. The mid point co-ordinate (X_c, Y_c) of the semantic region will be the semantic centre of the input image I_j. Further, we have divided the input colour region image Cl_j into n annular regions with centre (X_c, Y_c). Subsequently, from each annular region, we have computed the colour histograms of all colour components. Next, we have computed the probability histogram of all colour histograms based on a cumulative probability approach. Later, we have decomposed each probability histogram into k non-uniform bins for extraction of final colour features. The semantic centre based annular division supports the object based position invariance properties of background information. Usually, in the object region the texture property of the image dominates over colour features. So, most of the colour features become background information as shown in

Fig. 3. This proposed semantic centre based annular division scheme will control and preserve the distribution of the colour features with respect to the object region. Let Cl_j is the colour region of the image I_j with (x_c, y_c) as the semantic centre of the image. Here, the maximum possible radius ($D_{\max}$) of a circular annular region is defined as

$$D_{TL} = \sqrt{(X_c - 0)^2 + (Y_c - 0)^2} \tag{12}$$

$$D_{TR} = \sqrt{(X_c - X_{\max})^2 + (Y_c - 0)^2} \tag{13}$$

$$D_{BL} = \sqrt{(X_c - 0)^2 + (Y_c - Y_{\max})^2} \tag{14}$$

$$D_{BR} = \sqrt{(X_c - X_{\max})^2 + (Y_c - Y_{\max})^2} \tag{15}$$

$$D_{\max} = \text{maxof}(D_{TL}, D_{TR}, D_{BL}, D_{BR}) \tag{16}$$

Here, $(0, 0)$, $(X_{\max}, 0)$, $(0, Y_{\max})$, and $(X_{\max}, Y_{\max})$ are the co-ordinates of the four corners of the image, i.e. top-left, top-right, bottom-left, and bottom-right corners, respectively. Similarly, D_{TL}, D_{TR}, D_{BL}, and D_{BR} are the distances of the corner points from the semantic centre (X_c, Y_c).

Further, we have divided the whole image into n annular regions with the distance range of $D_{\max}/n$. The range of ith annular region A_i has been defined as

$$\text{Range}(A_i) = A_i - A_{i-1} \tag{17}$$

where annular regions A_{i-1} and A_i are the circular regions with $(D_{\max} \times i)/n$ and $(D_{\max} \times (i-1))/n$ as the radius. So, every pixel $P(x, y)$ must belong to any one of the annular regions, which is defined as follows:

$$P(x, y) \in \begin{cases} A_1, & \text{If } 0 < D_p \le (D_{\max} \times 1/n) \\ A_2, & \text{If } (D_{\max} \times 1/n) < D_p \le (D_{\max} \times 2/n) \\ A_3, & \text{If } (D_{\max} \times 2/n) < D_p \le (D_{\max} \times 3/n) \\ \dots & \\ A_n, & \text{If } (D_{\max} \times (n-1)/n) < D_p \le D_{\max} \end{cases} \tag{18}$$

where $D_p = \sqrt{|x_c - x|^2 + |y_c - y|^2}$. Now, compute the colour histogram of all the annular regions and calculate the probability histograms. The ith coefficient of a probability histogram PH_j of any colour histogram H_j is defined as follows:

$$PH_j(i) = \frac{i\text{th coefficient of the colour histogram } H_j}{\text{Total sum of all coefficients of } H_j} \tag{19}$$

Further, every probability histograms have been divided into $k = 10$ non-uniform bins where the cumulative probability of every bin must be less than or equal to 0.1. Finally, some statistical parameters like mean and standard deviation have been computed from every bin of all probability histograms and stored in a form of a colour feature vector. Fig. 4 shows the diagram of annular region decomposition of a colour region image. Next, Fig. 5 shows an example of a query image and its corresponding probability annular colour histogram with four annular regions.

3.3 Texture and shape feature extraction

The texture region Tr_j of any input image I_j possess rich textural and structural information since this region carries collection of huge heterogeneous pixels. It means the local features of this texture region Tr_j varies drastically in different local areas of the Tr_j. To capture these diverse textural and structural features, we have adopted the 2D DT-CWT along with the SIFT. The 2D DT-CWT will extract the texture feature along six directions, i.e. ±15°, ±45°, and ±75°. At the same time, the SIFT feature ensures the detection of the unique structural key points of the texture region Tr_j. We have employed the 2D DT-CWT for three levels of

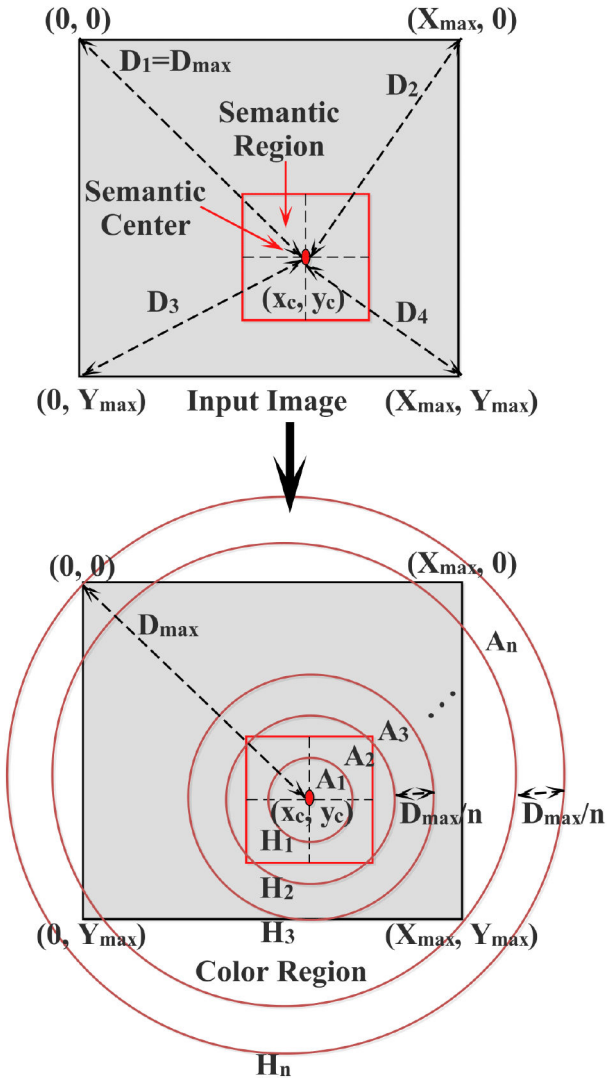

Fig. 4 *Annular region decomposition of a colour region image*

decompositions. This will extract the total six directional wavelet coefficients in each level of decomposition with two approximation wavelet coefficients from the final level of decompositions. Further, we have computed the first-order statistical features from the each wavelet coefficient. Simultaneously, we have extracted the 128 key structural features from the Tr_j using SIFT. The final texture region feature vector ft combines the directional and structural features all together.

3.4 Weighted similarity matching

Let

$$ft = (ft_1, ft_2, ft_3, \dots, ft_x)$$

and

$$fc = (fc_1, fc_2, fc_3, \dots, fc_y)$$

represent the extracted texture and colour feature vectors of the query image I_j. Similarly,

$$ft(DB)_j = (ft(DB)_{j1}, ft(DB)_{j2}, ft(DB)_{j3}, \dots, ft(DB)_{jx})$$

and

$$fc(DB)_j = (fc(DB)_{j1}, fc(DB)_{j2}, fc(DB)_{j3}, \dots, fc(DB)_{jy})$$

represent the texture and colour feature vectors of every database image, where $j = 1, 2, 3, \dots, |Z|$ and Z is the digital image database.

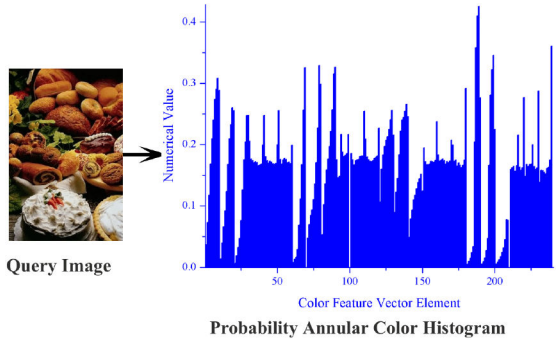

Fig. 5 *Example query image and its probability annular colour histogram representation with four annular regions*

Input: An RGB colored image I_j.
Output: Top L most similar images from image database Z.
1: Load colored query image.
2: Decompose the query image into texture and color region image.
3: Apply 2D DT-CWT in all three color components of texture region image for 3 level of decomposition. Simultaneously, compute the mean and standard deviation from all directional and last approximation coefficients.
4: Find the semantic center (X_c, Y_c) of the image I_j as the center of the main object region (semantic region) of the image I_j.
5: Apply probability annular histogram technique with center (X_c, Y_c) on the color region image and construct the color feature vector.
6: Apply scale-invariant feature transform (SIFT) [36] on the gray-scale texture region image and compute the top 128 key points of the geometrical structures.
7: Apply 2D DT-CWT on each color plane of the texture image for 3 decomposition level to extract the textural features of the texture image.
8: Construct the texture and shape feature vector which combines the normalized 2D DT-CWT and SIFT features.
9: Similarly, compute the texture and color feature vectors for all database images.
10: Perform the weighted similarity matching scheme between the query image and database image feature vectors.
11: Retrieve the top L most similar images.

Fig. 6 *Algorithm 2: CBIR framework*

In the process of similarity matching, our main aim is to retrieve top L most similar images from the image database. In the first step, we have used Euclidean distance to measure the similarity between texture and colour feature vectors of query image and database images. The similarity between the query image feature vectors and jth database image feature vectors is defined as follows:

$$D_j^T = \left(\sum_{m=1}^{x} \left| ft_m - ft(DB)_{jm} \right| \right)^{1/2} \tag{20}$$

$$D_j^C = \left(\sum_{m=1}^{y} \left| fc_m - fc(DB)_{jm} \right| \right)^{1/2} \tag{21}$$

Here D_j^T and D_j^C represents the texture and colour feature similarity between the query image I_j and jth database image. Further, the total feature similarity D_j has been defined as follows:

$$D_j = w_1 \times D_j^T + w_2 \times D_j^C \tag{22}$$

Here, w_1 and w_2 are the weight factors which show the semantic importance of texture feature and colour feature of a query image. w_1 and w_2 have been defined as follows:

$$w_1 = \frac{E(\partial TF_j)}{E(\partial TF_j) + E(\partial Cl_j)} \tag{23}$$

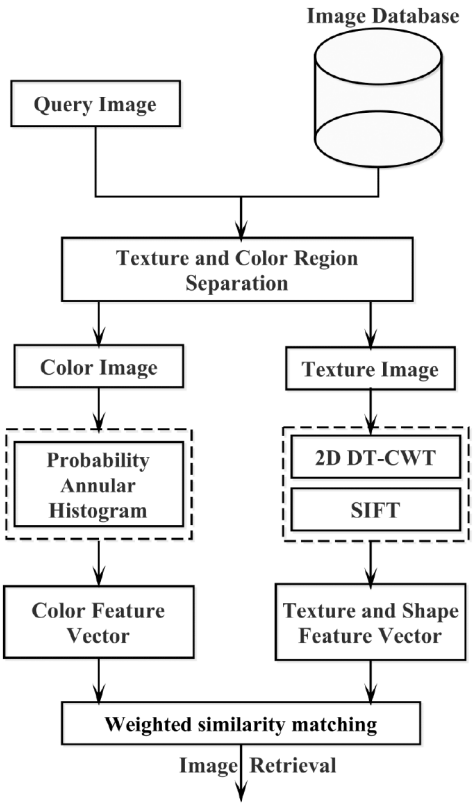

Fig. 7 *Schematic flow diagram of the proposed CBIR framework*

$$w_2 = \frac{E(\partial Cl_j)}{E(\partial TF_j) + E(\partial Cl_j)} \tag{24}$$

Here, $E(\cdot)$ is the entropy function and TF_j and Cl_j are the texture and colour regions of the query image I_j. Simultaneously, ∂A is defined as follows:

$$\begin{aligned} \partial A(x,y) \quad = \quad & \frac{1}{2}(A(x+1,y) \\ & -A(x-1,y)) + \frac{1}{2}(A(x,y+1) - A(x,y-1)) \end{aligned} \tag{25}$$

Finally, top L database images which have minimum D_j values will be retrieved as the final outcome of the CBIR process.

3.5 Proposed framework

Algorithm 2 (see Fig. 6 [37]) depicts the basic steps of the proposed CBIR scheme and Fig. 7 shows the schematic flow diagram of the proposed CBIR framework.

4 Experimental results and discussion

We have conducted several retrieval experiments on seven different types of image databases. The detailed performance analyses have been presented in the following subsections. The performance of the image retrieval experiments has been measured in terms of precision ($\eta(p)$), recall ($\eta(r)$), and f-score ($\eta(fs)$) parameters.

4.1 Performance evolution parameters

All of these three performance evolution parameters have been defined as follows:

$$\eta(p)\% = \frac{1}{|Z|}\left(\sum_{j=1}^{|Z|} \frac{R_j}{R_j + N_j} \times 100 \right) \tag{26}$$

$$\eta(r)\% = \frac{1}{|Z|}\left(\sum_{j=1}^{|Z|} \frac{R_j}{R_j + ND_j} \times 100 \right) \tag{27}$$

$$\eta(fs)\% = \left(\frac{2 \times \eta(p) \times \eta(r)}{\eta(p) + \eta(r)} \times 100\right) \quad (28)$$

Here, $|Z|$ is the size of the image dataset and R_j is the relevant set of the image in all retrieved images. Similarly, N_j is the non-relevant set of images in all retrieved images and ND_j are the total relevant image set present in the image databases apart from the retrieved set of images.

4.2 Database description

In all image retrieval applications, we have considered different image databases for fair performance analysis. In these databases, first four are natural, two are the object, and one is texture image dataset. The brief introduction about all these different databases has been shown in Table 1.

4.3 Execution time and complexity analysis

We have conducted all the experiments on a Matlab R2013a application software. We have used a personal computer system having Windows 10 operating system with 4GB RAM and 1TB Hard Drive with Intel (R) Core(TM) i7-4770 CPU @ 3.40 G Hz 64-bit processor. The execution time is the CPU time taken by any process for the execution. Execution time shows that how fast any process is working. The overall retrieval speed of CBIR scheme depends on the total execution time taken by the different process of the CBIR system. In our proposed CBIR system, there are total five different processes have been used. The first process (P_1) is separation of texture and colour regions from the input image. The other four processes are texture feature extraction (P_2), colour feature extraction (P_3), weighted similarity matching (P_4), and image retrieval (P_5). Table 2 shows the CPU time taken by all these five processes in the image retrieval experiment performed on seven different image databases.

In Table 2, we can see that the execution time for the different processes of the proposed CBIR are in the considerable limits. Further, we have also shown the time complexity of each process of the proposed CBIR system. Here, in our proposed scheme, the first process P_1 is the texture and colour region separation task. Here we have suggested an iterative approach for the texture and colour region separation. In this process, we have computed the intensity gradient, saturation gradient, binary Sobel edge map, and binary fuzzy edge map. For an $n \times n$ input query image, the computation cost of intensity and saturation gradients are $O(n^2)$, respectively. Next, binary Sobel and fuzzy edge map generation also bears $O(n^2)$ computation cost. So, total computation cost for one iteration of P_1 bears $O(n^2) + O(n^2) + O(n^2) + O(n^2) \simeq O(n^2)$. Hence, for k (where k is a constant) iterations of P_1 will have $O(k \times n^2)$ computation cost. Next, in process P_2 we have extracted the SIFT and 2D DT-CWT texture features. Here, SIFT bears $O(n^2)$ computation cost. DT-CWT computes two real DWT and one level of DWT decomposition takes $O(n^2)$ time. So, k levels of DT-CWT decomposition also takes $O(k \times n^2) + O(k \times n^2) \simeq O(k \times n^2)$ time. Next, in process P_3 we have computed the annular colour histogram. Here, for an $n \times n$ input image, the probability annular histogram computation with k annular regions take $O(k \times n^2)$ time. Next, process P_4 is the weighted similarity matching and it takes $O(n^2)$ time. Finally, in process P_5 we have performed the final image retrieval. It uses a distance (similarity value) sorting algorithm which runs on $O(n\log n)$ time. So, the overall time complexity of the proposed CBIR is $O(n^2) + O(k \times n^2) + O(k \times n^2) + O(n^2) + O(n\log n) \simeq O(k \times n^2)$. Further, we have also shown the time complexity comparisons for an $n \times n$ image in Table 3. Here, we have selected four recent CBIR schemes which are GMM+CQ [6], MFC-BTC [21], PTD-IR +CEF [18], and TT+JEH+CCCH [19]. In Table 3, we can see that the time complexity of the proposed scheme is lower than that of three out of four compared schemes. This execution and computation time performance analysis shows that the proposed CBIR is fast and reliable.

4.4 Performance of texture and colour image separation scheme

Here, we have performed the different number of iterations in the proposed texture and colour region separation algorithm and we have found that six iterations are sufficient for the separation. Here, Fig. 8 shows the output of the separation algorithm with six iterations.

Table 1 Brief description about all seven image databases use in the retrieval experiment

Dataset	Nature	Total image	Class	Images
COREL 1000 [38]	natural	1000	10	100
GHIM-10K [39]	natural	10000	20	500
COREL-5K [39]	natural	5000	50	100
COREL-10K [39]	natural	10000	100	100
COIL-100 [40]	object	7200	100	72
PRODUCE-1400 [41]	object	1400	14	100
OUTEX [42]	texture	4320	24	180

Table 2 CPU time taken (in seconds) by different process for execution

Dataset	P_1	P_2	P_3	P_4	P_5
COREL-1000	0.246	0.049	2.013	0.010	0.021
GHIM-10K	0.440	0.037	1.917	0.010	0.074
COREL-5K	0.086	0.036	1.922	0.010	0.035
COREL-10K	0.089	0.047	2.077	0.010	0.077
COIL_100	0.082	0.021	1.404	0.010	0.062
PRODUCE-1400	0.522	0.071	2.101	0.010	0.037
OUTEX	0.321	0.027	1.052	0.010	0.041

Table 3 Time complexity comparisons between proposed and four recent CBIR schemes

Methods	GMM + CQ	MFC -BTC	PTD-IR + CEF	TT + JEH + CCCH	Proposed
Time complexity	$O(n^3)$	$O(n^2)$	$O(n^2\log n^2)$	$O(n^3)$	$O(kn^2)$

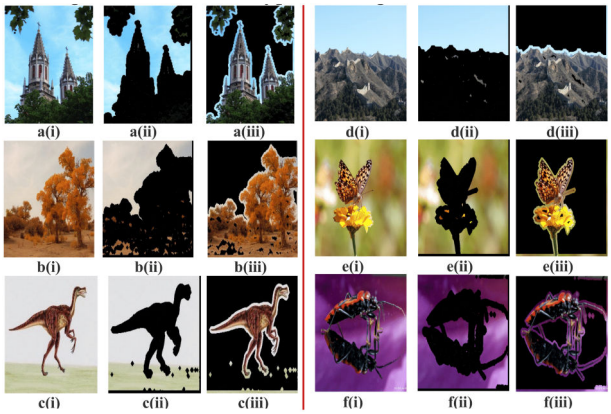

Fig. 8 *Final extracted texture and colour regions of six different images after applying six iterations on separation algorithm* ***(a(i))-(f(i))*** The input query images, ***(a(ii))-(f(ii))*** The output colour region of the images, ***(a(iii))-(f(iii))*** The output texture region of the images

Table 4 Image retrieval performance in terms of average precision for seven different image datasets

Dataset	Average precision ($\eta(p)$%)			
	$n < 5$	$n < 10$	$n < 15$	$n < 20$
COREL-1000	94.67	88.50	84.23	81.03
GHIM-10K	83.12	80.48	79.67	78.56
COREL-5K	84.02	77.67	75.45	72.14
COREL-10K	81.23	76.45	72.67	70.82
COIL-100	90.42	86.67	82.78	79.67
PRODUCE-1400	99.12	96.73	92.22	90.58
OUTEX	96.31	94.72	93.44	89.46

Table 5 Image retrieval performance in terms of average recall for seven different image datasets

Dataset	Average recall ($\eta(r)$%)			
	$n < 5$	$n < 10$	$n < 15$	$n < 20$
COREL-1000	4.73	8.85	12.63	16.21
GHIM-10K	0.83	1.61	2.39	3.14
COREL-5K	4.20	7.77	11.32	14.23
COREL-10K	4.06	7.65	10.90	14.16
COIL-100	6.28	12.04	17.25	22.13
PRODUCE-1400	6.88	13.44	19.21	25.16
OUTEX	2.68	5.26	7.76	9.94

In Fig. 8, first a(i) to f (i) show the original image, a(ii) to f (ii) show the output colour region image, and a(iii) to f (iii) show the output texture region image. In Fig. 8, we can see that the proposed scheme is working well for different types of images.

4.5 Feature vector description

We have constructed two types of feature vectors, i.e. texture *ft* and colour *fc* feature vector. Texture feature vector *ft* contains 128 key points generated from SIFT operation and 120 statistical features computed from 18 directional wavelets and 2 approximation wavelets of each colour component of the texture image. So, the total length of the texture feature vector ft is 248. Simultaneously, we have used $n = 4$ for the annular region decomposition. Though, set of four annular probability histograms have been computed for each colour component of the colour region image. Every probability histogram has been divided into 10 non-uniform bins and 20 statistical parameters have been computed in total. Therefore, length of colour feature vector *fc* is 240.

4.6 Retrieval performance analysis

In this section, different image retrieval experiments have been carried out on all previously discussed seven image datasets. In these retrieval experiments, we have retrieved different numbers (n) of similar images from the different image datasets with respect to a given query image. The retrieval performances have been measured in terms of the overall precision, recall, and f -score. We have chosen these three performance evolution parameters because precision gives system accuracy in retrieval, recall gives the robustness of the system, and f -score gives the total performance of the CBIR system. Here, Table 4 shows the overall retrieval precision of the proposed CBIR scheme. In Table 4, we have shown the average retrieval precision for $n < 5$, $n < 10$, $n < 15$, and $n < 20$ (where n is the number of retrieved images) retrieved similar images from seven different image datasets. Similarly, Tables 5 and 6 show the image retrieval performance in terms of average recall and f -score. In all the image retrieval experiments we have used the proposed weighted similarity matching approach.

In Tables 4–6, we can see that the retrieval precisions are more than 70% in most of the cases. Here, the average retrieval precision, recall, and f -score are high for Produce and Outex image dataset because Produce is an object dataset in which images have clear and definite texture and colour regions. Simultaneously, Outex is a texture image dataset in which same class images possess high textural similarities. At the same time, retrieval performance for COREL-10K is relatively low because COREL-10K is an natural image dataset in which huge diversity is present among the same class images also. Further, Fig. 9 demonstrates the average retrieval performances for different numbers of retrieved images from COREL-1000 dataset. Next, the

Table 6 Image retrieval performance in terms of average *f*-score for seven different image datasets

Dataset	Average *F*-score ($\eta(fs)$%)			
	$n < 5$	$n < 10$	$n < 15$	$n < 20$
COREL-1000	9.02	16.09	21.97	27.01
GHIM-10K	1.65	3.16	4.64	6.04
COREL-5K	8.00	14.12	19.68	24.05
COREL-10K	7.74	13.90	19.96	23.61
COIL-100	11.74	21.14	28.54	34.64
PRODUCE-1400	12.87	23.59	31.80	39.38
OUTEX	5.21	9.97	14.32	17.89

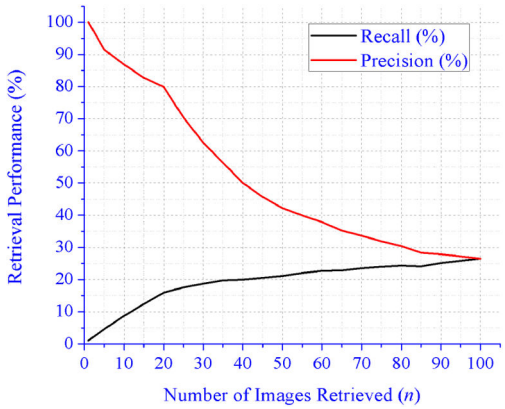

Fig. 9 *Image retrieval performances on COREL-1000 image dataset for different n values*

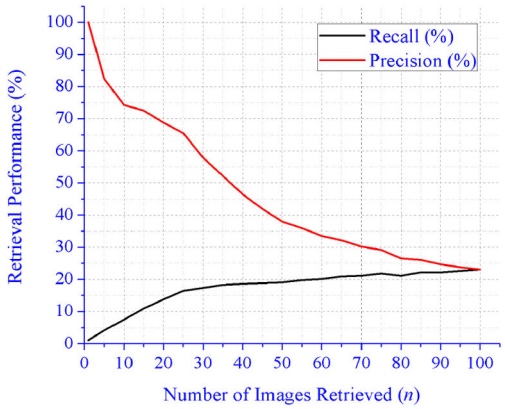

Fig. 11 *Image retrieval performances on COREL-5000 image dataset for different n values*

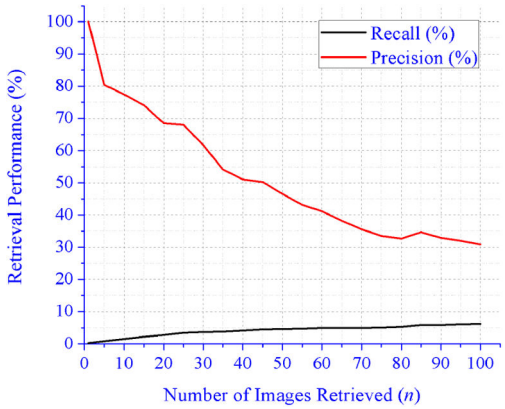

Fig. 10 *Image retrieval performances on GHIM image dataset for different n values*

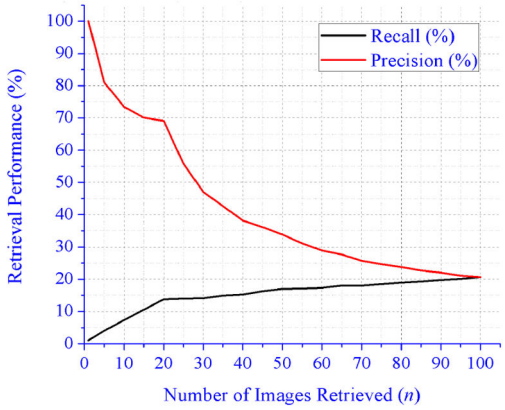

Fig. 12 *Image retrieval performances on COREL-10K image dataset for different n values*

detailed average retrieval performances for GHIM and COREL-5K datasets have been presented in Figs. 10 and 11.

Next, Fig. 12 demonstrates the detailed retrieval performances for different numbers of retrieved image from COREL-10K image dataset. Subsequently, Figs. 13 and 14 show the detailed retrieval performances for COIL and Produce image datasets for different n values. Here, in Figs. 9, 11, 12, and 14, we can observe that the average precision and recall is same for $n = 100$ because all these datasets contain 100 images in each class. In Fig. 10, the average precision and recall will become same at $n = 500$ since GHIM has 500 different images in each image class. In Fig. 10, the retrieval recall for smaller n values are very low because GHIM has 500 images in each class and the recall value also inversely depends on the total number of images present in the dataset. GHIM has five times more numbers of images in each class as compared to the COREL datasets. As a result, the retrieval recall of GHIM for smaller n values are lower than the retrieval recall values of the COREL datasets. Similarly, in Fig. 13, the average precision and recall is same for $n = 72$ because COIL contains 72 images in every image class.

Later, Fig. 15 shows the detailed retrieval performances for texture image dataset (i.e. Outex) for different n values. Here, in Fig. 15, average precision and recall will become same at $n = 180$ because the Outex dataset has 180 different images in each category of images. In this section, we have considered seven different image datasets of Corel, object, and texture images. In total, these datasets contain 38,920 different images which belong to 318 different classes. These numbers of images are significant to validate the retrieval performance of the proposed CBIR system. In the above retrieval results, we can see that the proposed CBIR system is demonstrating satisfactory retrieval performances in all different scenarios.

Figs. 9–15 show the efficiency as well as robustness of the proposed CBIR scheme for different kinds of image datasets. In all these figures, we can observe that the retrieval precisions are better for lower n values whereas retrieval recalls are better for higher n values. These statistical results show that the proposed CBIR scheme gives maximum analogous images as a retrieval outcome for every query image when we have to retrieve less number of images from image datasets. At the same time, good recall values also confirm that the proposed CBIR system also extracts

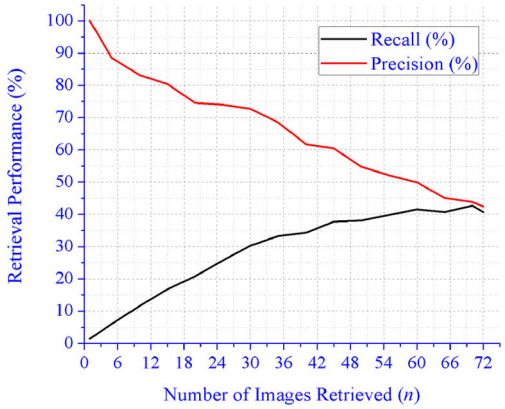

Fig. 13 *Image retrieval performances on COIL image dataset for different n values*

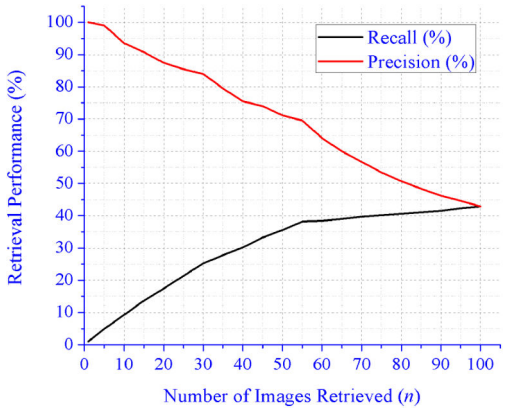

Fig. 14 *Image retrieval performances on Produce image dataset for different n values*

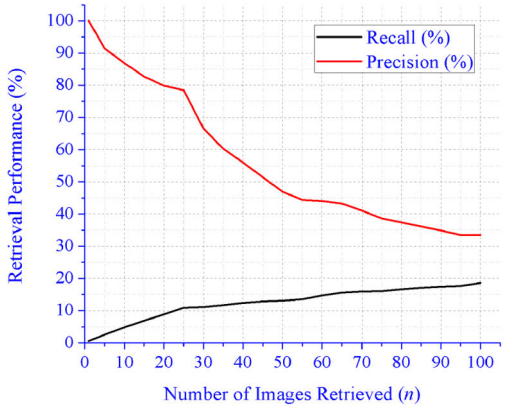

Fig. 15 *Image retrieval performances on Outex image dataset for different n values*

significant quantity of similar images in retrieval outcome for larger n values. Together the acceptable average precision and recall values of the proposed method for seven different image datasets validate that the proposed CBIR system is efficient and robust for all kind of image retrieval applications.

4.7 Performance analysis of weighted similarity scheme

Here, we have performed image retrieval experiments on selected seven different image datasets and retrieved $n < 5$ and $n < 10$ images. We have also considered different possible fixed values of w_1 and w_2 for similarity matching purpose. Here, w_1 and w_2 are the weights of texture and colour based separate similarity values to compute the resultant similarity value for image retrieval. Further, we have compared the image retrieval performances of different fixed values of w_1 and w_2 with proposed weighted similarity matching scheme. We, have selected the different values of w_1 and

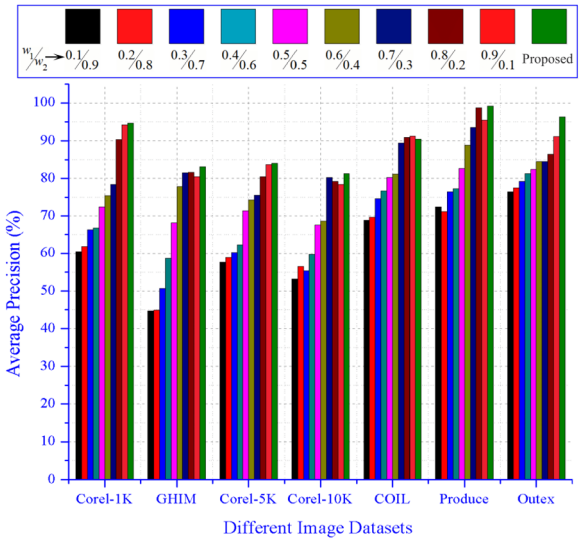

Fig. 16 *Image retrieval performance comparison between different values of w_1 and w_2 for $n < 5$*

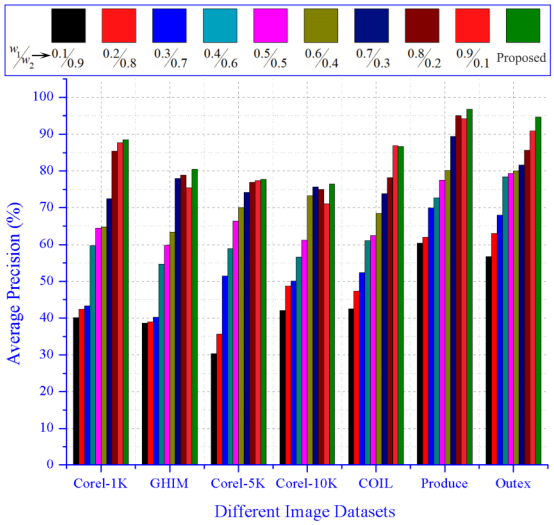

Fig. 17 *Image retrieval performance comparison between different values of w_1 and w_2 for $n < 10$*

w_2 between 0 and 1. Here, Figs. 16 and 17 are showing the retrieval performance comparison between different values of w_1 and w_2.

In these figures, w_1/w_2 show the ratio between selected vales of w_1 and w_2. Here, in Figs. 16 and 17 we can see that proposed weighted similarity scheme is performing better than fixed weight similarity scheme in most of the cases.

4.8 Performance comparison

In this paper, we have also performed image retrieval experiments from COREL-5000 image dataset using six different state-of-arts CBIR schemes. Further, we have computed retrieval performance of these seven CBIR schemes in terms of average precision, recall, and f-score. These seven CBIR schemes are MTH [10], MSD [15], CDH [16], GMM+CQ [6], NCH+CWF [17], PTD-IR+CEF [18], and TT+JEH+CCCH [19]. All these schemes have already been discussed in the related work section. All these schemes have been proposed in last few years. These seven CBIR schemes use either primitive visual features separately or combinedly. We have selected these schemes for comparisons because all these discussed schemes cover texture, shape, colour, structural, geometrical, co-relational, and local image feature based CBIR application. Though, for a fair comparison, we have selected these seven different schemes and performed the retrieval experiments on the COREL-5K and COREL-10K image datasets. We have compared the retrieval performance of these seven schemes with our

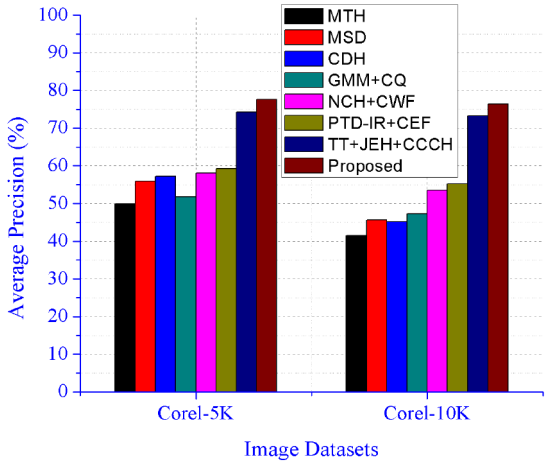

Fig. 18 *Average retrieval precision comparison between the proposed CBIR scheme and the seven different CBIR schemes*

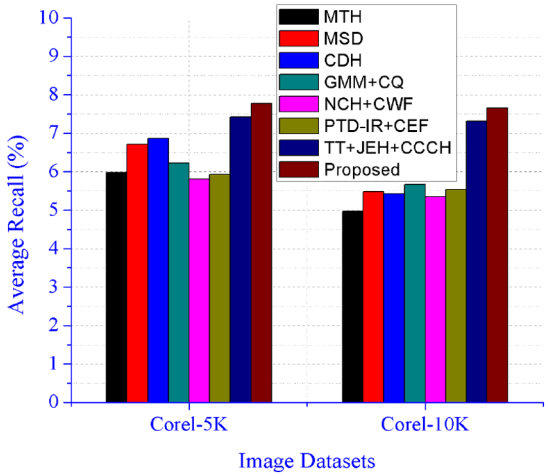

Fig. 19 *Average retrieval recall comparison between the proposed CBIR scheme and the seven different CBIR schemes*

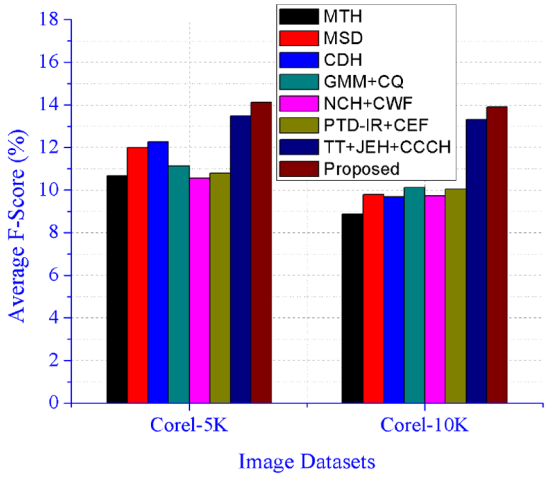

Fig. 20 *Average retrieval f-score comparison between the proposed CBIR scheme and the seven different CBIR schemes*

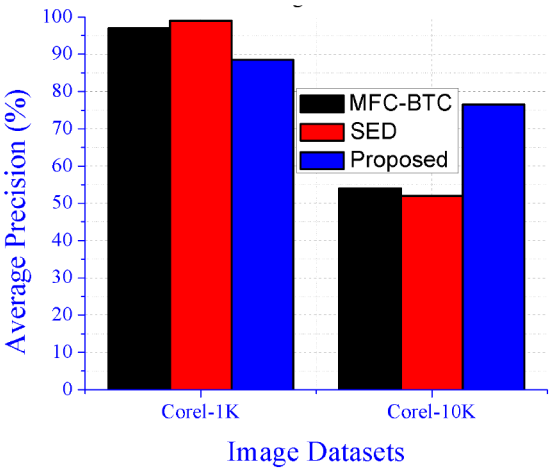

Fig. 21 *Average retrieval precision comparison between the proposed CBIR scheme and other related CBIR schemes*

proposed scheme and our scheme is performing better than these state-of-art methods.

The state-of-the-arts CBIR schemes choose the query images from every image category of the COREL-5K and COREL-10K image datasets. In this experiment, all images of the image dataset are possible contender for the query image. So, for a impartial performance comparison we have considered the average retrieval performances in terms of average precision, recall, and f -score. Here, Fig. 18 shows the detailed comparisons between proposed CBIR and different state-of-arts CBIR systems in terms of average precision values. Subsequently, the detailed comparisons for average recall and f -score values have been shown in Figs. 19 and 20. Among all these different CBIR systems, TT+JEH+CCCH is showing best average precision values, i.e. 74.22% for COREL-5K and 73.24% for COREL-10K. Whereas, our proposed CBIR scheme has shown improvements of 4.65 and 4.38% by giving 77.67% for COREL-5K and 76.45% for COREL-10K. At the same time, TT+JEH+CCCH is showing best average recall values, i.e. 7.42% for COREL-5K and 7.32% for COREL-10K. But our proposed CBIR scheme is giving 7.77% for COREL-5K and 7.65% for COREL-10K with the improvement of 4.72 and 4.51%. Simultaneously, for average retrieval recall also TT+JEH+CCCH is showing best results, i.e. 13.49% for COREL-5K and 13.31% for COREL-10K. In this case also, our proposed CBIR scheme has achieved improvements of 4.67 and 4.32% by giving 14.12% for COREL-5K and 13.90% for COREL-10K. From all these different comparisons we can observe that the proposed CBIR scheme has shown significant improvements over other different CBIR schemes.

Further, for more detailed retrieval performance evolution, we have also compared our proposed approach with two well-known colour image retrieval schemes which are MFC-BTC [21] and SED [20]. These two schemes uses combination of colour, textural, and structural image features for CBIR applications. In this case, we have also used COREL-1K and COREL-10K image datasets for fair comparisons. Here, in MFC-BTC and SED, the authors have selected ten images randomly from each image category from COREL-1K for query image. At the same time, they have selected 20 random image categories with 10 random images from COREL-10K image datasets for query images. Whereas in our proposed approach, every dataset image is a possible contender for query image. Though, MFC-BTC and SED show better precision for COREL-1K image datasets which are 97 and 99%. For COREL-1K, our proposed CBIR approach has also shown competitive performance. At the same time, for COREL-10K our proposed approach has shown significantly high performances as compared to the MFC-BTC and SED. Here, Fig. 21 shows the detailed comparison in terms of average precision values between MFC-BTC, SED, and proposed CBIR for COREL-1K and COREL-10K image datasets. Later, the detailed comparisons of average recall and f -score values for COREL-1K and COREL-10K have been shown in Figs. 22 and 23.

Further, we have also compared our proposed CBIR approach with well-known VGG-16 CNN [26] based CBIR approaches. Here, we have considered four variants of VGG-16 CNN based CBIR systems which are VGG-16(A), VGG-16(B), VGG-16(C), and VGG-16(D). Among these four systems, the first three CBIR approach uses pre-trained VGG-16 CNN model for image classification and retrieval. Initially, VGG-16 CNN has been used to identify the class of the query image. Next, the image retrieval has been performed from the identified image class only. This scheme significantly reduces the final image search space. Here,

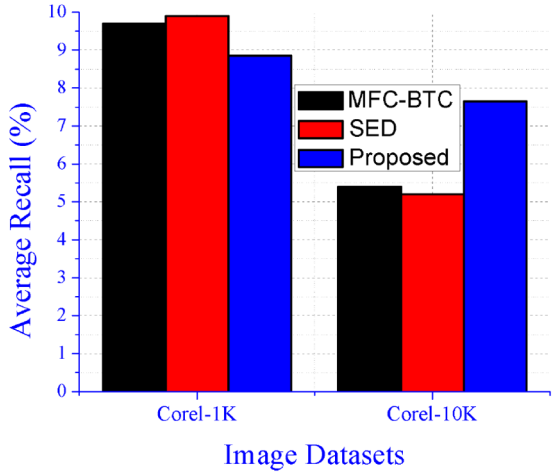

Fig. 22 *Average retrieval recall comparison between the proposed CBIR scheme and other related CBIR schemes*

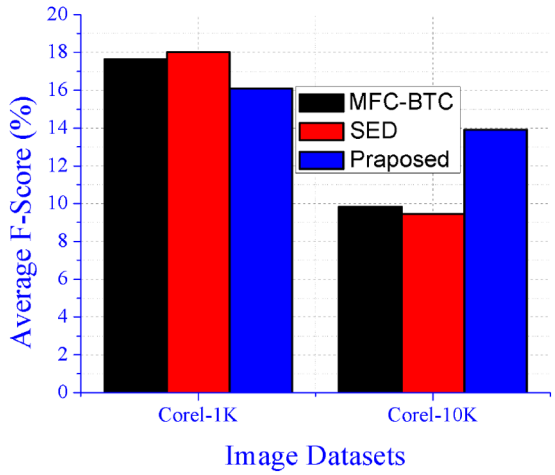

Fig. 23 *Average retrieval f -score comparison between the proposed CBIR scheme and other related CBIR schemes*

VGG-16(A) uses the actual extracted image features with a feature vector of length 49,152. In VGG-16(B) and (C), we have used a principal component analysis for post-feature selection. So, VGG-16(B) and (C) use reduced feature set of length 400 and 200, respectively. VGG-16(D) does not use pre-trained VGG-16 network so we have used 60% of COREL-1K images for training and rest 40% for testing. Here, Fig. 24 shows the detailed comparison of the proposed CBIR system w.r.t. the different VGG-16 based CBIR systems.

In Fig. 24, we can see that the performance of VGG-16(A) is better for most of the cases because it uses per-trained VGG network with actual extracted image features. But the main problem with this VGG-16(A) is its training time and feature vector length which adversely affects the retrieval time. At the same time, our proposed system has shown better performance w.r.t. VGG-16(A) when number of retrieved images are less than 10. Next, our system has shown better retrieval performances as compared to VGG-16(B) and (C) when number of retrieved images are less than 15. Simultaneously, our system has shown better results than the VGG-16(D) when number of retrieved images are less than 35. These above statistical results confirm that the proposed approach has demonstrated the acceptable and reliable image retrieval performances.

5 Conclusions

Most of the images possess well-defined separate texture and colour dominant regions. The texture region carries rich textural and shape feature information whereas colour region carries rich colour feature information. Usually, researchers perform feature extraction techniques on whole image itself which leads to an

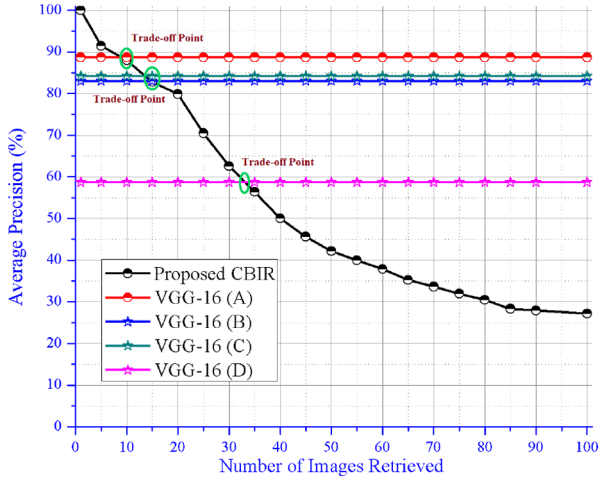

Fig. 24 *Average retrieval precision comparison between the proposed CBIR scheme and VGG_16 based CBIR schemes for Corel-1K image dataset*

inaccurate feature extraction. Since, the extracted features carry overlapped texture, shape, and colour features, which ultimately degrades the retrieval accuracy of the CBIR system. To address this problem, a CBIR system must extract the different visual features from their corresponding dominant regions. In this paper, an iterative approach has been proposed to separate the colour and texture regions of an image. This approach uses a combination of the intensity and saturation gradient maps along with the detailed edge map. This scheme will not only extract the texture regions but it also preserves the spatial structural features of the image. Later, we have extracted the colour features from the colour regions and texture and shape features from texture regions only. As a result, our proposed CBIR framework emphasises the quality of the extracted image features. We have used the combination of 2D DT-CWT along with SIFT and proposed semantic centred probability annular histogram to extract the texture, structural, and colour features from texture and colour region separately. Here, the combination of 2D DT-CWT and SIFT features will extract the six directional texture features along with 128 key structural points which represent the vital shape information. Simultaneously, semantic centred probability annular histogram extracts the object position invariant colour informations. The colour, texture, and shape information varies with respect to the images. So, it is not feasible to assign equal weights to all the low-level features in similarity matching process. Our proposed weighted similarity matching scheme computes the adaptive weights of colour and texture features based on its semantic importance to the image. We have used seven different and standard image datasets in retrieval experiments to validate the robustness and effectiveness of the proposed framework. Also, we have compared our proposed CBIR system with current sate-of-the-art CBIR systems. The experimental outcomes and the comparative analysis show that the proposed CBIR system is performing better in most of the cases.

6 References

[1] Gong, Y., Zhang, H., Chuan, H.C., *et al.*: 'An image database system with content capturing and fast image indexing abilities'. 1994 Proc. of IEEE Int. Conf. on Multimedia Computing and Systems, Boston, USA, 1994, pp. 121–130

[2] Guo, J.M., Prasetyo, H.: 'Content-based image retrieval using features extracted from halftoning-based block truncation coding', *IEEE Trans. Image Process.*, 2015, **24**, (3), pp. 1010–1024

[3] Swain, M., Ballard, D.: 'Indexing via color histograms'. Third Int. Conf. on Computer Vision, Maratea, Italy, 1994, pp. 390–393

[4] Huang, J., Kumar, S.R., Mitra, M., *et al.*: 'Image indexing using color correlograms'. Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 1997, pp. 762–768

[5] Rui, Y., Huang, T.S., Chang, S.F.: 'Image retrieval: current techniques, promising directions, and open issues', *J. Vis. Commun. Image Represent.*, 1999, **10**, (1), pp. 39–62. Available at http://www.sciencedirect.com/science/article/pii/S1047320399904133

[6] Zeng, S., Huang, R., Wang, H., *et al.*: 'Image retrieval using spatiograms of colors quantized by gaussian mixture models', *Neurocomputing*, 2016, **171**,

pp. 673–684. Available at http://www.sciencedirect.com/science/article/pii/S0925231215009820
[7] Beura, S., Majhi, B., Dash, R.: 'Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer', *Neurocomputing*, 2015, **154**, pp. 1–14. Available at http://www.sciencedirect.com/science/article/pii/S0925231214016968
[8] Chun, Y.D., Seo, S.Y., Kim, N.C.: 'Image retrieval using bdip and bvlc moments', *IEEE Trans. Circuits Syst. Video Technol.*, 2003, **13**, (9), pp. 951–957
[9] Wang, X.y., Chen, Z.f., Yun, J.j.: 'An effective method for color image retrieval based on texture', *Comput. Stand. Interfaces*, 2012, **34**, (1), pp. 31–35
[10] Liu, G.H., Zhang, L., Hou, Y.K., *et al.*: 'Image retrieval based on multi-texton histogram', *Pattern Recognit.*, 2010, **43**, (7), pp. 2380–2389
[11] Krommweh, J.: 'Tetrolet transform: a new adaptive haar wavelet algorithm for sparse image representation', *J. Vis. Commun. Image Represent.*, 2010, **21**, (4), pp. 364–374. Available at http://www.sciencedirect.com/science/article/pii/S1047320310000313
[12] Amanatiadis, A., Kaburlasos, V., Gasteratos, A., *et al.*: 'Evaluation of shape descriptors for shape-based image retrieval', *IET Image Process.*, 2011, **5**, (5), pp. 493–499
[13] Ziou, D., Tabbone, S.: 'Edge detection techniques an overview', *Int. J. Pattern Recognit.*, 1998, **8**, (4), pp. 537–559
[14] Singha, M., Hemachandran, K., Paul, A.: 'Content-based image retrieval using the combination of the fast wavelet transformation and the colour histogram', *IET Image Process.*, 2012, **6**, (9), pp. 1221–1226
[15] Liu, G.H., Li, Z.Y., Zhang, L., *et al.*: 'Image retrieval based on micro-structure descriptor', *Pattern Recognit.*, 2011, **44**, (9), pp. 2123–2133
[16] Liu, G.H., Yang, J.Y.: 'Content-based image retrieval using color difference histogram', *Pattern Recognit.*, 2013, **46**, (1), pp. 188–198
[17] Varish, N., Pradhan, J., Pal, A.K.: 'Image retrieval based on non-uniform bins of color histogram and dual tree complex wavelet transform', *Multimedia Tools Appl.*, 2017, **76**, (14), pp. 15885–15921. Available at https://doi.org/10.1007/s11042-016-3882-4
[18] Pradhan, J., Pal, A.K., Banka, H.: 'Principal texture direction based block level image reordering and use of color edge features for application of object based image retrieval', *Multimedia Tools Appl.*, 2018, **82**, pp. 1–33
[19] Pradhan, J., Kumar, S., Pal, A.K., *et al.*: 'A hierarchical CBIR framework using adaptive tetrolet transform and novel histograms from color and shape features', *Digital Signal Process.*, 2018, **82**, pp. 258–281
[20] Wang, X., Wang, Z.: 'A novel method for image retrieval based on structure elements' descriptor', *J. Vis. Commun. Image Represent.*, 2013, **24**, (1), pp. 63–74
[21] Wang, X., Wang, Z.: 'The method for image retrieval based on multi-factors correlation utilizing block truncation coding', *Pattern Recognit.*, 2014, **47**, (10), pp. 3293–3303
[22] Unar, S., Wang, X., Zhang, C.: 'Visual and textual information fusion using kernel method for content based image retrieval', *Inf. Fusion*, 2018, **44**, pp. 176–187
[23] dos Santos, J.M., de Moura, E.S., da Silva, A.S., *et al.*: 'Color and texture applied to a signature-based bag of visual words method for image retrieval', *Multimedia Tools Appl.*, 2017, **76**, (15), pp. 16855–16872. Available at https://doi.org/10.1007/s11042-016-3955-4
[24] Liu, P., Guo, J.M., Wu, C.Y., *et al.*: 'Fusion of deep learning and compressed domain features for content-based image retrieval', *IEEE Trans. Image Process.*, 2017, **26**, (12), pp. 5706–5717
[25] Meng, A., Shan, D., Shi, R., *et al.*: 'Merged region based image retrieval', *J. Vis. Commun. Image Represent.*, 2018, **55**, pp. 572–585 https://doi.org/10.1016/j.jvcir.2018.07.003, (http://www.sciencedirect.com/science/article/pii/S1047320318301627)
[26] Simonyan, K., Zisserman, A.: 'Very deep convolutional networks for large-scale image recognition'. arXiv preprint arXiv:14091556, 2014
[27] Liu, C., Wechsler, H.: 'Robust coding schemes for indexing and retrieval from large face databases', *IEEE Trans. Image Process.*, 2000, **9**, (1), pp. 132–137
[28] Wang, X.Y., Wang, Y.X., Yun, J.J.: 'An improved no-search fractal image coding method based on a fitting plane', *Image Vis. Comput.*, 2010, **28**, (8), pp. 1303–1308
[29] Wang, X.Y., Chen, Z.F.: 'A fast fractal coding in application of image retrieval', *Fractals*, 2009, **17**, (4), pp. 441–450
[30] Wang, C., Wang, X., Li, Y., *et al.*: 'Quaternion polar harmonic fourier moments for color images', *Inf. Sci.*, 2018, **450**, pp. 141–156
[31] Wang, C., Wang, X., Xia, Z., *et al.*: 'Ternary radial harmonic fourier moments based robust stereo image zero-watermarking algorithm', *Inf. Sci.*, 2019, **470**, pp. 109–120
[32] Lowe, D.G: 'Object recognition from local scale-invariant features'. The Proc. of the Seventh IEEE Int. Conf. on Computer Vision, 1999, Corfu, Greece, 1999, vol. 2, pp. 1150–1157
[33] Dou, J., Qin, Q., Tu, Z.: 'Robust image matching based on the information of sift', *Optik*, 2018, **171**, pp. 850–861
[34] Selesnick, I.W., Baraniuk, R.G., Kingsbury, N.C.: 'The dual-tree complex wavelet transform', *IEEE Signal Process. Mag.*, 2005, **22**, (6), pp. 123–151
[35] Zhu, N., Wang, G., Yang, G., *et al.*: 'A fast 2d otsu thresholding algorithm based on improved histogram'. 2009 Chinese Conf. on Pattern Recognition, Nanjing, China, 2009, pp. 1–5
[36] Pradhan, J., Pal, A.K., Banka, H: 'A prominent object region detection based approach for cbir application'. 2016 Fourth Int. Conf. on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, India, 2016, pp. 447–452
[37] Giveki, D., Soltanshahi, M.A., Montazer, G.A.: 'A new image feature descriptor for content based image retrieval using scale invariant feature transform and local derivative pattern', *Optik - Int. J. Light Electron Opt.*, 2017, **131**, pp. 242–254. Available at http://www.sciencedirect.com/science/article/pii/S0030402616313766
[38] Wang, J.Z., Li, J., Wiederhold, G.: 'Simplicity: semantics-sensitive integrated matching for picture libraries', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (9), pp. 947–963
[39] Liu, G.H., Yang, J.Y., Li, Z.: 'Content-based image retrieval using computational visual attention model', *Pattern Recognit.*, 2015, **48**, (8), pp. 2554–2566
[40] Nene, S.A., Nayar, S.K., Murase, H., *et al.*: 'Columbia object image library (coil-20)', 1996
[41] 'tropical-fruits-db-1024x768.tar.gz'. Accessed: 18 August 2017. http://www.ic.unicamp.br/rocha/pub/downloads/tropical-fruits-DB-1024x768.tar.gz/
[42] 'site www, vision & image', lagis-vi.univ-lille1.fr, 2017'. accessed: 18 August 2017. Available at http://lagis-vi.univlille1.fr/datasets/outex.html