IET Image Processing

Review Article

IET Journals
The Institution of Engineering and Technology

Hybrid CBIR method using statistical, DWT-Entropy and POPMV-based feature sets

ISSN 1751-9659
Received on 2nd July 2018
Revised 18th May 2019
Accepted on 1st July 2019
E-First on 15th August 2019
doi: 10.1049/iet-ipr.2018.5797
www.ietdl.org

D. Latha[1] ✉, *Y. Jacob Vetha Raj*[2]

[1]*NMC College, Affiliated to Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India*
[2]*NMC College, Affiliated to Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India*
✉ *E-mail: lathad_ann@yahoo.com*

Abstract: Content-based image retrieval (CBIR) is an image retrieval technique that can retrieve images by matching its feature set values. This research focuses on a novel CBIR method namely hybrid CBIR method using statistical, Discrete Wavelet Transform (DWT)-Entropy and Peak-oriented Octal Pattern-derived Majority Voting (POPMV)-based feature sets (CBIR_SWPOPMV) to efficiently extract the relevant colour images from the colour image dataset. The doctrine of the proposed method is influenced by a novel texture descriptor namely POPMV which is an octal pattern based on the histogram peak information, to bring about a majority voting-based feature set and three histogram-based feature sets. Furthermore, to improve the retrieval accuracy, the DWT-based Entropy feature set and the statistical feature set are also included. Finally, the Euclidean distance-based matching process brings more favourable relevant images with respect to the query image. The proposed methodology is experimentally compared with the existing recent CBIR versions by using seven standard databases such as Corel-1k, USPTex, MIT-VisTex, KTH-TIPS, KTH-TIPS2a, KTH-TIPS2a, Colored Brodatz and a user-contributed database named DB_VEG.

1 Introduction

Digital images perform a wonderful role in exploring and propagating the pictorial information. Hence, image databases are used in tremendous applications including criminal identification, geographic information system and medical imaging [1]. In the recent years, because of the availability of powerful digital cameras, high-speed internet connections and the falling price of storage devices, the magnitude of the digital images in the databases of these applications has been growing rapidly [2]. This image database consists of thousands of images and is very difficult for the user to browse through the entire database to retrieve an image. Therefore, an efficient and automated image retrieval system is needed. Content-based image retrieval (CBIR) [3–5] is a process that can retrieve similar images from huge databases by matching its feature descriptors instead of matching two images. It means that the performance of CBIR heavily depends on the matching of the image feature descriptors [6].

In the research community, mostly colour, texture and shape are used as the basic types of feature to describe the images as image descriptors [7]. Owing to the simplicity of extracting colour information from the images, the colour feature is used as one of the most common and determinant features in image retrieval systems and is also stable against direction variations, size of image and background complexity [4]. Texture feature [8] refers to the visible patterns containing surface structure and their relation to the environment. Recently, local pattern-based descriptors have been used for extracting the texture feature of an image. Local binary pattern (LBP) [9] has gained fame among all local pattern-based descriptors due to its simplicity and efficiency in several applications. Shape feature is more useful in semantically characterising the content of an image than other features. Shapes [10] are commonly determined by applying segmentation or edge detection on the input image. Some methods use shape filters to identify the given shapes of an image. An evaluation of MPEG-7-based shape descriptors has been expressed in [11, 12]. In [7] the MPEG-7 standard-based visual descriptors have been used to retrieve the relevant images.

Tang *et al.* [13] introduce a neighbourhood discriminate hashing method for image retrieval by exploiting the local discriminative information and also using compact hashing codes. Guo *et al.* [14] present two feature descriptors called Colour Histogram Feature and Bit Pattern Histogram Feature which are derived from Error Diffusion Block Truncation Coding (EDBTC) compression data stream.

A new logical compact LBP co-occurrence matrix for texture analysis [15] is demonstrated by Sujatha *et al.* This method first evaluates the micro textured features using textons. Then logical compact LBP with logical OR operation is evaluated on textons to make the texture feature invariant in terms of change of illuminations and image rotation. Finally, the grey level co-occurrence matrix is calculated. Kwitt *et al.* [16] propose a joint statistical modelling of Discrete Wavelet Transform (DWT)/ Discrete Cosine Transform (DCT) coefficients coupled with the Bayesian framework which improves the retrieval performance.

Bian and Tao [17] have used colour, texture and shape features for image retrieval. Under the colour space domain, 128-D colour coherence vectors in the Lab colour space and a 9-D colour moment feature in the Lab colour space are considered. Under the texture domain, pyramidal wavelet transforms for Y component is computed in the YCrCb colour space. For shape feature, an edge directional histogram is computed from Y component in the YCrCb colour space. Texton co-occurrence matrix [18] is proposed by Liu and Yang. The statistical information of texton is calculated for quantised image and is used as the image feature.

Ptino-Escarcina and Costa [19] propose an arrangement of low-level descriptors into hierarchy. In this work, only two descriptors namely Colour Layout Descriptor (CLD) and Edge Histogram Descriptor (EHD) are arranged into the hierarchical structure. The CLD is calculated by using DCT in the *YCbCr* colour space and the EHD is calculated by dividing the image into 4×4 blocks. Moghddam *et al.* [20] propose an algorithm that computes the wavelet coefficient of an image using a Gabor wavelet. Finally, images are retrieved by using wave correlogram; but in this method, different coefficients for each wavelet-resolution-level are not applied to increase the retrieval accuracy. A novel fusion approach is presented by Qi and Han [21], in which an image is first segmented into regions using colour clustering. Then two sets of features including colour and texture properties are derived. The region-based image-level similarity is then measured by applying

the fuzzy matching scheme to the fuzzified region based on colour and texture features. The major limitation of this method is that the retrieval performance depends on segmentation quality.

Liu *et al.* [22] suggest a promising descriptor namely Micro-Structure Descriptor (MSD). The MSD describes image features through micro-structures. The MSD is constructed based on computing edge orientation similarity and the underlying colours, so it can effectively represent the image local feature. Liu *et al.* [23] introduce a learning framework for CBIR, in which a soft hyper-graph is used to represent the relevance relationship among the vertices. Based on the matrix which is computed from the image features, each image is taken as a centroid vertex and forms a hyper-edge by a centroid and its nearest neighbours. In this way, image retrieval with relevance feedback is converted into a learning algorithm.

Dubey *et al.* [3] implement two multichannel decoded LBPs. They are Multichannel Adder Local Binary Pattern (MALBP) and Multichannel Decoder Local Binary Pattern (MDLBP). Basically, both MALBP and MDLBP have used the local information of three channels on the basis of the adder and decoder concepts. Based on the experimental results, this paper reports that the MALBP descriptor is not giving the best performance in most cases while MDLBP descriptor performs well than the existing multichannel-based descriptors.

A Local Bit-plane Decoded Pattern (LBDP)-based image feature description is introduced by Dubey *et al.* [5] for biomedical image retrieval. The local neighbourhood in each plane is transformed and then the relationship between the intensity values of the centre pixel and the transformed values is encoded to generate the LBDP binary pattern. Wang *et al.* [24] present a CBIR method by combining DCT intrinsic orientation and DCT Inspired Feature Transform. It first determines a DCT intrinsic orientation through DCT matrix and rotates the region, and then produces an optimised low-dimensional descriptor. In this paper, the dimension of the feature is reduced dramatically.

Fadaei *et al.* [25] extract colour features using Dominant Colour Descriptor. Several wavelet and curvelet features are used in this paper to extract texture feature. Finally, the Particle Swarm Optimisation algorithm has been used to optimally combine colour and texture features. This paper reports that the optimum combination of colour and texture features gives a better accuracy compared to the several existing CBIR systems. An unsupervised visual hashing approach called Semantic-Assisted Visual Hashing (SAVH) is proposed in the paper [26]. This paper suggests that a unified unsupervised framework can be achieved by simultaneously keeping the visual similarities of images, integrating the semantic assistance from auxiliary texts and characterizing the correlations between the images. An unsupervised visual hashing approach which is called as SAVH is described by Zhu *et al.* [26]. The prime scheme in this method is to extract the semantics embedded in auxiliary texts of images without using explicit semantic labels. A similarity learning method is described in [27] to maximise the top precision measure.

Although many methods were proposed since CBIR is originated in 1992, each method had drawbacks and inefficiencies predominantly in terms of retrieval accuracy and execution time. These drawbacks are mainly influenced by the less proficient image descriptors adopted in the existing methods. So, a new CBIR descriptor and methodology is proposed in this paper to minimise the drawbacks of the existing method. The proposed *CBIR_SWPOPMV* method is enriched by a descriptor namely POPMV which is generated from the octal pattern derived from the three histogram peak values and also considers majority voting. This POPMV image descriptor is processed to generate the first set of features which is used in the proposed *CBIR_SWPOPMV* method. They are the majority category feature set, the neighbour positive difference oriented feature set, the minimum oriented feature set and the maximum oriented feature set. DWT-based Entropy of an image is used to generate the second type of feature set in the *CBIR_SWPOPMV* method. The last feature set in the *CBIR_SWPOPMV* method is generated from the statistical features that are computed from the different colour spaces of an input image.

The remainder of this paper is structured as follows: Section 2 describes the proposed methodology of constructing an efficient image descriptor and feature sets. The performance of the proposed methodology is experimentally analysed and compared with other methodologies and is clearly exposed in Section 3. Finally, Section 4 precisely presents the conclusion of this paper with further enhancement.

2 Proposed methodology

This paper proposes a new method for an advanced CBIR system. A new descriptor pattern namely *Peak-oriented Octal Pattern-derived Majority Voting* (*POPMV*) is proposed in this paper to achieve better retrieval accuracy. This pattern is generated based on the histogram peak information and an octal-based data is stored in the POPMV descriptor. A majority voting-based feature is also extracted from the POPMV descriptor. The histogram-based features are derived from the POPMV image descriptor. The DWT features are derived from the input image and the statistical features are derived from the different colour spaces of the input image. Three feature sets are collected and integrated from the query input image to produce the final feature set vector. In the same manner, the database feature vectors are also computed based on the database images. Finally, Euclidean distance-based matching process is performed to find the relevant images with respect to the query image. Fig. 1 briefly describes the overall architecture of the proposed CBIR_SWPOPMV method. In the term CBIR_SWPOPMV, the character *S* indicates the statistical feature set and the character *W* indicates the wavelet feature. The characters 'POPMV' indicate the proposed POPMV-based feature set.

2.1 POPMV image descriptor

The POPMV descriptor converts the image into a new format by setting octal-based data. It can be treated as a new image representation which can be used as a source for feature extraction. Fig. 2 shows the block diagram of POPMV computation. The image information is read by pixel-wise and it is made into a two-dimensional matrix. An overlapping type window is chosen from the image based on a centralised processing pixel P_{ij} using (1). The order of processing locations can be marked based on Fig. 3:

$$\begin{aligned} Z_{\mathrm{W}} = \{&P_{i,j+1}, P_{i-1,j+1},\ P_{i-1,j}, P_{i-1,j-1}, P_{i,j-1},\\ &P_{i+1,j-1}, P_{i+1,j}, P_{i+1,j+1},\ P_{i,j}\}\\ &i \in [1,\ \mathrm{IH}-2]\\ &j \in [1,\ \mathrm{IW}-2] \end{aligned} \tag{1}$$

where Z_{W} is the window elements based on $P_{i,j}$, IH is the image height, IW is the image width.

In (1), the numerical 1 and 2 in $i \in [1,\ \mathrm{IH}-2]$ and $j \in [1,\ \mathrm{IW}-2]$ are utilised to solve the border issues in computing Z_{W}. The surrounded eight pixels of each and every element of Z_{W} are computed using (2) and (3):

$$Z^{q,k}_{\mathrm{NBR}} = \mathrm{FindNeighbours}(Z_{\mathrm{W}}) \tag{2}$$

$$\begin{aligned} Z^{q,k}_{\mathrm{DIFF}} &= Z^{q,k}_{\mathrm{NBR}} - Z^{q}_{\mathrm{W}}\\ q &\in [0,8]\\ k &\in [0,8] \end{aligned} \tag{3}$$

where FindNeighbours is the function to collect neighbour element, $Z^{q,k}_{\mathrm{NBR}}$ is the neighbour element set corresponding with qth position element, $Z^{q,k}_{\mathrm{DiFF}}$ is the difference value set corresponding with qth position element, Z^{q}_{W} is the 3×3 size window elements, k is the indicated surrounding index for qth location.

The computation of $Z^{q,k}_{\mathrm{DiFF}}$ generates nine vectors each having eight elements. The total element count is $9 \times 8 = 72$. Then the

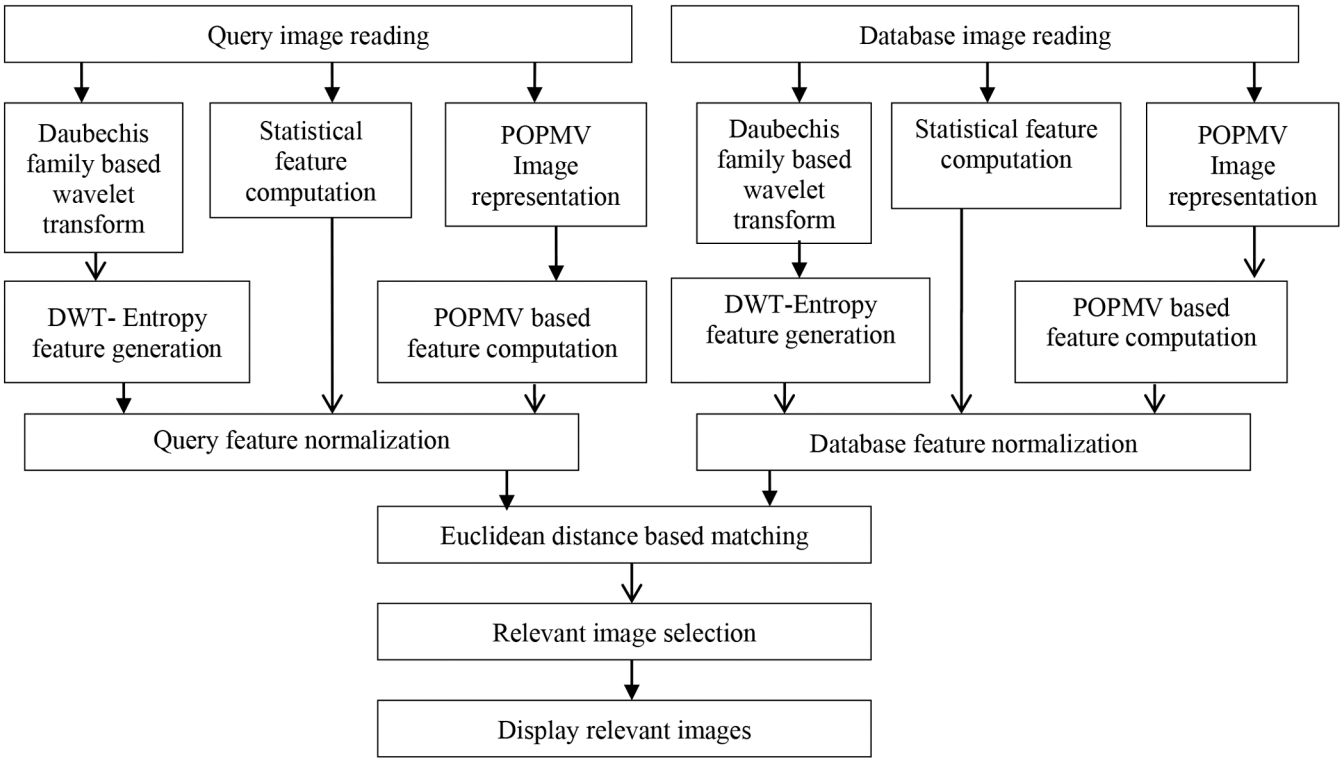

Fig. 1 *Architecture of the proposed CBIR_SWPOPMV method*

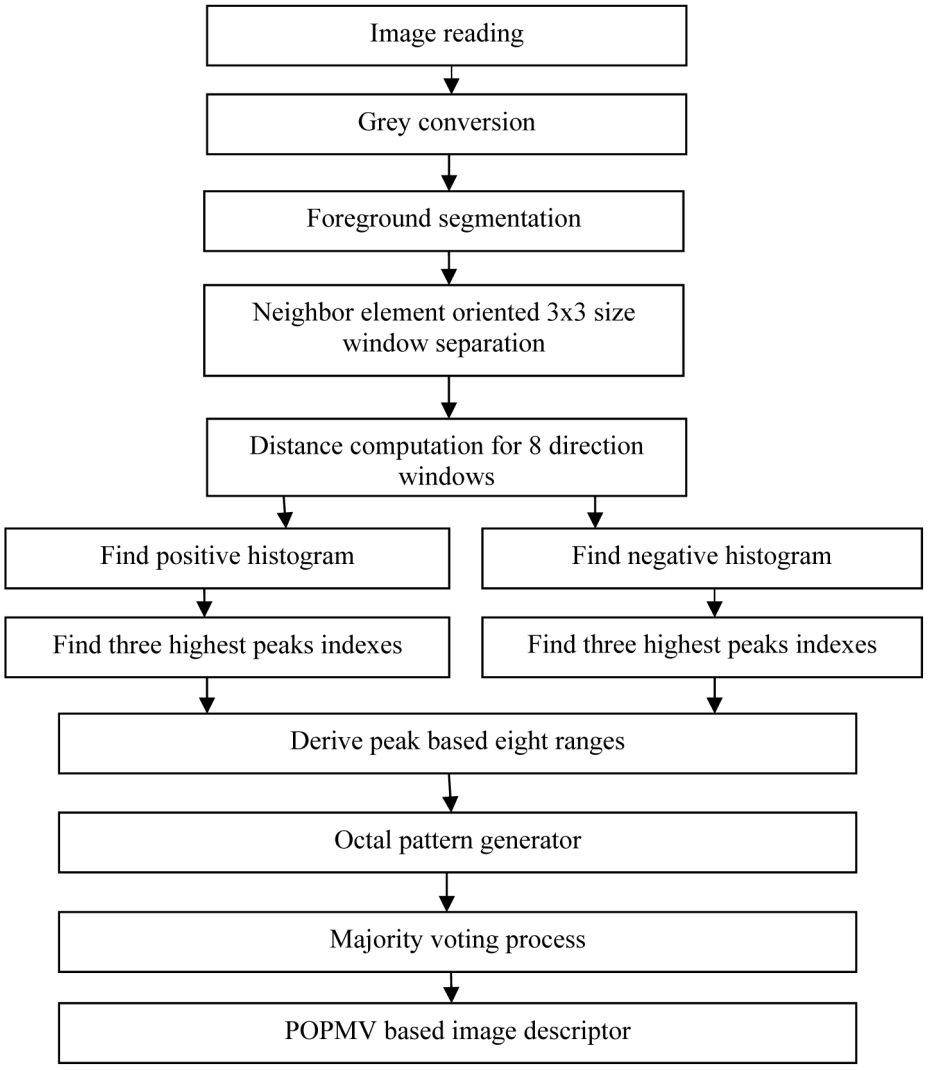

Fig. 2 *Block diagram of the proposed POPMV-based image descriptor*

3	**2**	**1**
4	$\boldsymbol{P_{ij}}$	**0**
5	**6**	**7**

Fig. 3 *Location map of overlapping 3 × 3 windows*

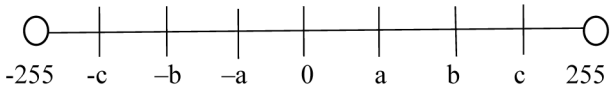

Fig. 4 *Description of sub-divisions in the histogram*

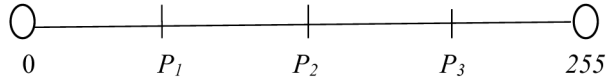

Fig. 5 *Demonstration of three peaks availability*

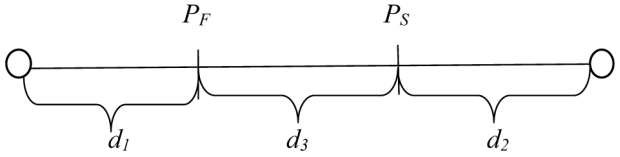

Fig. 6 *Demonstration of two peaks availability*

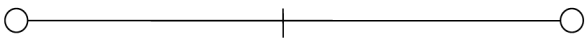

Fig. 7 *Demonstration of single peak availability*

histogram is found for these 72 elements and the histogram length is maintained in the range −255 to 255.

In the proposed method, there is a need for octal ranges or bins to get *Difference Category* component. The histogram is considered as a source of generating these eight ranges or bins since histogram acts as a compact representation of spatial data. These eight ranges can be derived from nine parameters of histogram based on Fig. 4. The parameters −255, 0 and 255 are built-in delimiters of the histogram. Three peaks are required in the positive direction of the histogram so that the histogram peak detection process searches for three peaks such as a, b and c in the positive histogram. Similarly, the negative histogram is also used to find three peaks such as $-a$, $-b$ and $-c$.

The range of histogram can be possibly divided into subdivisions as depicted in Fig. 4. Three peaks (a, b and c) in the positive direction histogram are found using any peak detection algorithm. In this work, peak detection is performed by MATLAB built in standard function:

$$a = P_1$$
$$b = P_2$$
$$c = P_3$$

where P_1 is the minimum peak, P_2 is the middle peak, P_3 is the maximum peak.

The values a, b and c are computed based on P_1, P_2 and P_3 as shown in Fig. 5 which shows the availability of three peaks.

Case 2: Two peaks are available.

The processing of two peaks availability case proceeds and the values of a, b and c are calculated using (4)–(9) as given in Fig. 6:

$$d_1 = |0 - P_F| \tag{4}$$

$$d_2 = |255 - P_S| \tag{5}$$

$$d_3 = |P_F - P_S| \tag{6}$$

$$a = \begin{cases} \frac{d_1}{2}, & \text{if } d_1 > d_2 \;\&\&\; d_1 > d_3 \\ P_F, & \text{otherwise} \end{cases} \tag{7}$$

$$b = \begin{cases} P_F, & \text{if } d_1 > d_2 \;\&\&\; d_1 > d_3 \\ P_S, & \text{if } d_2 > d_3 \\ P_F + \left(\frac{d_3}{2}\right), & \text{otherwise} \end{cases} \tag{8}$$

$$c = \begin{cases} P_S, & \text{if } d_1 > d_2 \&\& \; d_1 > d_3 \\ P_S + \left(\frac{d_2}{2}\right), & \text{if } d_2 > d_3 \\ P_S, & \text{otherwise} \end{cases} \tag{9}$$

where P_F is the first occurrence peak (MinPeak), P_S is the second occurrence peak (MaxPeak), d_1, d_2, d_3 are the distance values.

Case 3: Only one peak is available.

The values of a, b and c are calculated using (10)–(12) as shown in Fig. 7:

$$a = \frac{\text{PK}}{2} \tag{10}$$

$$b = \text{PK} \tag{11}$$

$$c = \text{PK} + \left(\frac{[\text{PK} - 255\}}{2}\right) \tag{12}$$

The values of a, b and c are calculated using (13)–(15):

$$R = 255$$
$$a = \frac{R}{4} \tag{13}$$

where PK is the peak value.

Case 4: No peaks are found.

$$b = a + \left(\frac{R}{4}\right) \tag{14}$$

$$c = b + \left(\frac{R}{4}\right) \tag{15}$$

where R is the maximum intensity value.

Case 1 explores the availability of three peaks and a, b and c values are assigned to *minimum peak*, *middle peak* and *maximum peak* values, respectively. Case 2 computations are built to fit the two peaks availability. The d_1, d_2 and d_3 computations are performed using (4)–(6). The term a is computed using (7) with the interaction of the variables d_1 and P_F. The computation of b is reached through (8) with the utilisation of the variables P_F, P_S and d_3. The variables P_S and d_2 are used to calculate c using (9). Case 3 reveals the condition of one peak availability. The equations (10)–(12) are utilised to compute a, b and c. Case 4 explains the condition of zero peak availability. The values of a, b and c are assigned with 64, 128 and 192, respectively, through the impact of (13)–(15). In the same way, the three components - a, - b and - c of negative direction histograms are computed based on Fig. 4.

The difference-based category formation is performed using a 3 × 3 window which is described in (16):

$$W_{m+1,n+1} = P_{i+m,j+n} \quad m \in [-1,\ 0,\ +1] \quad n \in [-1,\ 0,\ +1] \tag{16}$$

where W is the 3 × 3 window.

The difference calculation is found using the interaction of neighbour elements and centre pixel. Eight such difference values

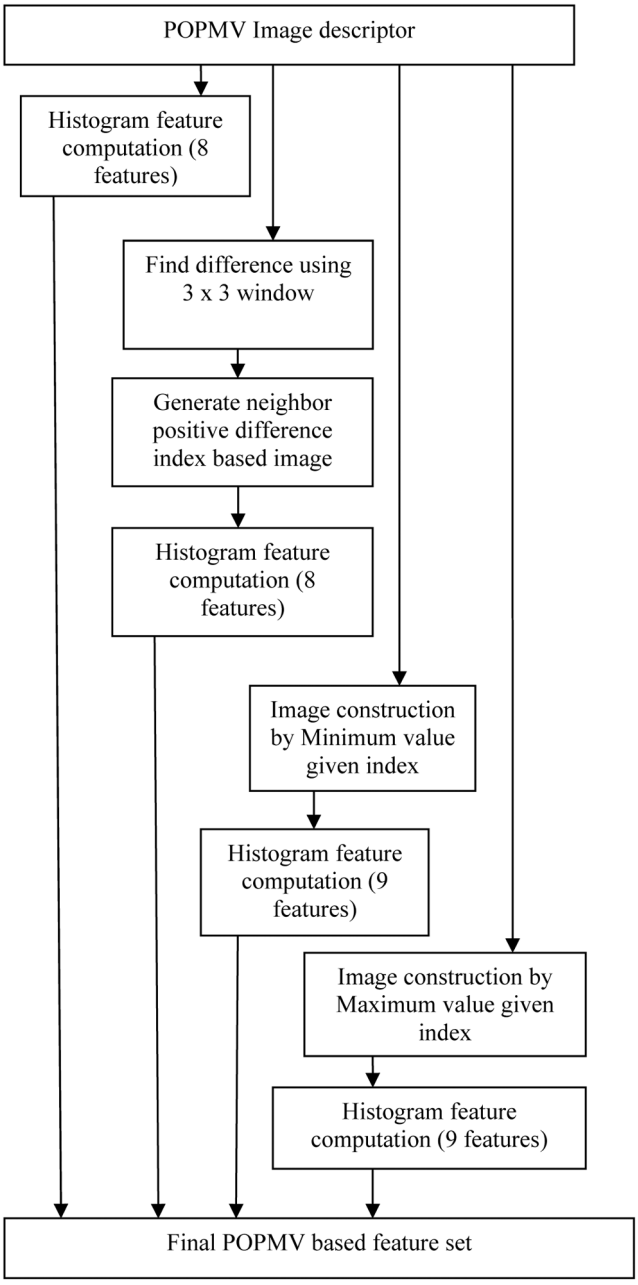

Fig. 8 *Block diagram for the computation of POPMV-based feature set generation*

are computed corresponding with the pixel $P_{i,j}$ and they are illustrated by (17)–(24):

$$D_0 = W_{0,0} - W_{1,1} \tag{17}$$

$$D_1 = W_{0,1} - W_{1,1} \tag{18}$$

$$D_2 = W_{0,2} - W_{1,1} \tag{19}$$

$$D_3 = W_{1,0} - W_{1,1} \tag{20}$$

$$D_4 = W_{1,2} - W_{1,1} \tag{21}$$

$$D_5 = W_{2,0} - W_{1,1} \tag{22}$$

$$D_6 = W_{2,1} - W_{1,1} \tag{23}$$

$$D_7 = W_{2,2} - W_{1,1} \tag{24}$$

where D is the difference value array.

The Difference Category is computed using (25):

$$\mathrm{DC}_K = \begin{cases} 0, & \text{if } D_K \in [-255,\ -c] \\ 1, & \text{if } D_K \in [-c+1,\ -b] \\ 2, & \text{if } D_K \in [-b+1,\ -a] \\ 3, & \text{if } D_K \in [-a+1,\ -1] \\ 4, & \text{if } D_K \in [0,\ a] \\ 5, & \text{if } D_K \in [a+1,\ b] \\ 6, & \text{if } D_K \in [b+1,\ c] \\ 7, & \text{if } D_K \in [c+1,\ 255] \end{cases} \tag{25}$$
$$K \in [0, 7]$$

where DC is the difference category array

The majority data in *difference category array* (DC) is calculated and stored in the POPMV feature image using (26):

$$\mathrm{POPMV}_{i,j} = \text{Majority (DC)} \tag{26}$$

where $\mathrm{POPMV}_{i,j}$ is the POPMV feature image.

Suppose the majority computation function is conflicted by multicompetitive candidates issue, then the averaging process on competitive candidates can be considered for majority computation. The averaged majority value is rounded to get an integer between 0 and 7. For example, if DC array is set as {3, 3, 3, 3, 6, 6, 6, 6} then the average value $(3+6)/2 = 4.5$ is fixed as the majority value. The numeric value 4.5 is further rounded to 5 for the final majority value. The POPMV feature image coincides with the dimension of the original image. The sample difference category computation of the POPMV descriptor for a pixel of $P_{i,j}$ in banana1 image is given in Appendix.

2.2 POPMV-based feature set generation

The feature sets for CBIR are generated from the POPMV image descriptor as shown in Fig. 8. Four feature sets are formulated from the POPMV image descriptor based on the histogram concept.

2.2.1 Majority category feature set: The data range of the majority category image is between 0 and 7. As the histogram is constructed using the majority category image, the length of this feature set is 8. This is described by (27) and (28):

$$H_{\mathrm{M}}(K) = 0, \quad K \in [0,\ 7] \tag{27}$$

$$\begin{aligned} H_{\mathrm{M}}(\mathrm{POPMV}_{i,j}) = {} & H_{\mathrm{M}}(\mathrm{POPMV}_{i,j}) + 1 \\ & i \in [0,\ IH] \\ & j \in [0,\ IW] \end{aligned} \tag{28}$$

where H_{M} is the histogram of majority categories.

The histogram $H_{\mathrm{M}}^{P \in [0,7]}$ is considered as the majority category feature set.

2.2.2 Neighbour positive difference oriented feature set: This feature is calculated using a 3×3 size window W centred by the pixel $P_{i,j}$ based on (16). The location indexes are mapped as shown in Fig. 9. In this figure, *italic* text notations are *location indexes* and the bold text notations are *window pixel coordinates*.

The differences D_0, D_1, D_2, D_3, D_4, D_5, D_6 and D_7 are computed using the equations (29)–(36):

$$D_0 = W_{0,0} - W_{1,1} \tag{29}$$

$$D_1 = W_{0,1} - W_{1,1} \tag{30}$$

$$D_2 = W_{0,2} - W_{1,1} \tag{31}$$

$$D_3 = W_{1,2} - W_{1,1} \tag{32}$$

$$D_4 = W_{2,2} - W_{1,1} \tag{33}$$

$$D_5 = W_{2,1} - W_{1,1} \tag{34}$$

$$D_6 = W_{2,0} - W_{1,1} \tag{35}$$

$$D_7 = W_{1,0} - W_{1,1} \tag{36}$$

The neighbour positive difference index-based image is found using (37) and (38):

$$\mathrm{IMG}_{\mathrm{NPDI}}^{i,j} = \sum_{k=0}^{7} f(D_k) \tag{37}$$

$$f(D_k) = \begin{cases} 2^k, & \text{if } D_k \geq 0 \\ 0, & \text{else} \end{cases} \tag{38}$$
$$k \in [0,\ 7]$$

where $\mathrm{IMG}_{\mathrm{NPDI}}$ is the neighbour positive difference index-based image, $f(D_k)$ is the function to compute the positive difference.

The index values are between 0 and 7, so the 2^k term [in (38)] generates values between 0 and 255. $\mathrm{IMG}_{\mathrm{NPDI}}$ contains the minimum value as 0 and the maximum value as 255. The histogram computation is performed on the image $\mathrm{IMG}_{\mathrm{NPDI}}$ using (39) and (40):

$$H_{\mathrm{NPDI}}(k) = 0, \quad k \in [0,\ q-1] \tag{39}$$

$$H_{\mathrm{NPDIC}}\left(IMG_{\mathrm{NPDI}}^{i,j}\right) = H_{\mathrm{NPDIC}}\left(IMG_{\mathrm{NPDI}}^{i,j}\right) + 1 \quad i \in [0, IH] \quad j \in [0, IW] \tag{40}$$

where H_{NPDI} is the neighbour positive difference-based histogram count, q is the length of the histogram (set as 256).

H_{NPDI} produces a 256 length histogram. Hence the length of the features is seemed to be abundant, the histogram is split up into eight bins (each bin has 32 elements). The majority value of each bin is stored as extracted features for the corresponding bin which results totally eight features to represent the whole image $\mathrm{IMG}_{\mathrm{NPDI}}$. This bin division process is performed to reduce the feature size. The new histogram $H_{\mathrm{NPDI}}^{q \in [0,7]}$ is known as neighbour positive difference oriented feature set.

2.2.3 Minimum oriented feature set: This feature computes its feature values based on the minimum number occurrence index in a 3×3 window W centred by $P_{i,j}$. The minimum value computation is done through (41):

$$\mathrm{IMG}_{\mathrm{MINI}}^{i,j} = f_{\mathrm{MINI}}(W) \tag{41}$$

where $\mathrm{IMG}_{\mathrm{MINI}}$ is the minimum given index-based image, f_{MINI} is the function to find minimum given index (between 0 and 8).

The histogram H_{MIN} is computed with the length 9 because the index marking lies between 0 and 8. The histogram $H_{\mathrm{MIN}}^{r \in [0,8]}$ is known as the minimum oriented feature set.

2.2.4 Maximum oriented feature set: This feature is constructed with the relationship of the maximum number occurrence index in a 3×3 size window W centred by $P_{i,j}$. The maximum value index is found via (42):

$$\mathrm{IMG}_{\mathrm{MAXI}}^{i,j} = f_{\mathrm{MAXI}}(W) \tag{42}$$

where $\mathrm{IMG}_{\mathrm{MAXI}}$ is the maximum given index-based image, f_{MAXI} is the function to find the maximum given index (between 0 and 8).

The image $\mathrm{IMG}_{\mathrm{MAXI}}$ is used to compute the H_{MAX} histogram which has the length 9 because the index lies in between 0 and 8.

0 **0,0**	*1* **0,1**	*2* **0,2**
7 **1,0**	**$P_{i,j}$** **1,1**	*3* **1,2**
6 **2,0**	*5* **2,1**	*4* **2,2**

Fig. 9 *Mapping of location index*

This histogram $H_{\mathrm{MAX}}^{s \in [0,8]}$ is known as the maximum oriented feature set.

2.3 DWT-based Entropy feature set

The DWT yields highly compacted energy information about the input image [28]. In this proposed work, the DWT process is performed through a 2D process based on the Daubechies wavelet family. The DWT [29] transform splits the image into four sub-images which are named as Low–Low (LL), Low–High (LH), High–Low (HL) and High–High (HH). The LH, HL and HH sub-bands are inbuilt with edge-oriented energies with themselves. These edges represent the shape features. The LL sub-band is omitted for further process because it does not have any shape-oriented information and has only energy-related data. The Entropy-based feature set is computed individually in the LH, HL and HH sub-bands using (43)–(45). Fig. 10 explores the block diagram of the Entropy feature set computation:

$$E_{\mathrm{LH}} = -\sum \mathrm{LH}^{p,q} * \log_2(\mathrm{LH}^{p,q}) \quad p \in [0,\ \mathrm{LH\_H} - 1] \quad q \in [0,\ \mathrm{LH\_W} - 1] \tag{43}$$

$$E_{\mathrm{HL}} = -\sum \mathrm{HL}^{p,q} * \log_2(\mathrm{HL}^{p,q}) \quad p \in [0,\ \mathrm{HL\_H} - 1] \quad q \in [0,\ \mathrm{HL\_W} - 1] \tag{44}$$

$$E_{\mathrm{HH}} = -\sum \mathrm{HH}^{p,q} * \log_2(\mathrm{HH}^{p,q}) \quad p \in [0,\ \mathrm{HH\_H} - 1] \quad q \in [0,\ \mathrm{HH\_W} - 1] \tag{45}$$

where E_{LH} is the Entropy feature for LH sub-band, E_{HL} is the Entropy feature for HL sub-band, E_{HH} is the Entropy feature for HH sub-band

The LH_H, HL_H and HH_H are the height of the corresponding sub-bands. The LH_W, HL_W and HH_W are the widths of the corresponding sub-bands. This feature set contains three values, and from those values, the DWT-based Entropy feature set is extracted as $F_{\mathrm{ENT}} = \{E_{\mathrm{LH}},\ E_{\mathrm{HL}},\ E_{\mathrm{HH}}\}$.

2.4 Statistical feature set

Statistical features are defined as the features derived from the whole or block of an image through the spatial domain using mathematical statistic function. In this paper, a statistic-based feature is used to improve the retrieval accuracy of the proposed CBIR system. Colour space is a specific organisation of colours in combination with the physical device profiling. It allows a reproducible representation of colours in both analogue and digital representation [30]. Here four colour spaces such as RGB, HSV, *YCbCr* and NTSC are used to derive the statistical features like *Mean*, *Standard deviation* (*Std*), *Entropy* and *Skewness*.

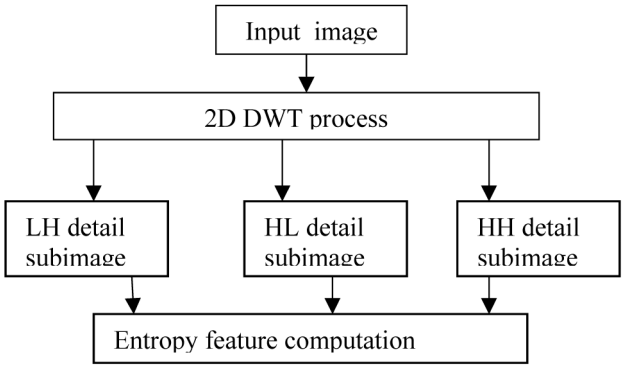

Fig. 10 *Block diagram of entropy feature set computation*

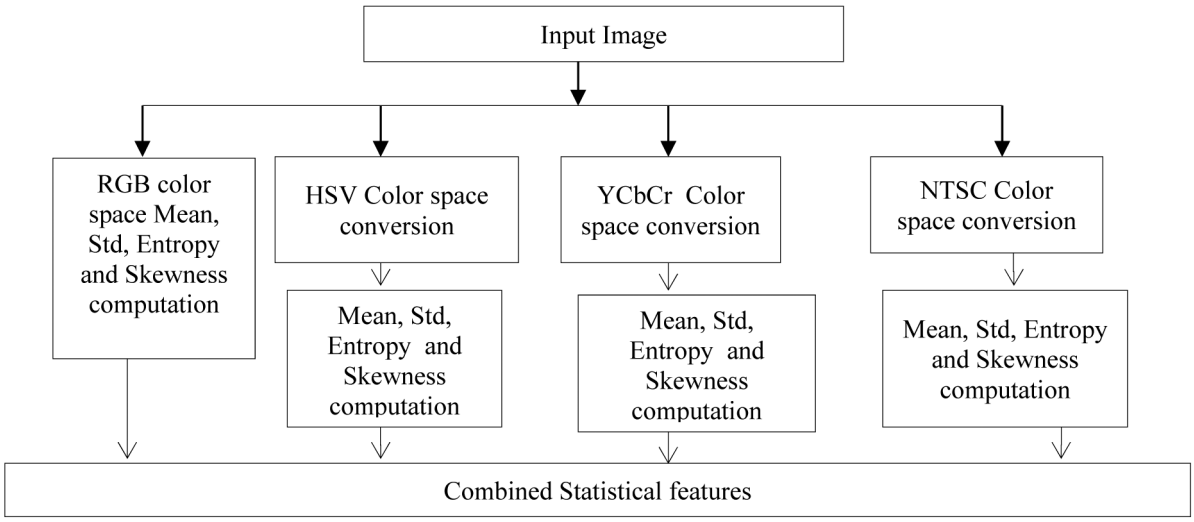

Fig. 11 *Block diagram of statistical feature computation*

One of the desirable characteristics of an appropriate colour space is its uniformity. Uniformity means that two colour pairs that are equal in similarity distance in a colour space are perceived as equal by viewers. In general, the measured proximity among the colours must be directly related to the psychology similarity among them. These four colour space integration may increase the computational cost, so the input image dimension is resized to 64 × 64. The image which is shrunk is used to employ the four colour spaces.

Fig. 11 explains the block diagram of the statistical feature computation in the proposed work. The RGB colour space is implemented in 24-bit implementation with 256 discrete levels of colours per channel. There are three colour channels in the RGB colour model such as Red, Green and Blue [30].

The Red, Green and Blue channels describe what kind of light needs to be emitted to produce a given colour. In this colour space, the importance is given to colour and not for shapes.

In this work, four statistical features are derived from the above mentioned channels (Red, Green and Blue) and totally 12 features are computed with reference to the RGB colour space. The mean statistical feature from the RGB colour space can be derived based on (46):

$$S_{\mathrm{RGB}}^{c,f=0} = \frac{1}{\mathrm{IH} * \mathrm{IW}} \sum_{i=0}^{\mathrm{IH}-1} \sum_{j=0}^{\mathrm{IW}-1} \mathrm{IMG}_{\mathrm{RGB}}^{c}(i,j) \quad c \in [0,1,2] \tag{46}$$

where $\mathrm{IMG}_{\mathrm{RGB}}^{c}$ is the Red, Green or Blue colour channel of an image corresponding with c, $S_{\mathrm{RGB}}^{c,f=0}$ is the RGB colour space statistical feature with respect to Mean feature ($f=0$), f is the feature type, c is the channel type, $c=0$ means Red channel, $c=1$ means Green channel, $c=2$ means Blue channel.

The Std-based statistical feature can be computed from the RGB colour space using (47):

$$S_{\mathrm{RGB}}^{c,f=1} = \sqrt{\frac{1}{\mathrm{IH} * \mathrm{IW}} \sum_{i=0}^{\mathrm{IH}-1} \sum_{j=0}^{\mathrm{IW}-1} \left(\mathrm{IMG}_{\mathrm{RGB}}^{c}(i,j) - S_{\mathrm{RGB}}^{c,f=0}\right)^2} \tag{47}$$

where $S_{\mathrm{RGB}}^{c,f=1}$ is the RGB colour space statistical feature with respect to Std feature ($f=1$).

Entropy is a statistical measure of randomness that can be used to characterise the texture of the input image. The Entropy feature for the RGB colour space can be derived using (48):

$$S_{\mathrm{RGB}}^{c,f=2} = -\sum_{i=0}^{\mathrm{IH}-1} \sum_{j=0}^{\mathrm{IW}-1} \mathrm{IMG}_{\mathrm{RGB}}^{c}(i,j) * \log(\mathrm{IMG}_{\mathrm{RGB}}^{c}(i,j)) \tag{48}$$

where $S_{\mathrm{RGB}}^{c,f=2}$ is the RGB colour space statistical feature with respect to Entropy feature ($f=2$).

The skewness is a measure of the asymmetry of data around the sample mean. If skewness is negative, the data are spread out more to the left of the *mean* than to the right. If skewness is positive, the data are spread out more to the right. The skewness of the normal distribution is known as zero. The skewness feature can be computed using (49):

$$S_{\mathrm{RGB}}^{c,f=3} = \frac{\frac{1}{\mathrm{IH} * \mathrm{IW}} \sum_{i=0}^{\mathrm{IH}-1} \sum_{j=0}^{\mathrm{IW}-1} \left(\mathrm{IMG}_{\mathrm{RGB}}^{c}(i,j) - S_{\mathrm{RGB}}^{c,f=0}\right)^2}{\sqrt{\frac{1}{\mathrm{IH} * \mathrm{IW}} \sum_{i=0}^{\mathrm{IH}-1} \sum_{j=0}^{\mathrm{IW}-1} \left(\mathrm{IMG}_{\mathrm{RGB}}^{c}(i,j) - S_{\mathrm{RGB}}^{c,f=0}\right)^2}} \tag{49}$$

where $S_{\mathrm{RGB}}^{c,f=3}$ is the RGB colour space statistical feature with respect to Skewness feature ($f=3$).

The HSV colour space is a transformation of a Cartesian RGB colour space and it includes the components as Hue, Saturation and Value. The RGB to HSV colour conversion can be obtained from the reference paper [31, 32]. The HSV statistical features $S_{\mathrm{HSV}}^{c,f=0}$, $S_{\mathrm{HSV}}^{c,f=1}$, $S_{\mathrm{HSV}}^{c,f=2}$ and $S_{\mathrm{HSV}}^{c,f=3}$ with respect to Mean, Std, Entropy and Skewness are computed using (46)–(49) with the colour channels – Hue, Saturation and Value instead of Red, Green and Blue channels. The Hue is invariant to the changes in illumination and camera direction and hence more suited to object or image retrieval [33].

The *YCbCr* colour space is used in the video and digital photography system with perceptually meaningful information where *Y* is the luminance and *Cb* is the blue-based chrominance and *Cr* is the red-based chrominance component [34]. The sky like blue components are strongly represented by *Cb* components and the reddish regions are well represented by *Cr* components. The formulae for converting from RGB to *YCbCr* are given below:

$$Y = (77/256)R + (150/256)G + (29/256)B \tag{50}$$

$$Cb = (44/256)R - (87/256)G + (131/256)B + 128 \tag{51}$$

$$Cr = (131/256)R - (110/256)G - (21/256)B + 128 \tag{52}$$

The *YCbCr* statistical features $S_{YCbCr}^{c,f=0}$, $S_{YCbCr}^{c,f=1}$, $S_{YCbCr}^{c,f=2}$ and $S_{YCbCr}^{c,f=3}$ which are related with Mean, Std, Entropy and Skewness are calculated based on (46)–(49). The *YCbCr* colour space represents a more shape-oriented statistical feature. Owing to this feature, the database images which have unique shape are retrieved more successfully.

The National Television System Committee (NTSC) colour space is used in the colour display device such as TV [35]. The NTSC colour space includes luminance (for monochrome display), hue and saturation. The RGB to NTSC conversion can be performed using the reference [35]. The NTSC colour space-oriented Mean feature $S_{\mathrm{NTSC}}^{c,f=0}$, Std feature $S_{\mathrm{NTSC}}^{c,f=1}$, Entropy feature $S_{\mathrm{NTSC}}^{c,f=2}$ and Skewness feature $S_{\mathrm{NTSC}}^{c,f=3}$ are found based on (46)–(49). The final statistical feature set F_{STAT} is found using (53). The NTSC colour space is used to improve the retrieval rate by comparing more colour-related data.

$$\begin{aligned} F_{\mathrm{STAT}} = & \left\{\left\{S_{\mathrm{RGB}}^{c,f}\right\}, \left\{S_{\mathrm{HSV}}^{c,f}\right\}\left\{S_{YCbCr}^{c,f}\right\}, \left\{S_{\mathrm{NTSC}}^{c,f}\right\}\right\} \\ & c \in [0,1,2] \\ & f \in [0,1,2,3] \end{aligned} \tag{53}$$

In the RGB colour space, 12 statistical features are extracted because of the availability of three colour channels and four kinds of statistical functions. Like this, the HSV, *YCbCr* and NTSC colour channels are also granted 12 features per colour space. There are totally 48 features obtained from four colour spaces and they are integrated in (53). The colour statistical features (colour moments) have been successfully used in many image retrieval systems [33, 36] especially when the image contains a major object. The colour statistical features have been proved to be efficient and effective in representing colour distributions of images [37]. The colour moments are the very compact representations compared to other colour features.

2.5 Image retrieval

This research uses three types of features such as POPMV feature set, DWT-based Entropy feature set and statistical feature set for image retrieval. The query image is given as input and the extracted feature sets are combined using (54) to get the final query feature set F_{QUERY}.

$$F_{\mathrm{QUERY}} = \left\{\begin{array}{c} \left\{H_{\mathrm{M}}^{P\in[0,7]}\right\}, \left\{H_{\mathrm{NPDI}}^{q\in[0,7]}\right\}, \left\{H_{\mathrm{MIN}}^{r\in[0,8]}\right\}, \left\{H_{\mathrm{MAX}}^{s\in[0,8]}\right\}, \left\{F_{\mathrm{ENT}}^{t\in[0,2]}\right\}, \\ \left\{F_{\mathrm{STAT}}^{u\in[0,47]}\right\} \end{array}\right\} \tag{54}$$

The image database contains *n* (assume) images and the three types of features such as POPMV feature set, DWT-based Entropy feature set and statistical feature set are extracted from each database images and it can be represented based on (55). The term DB represents the database.

$$\begin{aligned} F_{\mathrm{DBF}}^{i,j} = & \left\{\begin{array}{c} \left\{H_{\mathrm{DB_M}}^{P\in[0,7]}\right\}, \left\{H_{\mathrm{DB_{NPDI}}}^{q\in[0,7]}\right\}, \left\{H_{\mathrm{DB_{MIN}}}^{r\in[0,8]}\right\}, \left\{H_{\mathrm{DB_{MAX}}}^{s\in[0,8]}\right\}, \\ \left\{F_{\mathrm{DB_{ENT}}}^{t\in[0,2]}\right\}, \left\{F_{\mathrm{DB_{STAT}}}^{u\in[0,47]}\right\} \end{array}\right\} \\ & i \in [0, n-1] \\ & j \in [0, T-1] \end{aligned} \tag{55}$$

where F_{DBF} is the final database feature set, *T* is the total features ($T=85$).

Similarity measurement is a significant issue in CBIR. The similarity between the query image and the database images is calculated by finding the difference between the query feature vector and the database feature vectors by using distance metrics. The images with small distances are most similar to the query images. The Euclidean distance is a similarity measurement method used by most of the CBIR researches ([4, 38, 39]) because of its efficiency and image retrieval power. The Euclidean distance between the query features and the database image features is computed based on (56):

$$\begin{aligned} \mathrm{Dist}_{\mathrm{EQ}}^{i} = & \sqrt{\sum_{j\in[0,T-1]} \left(F_{\mathrm{QUERY}}^{j} - F_{\mathrm{DBF}}^{i,j}\right)^2} \\ & i \in [0, n-1] \end{aligned} \tag{56}$$

where $\mathrm{Dist}_{\mathrm{EQ}}^{i}$ is the Euclidean distance of the *i*th image of the database and the query image.

The $\mathrm{Dist}_{\mathrm{EQ}}$ values have undergone ascending sorting, and based on those results, the sorted image index is obtained. The lowest distance score indicates the most similar image.

3 Experiment results and analysis

This paper proposes a hybrid image retrieval system based on the features of statistical, DWT-Entropy and POPMV descriptor. This paper is implemented using MATLAB 10.0 and the assessment evaluation of the proposed method is designed using the following existing algorithms.

- Image retrieval using Biased Discriminant Euclidean Embedding (BDE) [17].
- Image retrieval using Neighbourhood Discriminates Hashing (NDH) [13].
- Image retrieval using EDBTC [14].
- Image retrieval using MDLBP [3].
- Image retrieval using Robust Aggregation of Local Descriptors (RALD) [40].

This section uses seven standard databases containing the natural scene, fruits, vegetables, flowers, buildings, vehicles and textures etc. and the other one is the user-contributed database named as DB_VEG containing full of vegetable images. Table 1 shows the information about the databases such as the total number of images, number of categories, number of images in each category. Table 2 explores the feature dimensionality occupied by different methods.

Fig. 12 expresses the sample test images used by this paper for the categories Corel-1k and DB_VEG. Figs. 13 and 14 depict the results of image retrieval for the query image from Corel-1k using the proposed CBIR_SWPOPMV method for ten retrieved images

Table 1 Database image information

Database name	No. of categories	No. of images per category	Total no. images
Corel-1k [41]	10	100	1000
USPTex [42]	191	12	2292
MIT-VisTex [43]	40	16	640
KTH-TIPS [44]	10	81	810
KTH-TIPS2a [44]	11	396	4608
Corel-Tex [41]	6	100	600
Colored Brodatz [45]	112	256	28,672
DB_VEG	14	15	210

Table 2 Feature dimensions

Methods	Feature dimension
BDE	728
NDH	128
BTC	256
MDLBP	8 × 256
RALD	256
proposed	85

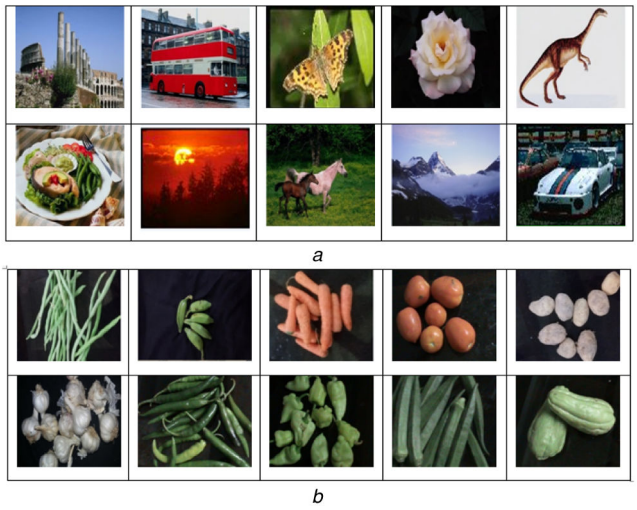

Fig. 12 *Sample images of databases used*
(a) Corel-1k, ***(b)*** DB_VEG

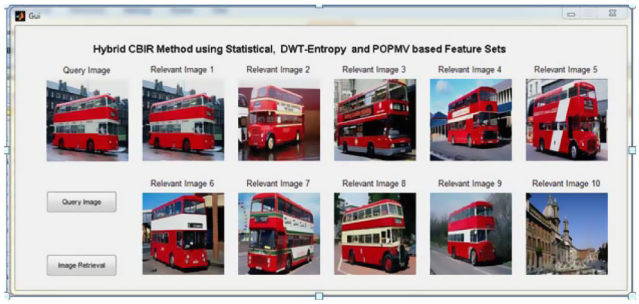

Fig. 13 *Image retrieval results for Corel-1k bus query image using the proposed method*

and Figs. 15 and 16 represent the image retrieval results for ten retrieved images from DB_VEG for the proposed CBIR_SWPOPMV method.

The effectiveness of the proposed approach can be experimentally examined by calculating the number of relevant images retrieved for the particular query image.

In this work, the proposed system's effectiveness for a variety of databases is estimated by computing the precision rate. Precision is the fraction of relevant images among the retrieved images. The method which produces higher precision value is considered as the

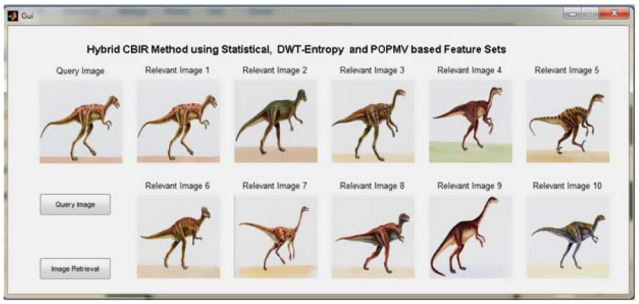

Fig. 14 *Image retrieval results for Corel-1k dinosaur query image using the proposed method*

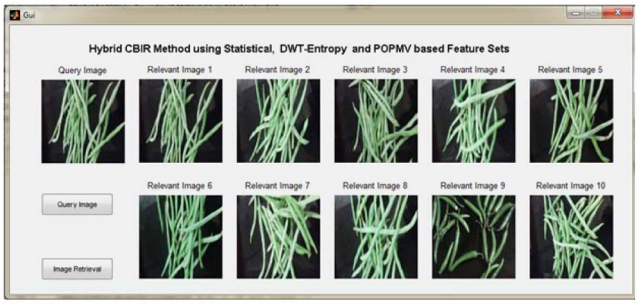

Fig. 15 *Image retrieval results for DB_VEG beans query image using the proposed method*

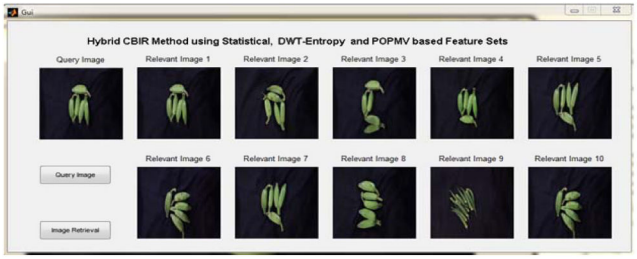

Fig. 16 *Image retrieval results for DB_VEG banana query image using the proposed method*

best method. In this work, the first ten retrieved images are used to calculate the precision rate based on (57):

$$\text{Precision} = \frac{\text{Number of Similar Images Retrieved}}{\text{Total Number of Images Retrieved}} \quad (57)$$

Table 3 reports the precision rate for ten categories of query images from Corel-1k database for ten retrieved images. Table 3 shows the proposed method's performance over the existing methods such as BDE, NDH, BTC, MDLBP and RALD. In this table, the proposed method gives 100% result for dinosaur category because the dinosaur images have a smooth background. For other categories of images also, the proposed method gives better results than the existing methods.

Similarly, the precision rate for ten retrieved images from the DB_VEG database is displayed in Table 4. Table 4 shows the assessment of the proposed method which gives a higher precision rate of 0.95 for banana because this category images show clear texture gap.

Another method used in this paper for calculating the effectiveness of the proposed work is *Recall*. Recall is the fraction of relevant images that have been retrieved over total relevant images in the whole database. The higher recall value producing method denotes the best method. In this work, the recall rate is estimated by using the first ten retrieved images based on (58):

$$\text{Recall} = \frac{\text{Number of Similar Images Retrieved}}{\text{Number of Similar Images in the Whole Database}} \quad (58)$$

Table 3 Precision analysis for Corel-1k database

Categories	BDE	NDH	BTC	MDLBP	RALD	Proposed
dinosaur	0.91	0.91	0.91	0.93	0.97	1.00
mountain	0.53	0.54	0.61	0.65	0.69	0.76
rose	0.55	0.61	0.65	0.67	0.71	0.81
building	0.61	0.64	0.68	0.72	0.75	0.87
bus	0.68	0.72	0.74	0.78	0.85	0.92
elephants	0.61	0.65	0.69	0.71	0.75	0.83
food	0.64	0.65	0.71	0.75	0.82	0.92
horses	0.58	0.61	0.65	0.72	0.76	0.80
Africans	0.53	0.58	0.58	0.62	0.64	0.70
beaches	0.65	0.68	0.68	0.78	0.83	0.91

Table 4 Precision analysis for DB_VEG database

Categories	BDE	NDH	BTC	MDLBP	RALD	Proposed
banana	0.67	0.72	0.77	0.81	0.85	0.95
potato	0.44	0.48	0.51	0.55	0.62	0.80
carrot	0.65	0.72	0.76	0.81	0.85	0.93
tomato	0.61	0.67	0.73	0.76	0.68	0.91
beans	0.62	0.66	0.71	0.77	0.83	0.93
brinjal	0.51	0.59	0.64	0.69	0.73	0.85
onion	0.48	0.56	0.61	0.66	0.70	0.79
chilly	0.43	0.48	0.56	0.63	0.71	0.83
dark chilly	0.46	0.54	0.62	0.65	0.77	0.88
cluster beans	0.40	0.46	0.50	0.56	0.67	0.75
chow-chow	0.41	0.52	0.56	0.61	0.65	0.74
green brinjal	0.45	0.54	0.58	0.63	0.69	0.82
garlic	0.47	0.52	0.57	0.65	0.75	0.87

Table 5 Recall analysis for Corel-1k database

Categories	BDE	NDH	BTC	MDLBP	RALD	Proposed
dinosaur	0.091	0.091	0.091	0.093	0.097	0.100
mountain	0.053	0.054	0.061	0.065	0.069	0.076
rose	0.055	0.061	0.065	0.067	0.071	0.081
building	0.061	0.064	0.068	0.072	0.075	0.087
bus	0.068	0.072	0.074	0.078	0.085	0.092
elephants	0.061	0.065	0.069	0.071	0.075	0.083
food	0.064	0.065	0.071	0.075	0.082	0.092
horses	0.058	0.061	0.065	0.072	0.076	0.080
Africans	0.053	0.058	0.058	0.062	0.064	0.070
beaches	0.065	0.068	0.068	0.078	0.083	0.091

The recall rate computed for Corel-1k and DB_VEG databases is shown in Tables 5 and 6. Similar with precision, for recall also *dinosaur* query image produces higher recall rate of 0.1 compared to the other methods which are tabulated in Table 5; but *africans* images produce low recall rate of 0.070 because of complex and overlapped textures of the images. The recall rate for DB_VEG database is also reported in Table 6. The *banana* category gives a higher recall rate of 0.63 but the recall value for the vegetables *beans* and *carrot* is 0.62. As the *cluster beans* and *chow-chow* images are not visually clear and the textures are not defined specifically, they produce 0.50 and 0.49 recall rate. For both databases, the proposed method gives a higher recall value than the existing methods. Thus, the proposed method's superiority is proved.

The time cost analysis is reported in Fig. 17 for Corel-1k database. The table shows that the time taken by the proposed method is less than the recently published methods RALD and MDLBP, so the performance of the proposed method is better than the other methods. Similarly Fig. 18 displays the time cost analysis for DB_VEG database. Similar to Corel-1k database, the proposed method's time cost analysis is less than the RALD and MDLBP methods.

Table 7 depicts the Average Retrieval Precision (ARP)-related comparative study. From the table, it can be noticed with ease that the proposed method shines with higher ARP while other methods lack in that. The second best high performance (in ARP) method is RALD. The least ARP given method is BDE. The higher ARP given dataset is Colored Brodatz.

Similarly Table 8 represents the Average Retrieval Recall (ARR) analysis for ten retrieved images and the proposed method proves its superiority on ARR analysis on entire databases. From the table, it can be identified that the proposed method's ARR value is higher for USPTex database and low for KTH-TIPS2a database. The second best high-performance (in ARR) method is RALD.

Fig. 19 also represents the Average Time Taken Analysis. Fig. 19 proves that the proposed method takes less time for all databases mentioned in the table compared to other methods. The second best method in terms of the average time taken is NDH.

The precision and recall analysis expresses the accuracy in two values. The complete accuracy can be computed in a single value by combining precision and recall for the effective image retrieval. This single value measurement is known as *F*-Score or *F*-Measure which is referred as the harmonic mean of the precision and recall [46]. The *F*-Score can be defined as

Table 6 Recall analysis for DB_VEG database

Categories	BDE	NDH	BTC	MDLBP	RALD	Proposed
banana	0.45	0.48	0.51	0.54	0.57	0.63
potato	0.29	0.32	0.34	0.37	0.41	0.53
carrot	0.43	0.48	0.51	0.54	0.57	0.62
tomato	0.41	0.45	0.49	0.51	0.45	0.61
beans	0.41	0.44	0.47	0.51	0.55	0.62
brinjal	0.34	0.39	0.43	0.46	0.49	0.57
onion	0.32	0.37	0.41	0.44	0.47	0.53
chilly	0.29	0.32	0.37	0.42	0.47	0.55
dark chilly	0.31	0.36	0.41	0.43	0.51	0.59
cluster beans	0.27	0.31	0.33	0.37	0.45	0.50
chow-chow	0.27	0.35	0.37	0.41	0.43	0.49
green brinjal	0.30	0.36	0.39	0.42	0.46	0.55
garlic	0.31	0.35	0.38	0.43	0.50	0.58

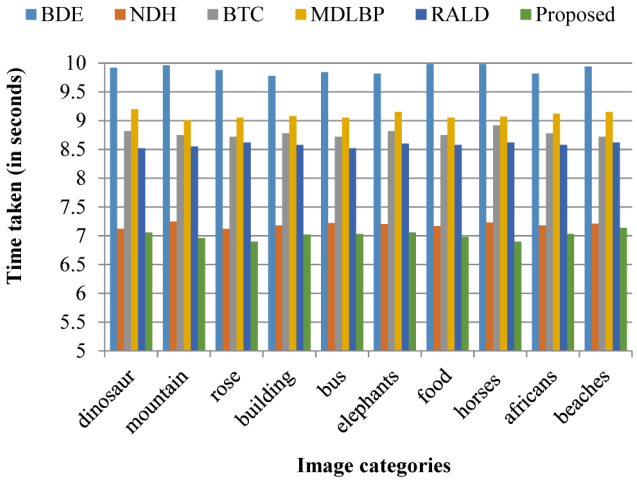

Fig. 17 *Time cost analysis for Corel-1k database*

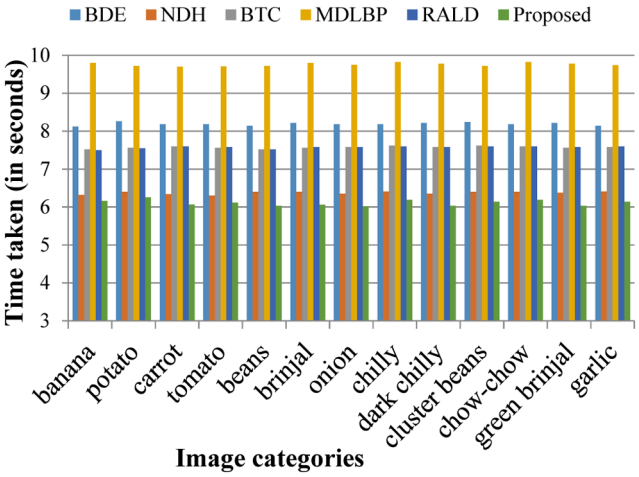

Fig. 18 *Time cost analysis for DB_VEG database*

Table 7 Average retrieval precision (ARP) analysis

Database	BDE	NDH	BTC	MDLBP	RALD	Proposed
Corel-1k	0.629	0.659	0.690	0.733	0.777	0.852
MIT-VisTex	0.439	0.559	0.670	0.708	0.733	0.801
USPTex	0.617	0.650	0.702	0.773	0.810	0.897
KTH-TIPS	0.432	0.682	0.690	0.702	0.752	0.887
KTH-TIPS2a	0.447	0.662	0.712	0.732	0.751	0.857
Corel-Tex	0.432	0.458	0.502	0.571	0.617	0.721
Colored Brodatz	0.552	0.672	0.692	0.702	0.752	0.902
DB_VEG	0.508	0.574	0.625	0.675	0.731	0.850

Table 8 Average retrieval recall (ARR) analysis

Database	BDE	NDH	BTC	MDLBP	RALD	Proposed
Corel-1k	0.063	0.066	0.069	0.073	0.078	0.085
USPTex	0.047	0.049	0.054	0.055	0.063	0.075
MIT-VisTex	0.350	0.356	0.400	0.406	0.469	0.519
KTH-TIPS	0.028	0.029	0.032	0.032	0.036	0.039
KTH-TIPS2a	0.014	0.018	0.016	0.016	0.019	0.021
Corel-Tex	0.055	0.052	0.072	0.076	0.080	0.090
Colored Brodatz	0.022	0.022	0.025	0.025	0.029	0.032
DB_VEG	0.341	0.386	0.419	0.452	0.486	0.566

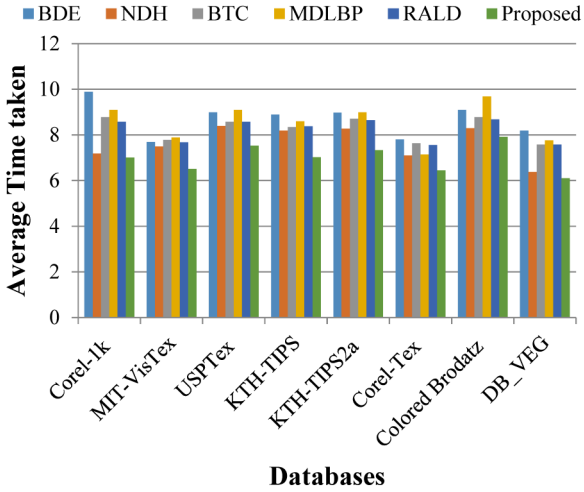

Fig. 19 *Average time taken analysis*

Table 9 *F*-score analysis for Corel-1k database

Categories	BDE	NDH	BTC	MDLBP	RALD	Proposed
dinosaur	0.165	0.165	0.165	0.169	0.176	0.182
mountain	0.096	0.098	0.111	0.118	0.125	0.138
rose	0.100	0.111	0.118	0.122	0.129	0.147
building	0.111	0.116	0.124	0.131	0.136	0.158
bus	0.124	0.131	0.135	0.142	0.155	0.167
elephants	0.111	0.118	0.125	0.129	0.136	0.151
food	0.116	0.118	0.129	0.136	0.149	0.167
horses	0.105	0.111	0.118	0.131	0.138	0.145
Africans	0.096	0.105	0.105	0.113	0.116	0.127
beaches	0.118	0.124	0.124	0.142	0.151	0.165

$$F = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{59}$$

The *F*-Score value acts as a single value descriptor for the overall effectiveness of the image retrieval. Table 9 represents the *F*-Score analysis for Corel-1K and Table 10 represents the *F*-Score analysis for DB_VEG database. Table 9 shows that the *dinosaur* category scores the highest *F*-Score value of 0.182 while *africans* category scores the low value of 0.127 and Table 10 shows that the *banana* category scores the highest *F*-Score value of 0.760.

The average precision versus recall analysis is indicated in Fig. 20 for Corel-1k database. It proves that the proposed method makes a plotting curve over the curve of RALD i.e. excellent performance due to the high gain in both the precision and recall terms. The chow-chow image produces worse retrieval results that indicate the potential drawbacks of this method.

4 Conclusion

In this paper, a hybrid CBIR method called CBIR_SWPOPMV is proposed to generate three kinds of image feature set. A new POPMV descriptor is proposed to generate the first type of feature set which comprises the majority category feature set, the neighbour positive difference oriented feature set, the minimum oriented feature set and the maximum oriented feature set. The second and third feature sets are generated by calculating DWT-based Entropy and statistical computation over the resized input image for different colour spaces, respectively. The performance of CBIR_SWPOPMV is analysed using precision, recall and run time against the subsistent methods such as BDE, NDH, BTC, MDLBP and RALD. The experimental results prove that the proposed method CBIR_SWPOPMV is superior over the existing methods. The proposed CBIR_SWPOPMV method has secured the higher average precision value of 0.852 and the average recall value of 0.085 for Corel-1k database. Moreover, this method requires less runtime than the existing methods. This preeminent method is fit for online image retrieval. In future, the neural network classifier such as ART can be integrated to support even more retrieval accuracy.

Table 10 *F*-score analysis for DB_VEG database

Categories	BDE	NDH	BTC	MDLBP	RALD	Proposed
banana	0.536	0.576	0.616	0.648	0.680	0.760
potato	0.352	0.384	0.408	0.440	0.496	0.640
carrot	0.520	0.576	0.608	0.648	0.680	0.744
tomato	0.488	0.536	0.584	0.608	0.544	0.728
beans	0.496	0.528	0.568	0.616	0.664	0.744
brinjal	0.408	0.472	0.512	0.552	0.584	0.680
onion	0.384	0.448	0.488	0.528	0.560	0.632
chilly	0.344	0.384	0.448	0.504	0.568	0.664
dark chilly	0.368	0.432	0.496	0.520	0.616	0.704
cluster beans	0.320	0.368	0.400	0.448	0.536	0.600
chow-chow	0.328	0.416	0.448	0.488	0.520	0.592
green brinjal	0.360	0.432	0.464	0.504	0.552	0.656
garlic	0.376	0.416	0.456	0.520	0.600	0.696

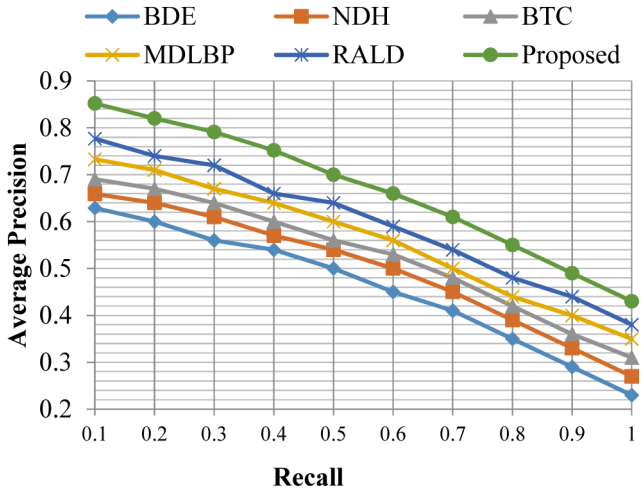

Fig. 20 *Average precision versus recall analysis*

5 References

[1] Furht, B.: '*Handbook of multimedia computing*' (CRC Press, Boca Raton, USA), p. 228

[2] Smeulders, A.W.M., Santini, S., Gupta, A., *et al.*: 'Content-based image retrieval at the end of the early years', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (12), pp. 1349–1380

[3] Dubey, S.R., Singh, S.K., Singh, R.K.: 'Multichannel decoded local binary patterns for content based image retrieval', *IEEE Trans. Image Process.*, 2016, **25**, (9), pp. 4018–4032

[4] Lai, C.-C., Chen, Y.-C.: 'A user-oriented image retrieval system based on interactive genetic algorithm', *IEEE Trans. Instrum. Meas.*, 2011, **60**, (10), pp. 3318–3325

[5] Dubey, S.R., Singh, S.K., Singh, R.K.: 'Local bit-plane decoded pattern: a novel feature descriptor for biomedical image retrieval', *IEEE. J. Biomed. Health. Inform.*, 2016, **20**, (4), pp. 1139–1147

[6] Li, Y., Zhang, D., Lu, G., *et al.*: 'A survey of content based image retrieval with high-level semantics', *Pattern Recognit.*, 2007, **40**, (1), pp. 262–282

[7] Sikora, T.: 'The MPEG-7 visual standard for content description – an overview', *IEEE Trans. Circuits Syst. Video Technol.*, 2001, **11**, (6), pp. 696–702

[8] Abbadeni, N.: 'Computational perceptual features for texture representation and retrieval', *IEEE Trans. Image Process.*, 2011, **20**, (1), pp. 236–246

[9] Ojala, T., Pietikainen, M., Maenpaa, T.: 'Multi resolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (7), pp. 971–987

[10] Persoon, E., Fu, K.: 'Shape discrimination using Fourier descriptors', *IEEE Trans. Syst. Man Cybern.*, 1997, **7**, (3), pp. 170–179

[11] Zhang, D., Lu, G.: 'Evaluation of MPEG-7 shape descriptors against other shape descriptors', *Multimedia Syst.*, 2003, **9**, (1), pp. 15–30

[12] Latecki, L.J., Lakamper, R., Eckhardt, T.: 'Shape descriptors for non-rigid shapes with a single closed contour', *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2000, **1**, pp. 424–429

[13] Tang, J., Li, Z., Wang, M., *et al.*: 'Neighborhood discriminant hashing for large-scale image retrieval', *IEEE Trans. Image Process.*, 2015, **24**, (9), pp. 2827–2840

[14] Guo, J.-M., Prasetyo, H., Chen, J.-H.: 'Content- based image retrieval using error diffusion block truncation coding features', *IEEE Trans. Circ. Syst. Video Technol.*, 2015, **25**, (3), pp. 466–481

[15] Sujatha, B., Kumar, V., Harini, P.: 'A new logical compact LBP co-occurrence matrix for texture analysis', *Int. J. Sci. Eng. Res.*, 2012, **3**, (2), pp. 1–5

[16] Kwitt, R., Meerwald, P., Uhi, A.: 'Efficient texture image retrieval using Copulas Bayesian Framework', *IEEE Trans. Image Process.*, 2011, **20**, (7), pp. 2063–2077

[17] Bian, W., Tao, D.: 'Biased Discriminant Euclidean Embedding for content-based image retrieval', *IEEE Trans. Image Process.*, 2010, **19**, (19), pp. 545–551

[18] Liu, G.-H., Yang, J.-Y.: 'Image retrieval based on the texton co-occurrence matrix', *Pattern Recognit.*, 2008, **41**, (12), pp. 3521–3527

[19] Ptino-Escarcina, R.E., Costa, J.A.F.: 'Content based image retrieval using a descriptors hierarchy'. Proc. IEEE Seventh Int. Conf. on Hybrid Intelligent Sys., Kaiserlautern, Germany, 2007, pp. 228–233

[20] Moghddam, A., Khajoie, T., Rouhi, A.: 'Wavelet correlogram: a new approach for image index retrieval', *Pattern Recognit.*, 2005, **38**, pp. 2506–2518

[21] Qi, X., Han, Y.: 'A novel fusion approach to content-based image retrieval', *Pattern Recognit.*, 2005, **38**, pp. 2449–2465

[22] Liu, G.-H., Li, Z.-Y., Zhang, L., *et al.*: 'Image retrieval based micro structure descriptor', *Pattern Recognit.*, 2011, **44**, (9), pp. 2123–2133

[23] Liu, Q., Huang, Y., Metaxas, D.N.: 'Hypergraph with sampling for image retrieval', *Pattern Recognit.*, 2011, **44**, (10–11), pp. 2255–2262

[24] Wang, Y., Shi, M., You, S., *et al.*: 'DCT inspired feature transform for image retrieval and reconstruction', *IEEE Trans. Image Process.*, 2016, **25**, (9), pp. 4408–4418

[25] Fadaei, S., Amirfattahi, R., Ahmadzadeh, M.R.: 'A new content-based image retrieval system based on optimized integration of DCD, wavelet and curvelet features', *IET Image Process.*, 2016, **11**, (2), pp. 89–98

[26] Zhu, L., Shen, J., Xie, L., *et al.*: 'Unsupervised visual hashing with semantic assistant for content-based image retrieval', *IEEE Trans. Knowl. Data Eng.*, 2017, **29**, (2), pp. 472–486

[27] Shi, L., Wang, H., Meng, J., *et al.*: 'Optimizing top precision performance measure of content based image retrieval by learning similarity function'. 23rd Int. Conf. on Pattern Recognition (ICPR), IEEE Conf., Cancun, Mexico, 2016, pp. 2954–2958

[28] Mistry, Y.D., Ingole, D.T.: 'Content-based image retrieval using DWT based feature extraction and texture, shape and color features', *Int. J. Res. Comput. Commun. Technol.*, 2014, **3**, (11), pp. 1510–1516

[29] Yadav, O., Suryawanshi, V.: 'CBIR evaluation using different distances and DWT', *Int. J. Comput. Appl.*, 2014, **93**, (16), pp. 36–40

[30] 'Color space types'. Available at: http:en.wikipedia.org/wiki/Color-Space (accessed 20 Feb 2018)

[31] 'HSV color space'. Available at: http:en.wikipedia.org/wiki/List-of-Color-space-and-their-uses-HSV-and-HSL (accessed 27 January 2018)

[32] Georgieva, L., Dimitrova, T., Angelov, N.: 'RGB and HSV color models in color identification of digital traumas images'. Int. Conf. on computer systems and Technology – CompSys Tech., Technical University, Varna, Bulgaria, 2005

[33] Foley, J.D., Dam, A.V., Feiner, J.F.: '*Hughes, computer graphics: principles and practice*' (Addison-Wesley, Boston, MA, USA, 1990, 2nd edn.)

[34] '*YCbcr* color space'. Available at: http:en.wikipedia.org/wiki/YCbCr (accessed 27 January 2018)

[35] 'Color space'. Available at: www.Compression.ru/download/articles/Color-space/ch03.pdf (accessed 02 February 2018)

[36] Flicker, M., Sawhney, H., Niblack, W., *et al.*: 'Query by image and video content: the QBIC system', *IEEE Comput.*, 1995, **28**, (9), pp. 23–32

[37] Stricker, O.: 'Similarity of color changes'. SPIE Storage and Retrieval for Image and Video Database III, San Jose, USA, 1995, vol. 2185, pp. 381–392

[38] Khosla, G., Rajpal, N., Singh, J.: 'Evaluation of Euclidean and Manhattan metrics in content based image retrieval system', *J. Eng. Res. Appl.*, 2014, **4**, (9), pp. 43–49

[39] Selvarasu, N., Nachiappan, A., Nandhitha, N.M.: 'Euclidean distance based color image segmentation of abnormality detection from pseudo color thermographs', *Int. J. Comput. Theory Eng.*, 2010, **2**, (4), pp. 514–516

[40] Husain, S., Bober, M.: 'Improving large-scale image retrieval through robust aggregation of local descriptors', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, **39**, (9), pp. 1783–1796

[41] 'Corel photo collection image database'. Available at: http://wang.ist.psu.edu/docs/realted/ (accessed 20 February 2018)

[42] Backes, A.R., Casanova, D., Bruno, O.M.: 'Color texture analysis based on fractal descriptors', *Pattern Recognit.*, 2012, **45**, (5), pp. 1984–1992
[43] 'MIT vision and modeling group, Cambridge'. Available at: http://vismod.media.mit.edu/pub/ (accessed 20 Feb 2018)
[44] 'KTH-TIPS texture image database', online Available at: http://www.nada.kth.se/cvap/databases/kth-tips/index.html (accessed 20 Feb 2018)
[45] 'Colored Brodatz database'. Available at: http://multibandtexture.recherche.usherbrooke.ca/colored%20_brodatz.html (accessed 20 Feb 2018)
[46] Mlik, F., Baharudin, B.: 'Analysis of distance metrics in content based image retrieval using statistical quantized histogram texture features in the DCT domains', *J. King Saud Univ., Comput. Inf. Sci.*, 2013, **25**, (2), pp. 207–208

6 Appendix

Sample category computation of the POPMV descriptor for a pixel of $P_{i,j}$ in banana1 image (Fig. 21).

The difference calculation order is followed by using Fig. 3:

$$Z_{\mathrm{DiFF}}^{1,k} = -34\ -17\cdot 1\ -19\cdot 11\ -5\cdot 8\cdot 13$$

$$Z_{\mathrm{DiFF}}^{2,k} = -23\ -12\cdot 4\ -17\cdot 18\ -2\cdot 17\cdot 28$$

$$Z_{\mathrm{DiFF}}^{3,k} = -4\ -6\cdot 5\ -4\cdot 17\ -3\cdot 15\cdot 34$$

$$Z_{\mathrm{DiFF}}^{4,k} = -1\cdot 0\ -2\ -1\cdot 4\ -1\cdot 1\cdot 19$$

$$Z_{\mathrm{DiFF}}^{5,k} = -2\ -1\cdot 3\ -2\cdot 18\cdot 0\cdot 14\cdot 32$$

$$Z_{\mathrm{DiFF}}^{6,k} = 32\ -16\ -14\cdot 4\ -14\cdot 18\ -1\cdot 18\cdot 33$$

$$Z_{\mathrm{DiFF}}^{7,k} = -32\ -14\cdot 5\ -18\cdot 13\cdot 0\cdot 15\cdot 15$$

$$Z_{\mathrm{DiFF}}^{8,k} = -27\ -8\cdot 3\ -13\cdot 5\cdot 2\cdot 2\cdot 3$$

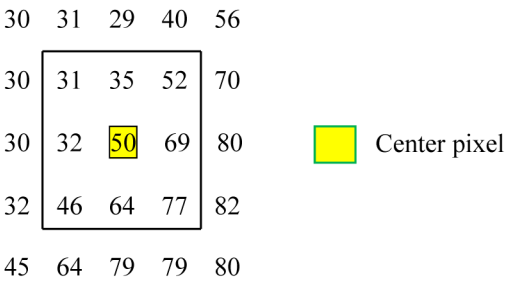

Fig. 21 *Sample input*

$$Z_{\mathrm{DiFF}}^{9,k} = -19\ -15\cdot 2\ -18\cdot 19\ -4\cdot 14\cdot 27$$

The three positive histogram peak values (a, b and c) are given below:

$$a = 3$$
$$b = 16$$
$$c = 19$$

The three negative histogram peak values ($-a$, $-b$ and $-c$) are given below:

$$-a = -4$$
$$-b = -14$$
$$-c = -17$$

The Difference Category (DC_K) is computed using (25) and the values are given below:

$$\mathrm{DC}_K = -0\cdot 1\cdot 4\cdot 0\cdot 6\cdot 2\cdot 5\cdot 7$$