IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 27, 2019, accepted April 14, 2019, date of current version May 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911954

Correlated Microstructure Descriptor for Image Retrieval

HUSSAIN DAWOOD[1], MONAGI H. ALKINANI[2], AHMAD RAZA[3], HASSAN DAWOOD[3], RUBAB MEHBOOB[3], AND SIDRA SHABBIR[3]

[1]Department of Computer and Network Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah 21589, Saudi Arabia
[2]Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah 21589, Saudi Arabia
[3]Department of Software Engineering, University of Engineering and Technology, Taxila 47050, Pakistan

Corresponding author: Hussain Dawood (hdaoud@uj.edu.sa)

This work was funded by the Deanship of Scientific Research (DSR), University of Jeddah, Jeddah, under grant No. (UJ-38-18-DR). The authors, therefore, acknowledge with thanks DSR technical and financial support.

ABSTRACT In this paper, a novel feature descriptor named ''correlated microstructure descriptor (CMSD)'' is proposed for image retrieval. CMSD represents high level semantics by identifying microstructures via establishing correlations between texture orientation, color, and intensity features. The proposed CMSD follows a two-staged approach for feature extraction and their integration. Color information is extracted by exclusively quantizing each component of HSV color space. Richer edge orientation information is extracted by using the multi-directional Sobel operator. Local contrast information is obtained by quantizing V component of HSV. Correlated microstructures are then identified by correlations established on the basis of proximity and continuation relations. The identified microstructures are then mapped to color, texture orientation, and intensity features exclusively to obtain micro-color, micro-orientation, and micro-intensity maps, respectively. Proposed three concatenated micro-maps are represented through a 2-D histogram, incorporate information regarding semantic and spatial contents of an image. The experiments are performed on standard datasets, i.e., Corel 1k, Corel 5k, and Corel 10k. The results evaluation depicts the outstanding performance of proposed CMSD as compared to state-of-the-art methods including TCM, GLCM, MTH, MSD, MTSD, and SED.

INDEX TERMS Content based image retrieval, correlated low-level visual features, correlated primary visual features, micro-maps, microstructures.

I. INTRODUCTION

Progressive development in multimedia tools and applications has promoted ease of accessibility to wider range of images containing diverse content. The rapid digital growth intends to reduce semantic gap between human and machine for retrieval of relevant images through effective retrieval systems. This prevalent variation of digital media has enormously expanded multimedia databases in different fields [1]. Currently image retrieval systems are used in various fields such as Education [2], medical diagnostics [3], online biometric systems [4], social networks [1], search engine [5] and many more. Content based image retrieval systems (CBIR) are actively playing role in improving the performance of retrieval systems. However, these systems are primarily based on visual features and cannot reach human visual perception level. Therefore, high level semantic information is required for achieving better perception.

Existing image retrieval techniques such as text-based (TBIR) methods for image retrieval usually use keywords annotated on images to search a particular image in a database. Such methods are quite popular in various applications such as Google, Bing, MSN etc. However, manual annotations increase time complexity. TBIR ignores the perceptual significant primary visual features i.e. color, shape and texture etc. In order to reduce the semantic gap, ontology-based image retrieval approaches, map vocabularies to define the object of interest [6]–[8]. The domain independent visual concept ontology is proposed to support automatic recognition [9], [72]. However, such approaches required detail annotations and are also time consuming. Sketch-based image retrieval proposed in Histogram of Line Relationship (HLR) [10], [11] bridge the appearance gap between the images and sketches. HLR describes the shape

The associate editor coordinating the review of this manuscript and approving it for publication was Antonio J. Plaza.

of object by selecting respective tokens therefore, minimizes the impact of noisy edges. The impreciseness of annotations in text-based, ontology-based approaches and perceptive subjectivity of sketch-based approaches may cause unrecoverable mismatches in image retrieval [53]. Relevance Feedback (RF) introduced in [12], [14] has considerably improved the performance of CBIR by incorporating human perception. A hybrid approach for image retrieval proposed in [15] minimizes the semantic gap by using color and query alteration. User intentions exploited in Biased Discriminative Euclidean Embedding (BDEE) [16], Biased Discriminant Analysis (BDA) [17], Case Based-Long Term Learning (CB-LTL) [18] have increased the retrieval performance and users' satisfaction. However, underlying complex similarity functions in RF techniques required powerful hardware resources. Image retrieval technique in [19], [70] utilized the Bag of Visual Words (BoVW) by incorporating textural and color characteristics of an image for retrieving relevant images. The method in [20] used sparse representation of data with BoVW to reduce curse of dimensionality. BoVW-based methods perform better but, they are computationally expensive.

Generally, semantic contribution of CBIR methods is comparatively low than semantic based image retrieval (SBIR) methods. CBIR descriptors are subdivided into two types of descriptors i.e. local descriptors and global descriptors. Local descriptors extract primary visual information from local patches of image while, global descriptors consider whole image to extract meaningful information of an image. Local descriptors encode the local content of image for retrieval by using the set of detected regions of interest. Scale-Invariant Feature Transform (SIFT) is the most commonly used local descriptor because of its scale and rotation invariant characteristics [21]. High dimensions of SIFT are reduced by PCA SIFT [22]. Root SIFT [23] is also used in SIFT-based retrieval. Time complexity of Speeded-up Robust Features (SURF) is significantly reduced due to Hessian-Laplace detector and local histogram [24]. However, SURF is sensitive to image rotation. Histogram of Oriented Gradients (HOG) is invariant to illumination and shadowing effects because of locally normalized contrast [25]. Inspired by the recognition accuracy, and simplicity, several variants of Local Binary Patters (LBP) have been proposed [26], [29]. These approaches are introduced for gray-scale (8-bit) images but natural images comprised of richer color content. Multi-channel decoded LBP provides joint information of each color channel for retrieval [30]. Local intensity differences are captured by combining both LBP and Local Neighborhood Difference Pattern (LNDP) for texture based image retrieval [31]. Local tri Directional Pattern (LDP) captures local intensity difference across local neighborhood along three dimensions [32]. Local Structure Descriptor (LSD) achieves low dimensionality and high indexing by combining color, shape and texture for image retrieval [33], [71]. Local Ternary Pattern (LTP) captures the spatial information whereas LBP captures the local textural information [34]. Edge orientation with underlying colors and spatial layout are effectively combined to represent image features by microstructures [35]. Local Texture-based Color Histogram (LTCH) effectively combines color and texture [36]. Color distribution is defined by microstructure image and Local Trinary Pattern (LTP) extract the textural features. Probabilistic latent semantic analysis (PLSA) and microstructures efficiently retrieve multimodal images [37]. However, In [35], [37] global properties of image are not considered and relation between spatial layout of different objects in an image is neglected.

Global descriptors consider low level visual information such as texture, color, shape and spatial information to represent an image. Color features are widely used in CBIR and considered as more efficient features. In MPEG7 [38], various types of color descriptors incorporating color layout, dominant color, color structure and scalable color descriptor are proposed. Color Histogram (CH) being effective and least computationally extensive is the commonly used to extract color information but, it lacks spatial information [39]. Image retrieval by using the Spatiogram of Colors quantized by Gaussian Mixture Models (SoC-GMM) [40] incorporate spatial information to the basic CH. In SoC-GMM, spatiograms are constructed by GMM, which is in turn used as a similarity function. Color Scale-invariant Feature Transform (Color-SIFT) by using weighted codebook distribution is proposed in Edge Orientation Difference Histogram (EODH) [41], [42]. The method in [43] extracted image features in YCbCr color space by using canny edge detector and discrete wavelet transform for effective image retrieval. However, color-based descriptors being invariant to scale and image rotation represent the high-level semantics of image. Texture being a global feature represents internal spatial information of an image and is described in terms of coarseness, homogeneity and roughness. Many descriptors such as Gabor transform [44], [45], Grey Level Co-occurrence Matrix (GLCM) [46], Markov Random Field (MRF) [47] and Edge Histogram Descriptor (EHD) [48] are proposed to extract textural characteristics of an image. Edge Histogram Descriptor (EDH) proposed in [43], [49] constructs the histogram of color and texture based features. Image retrieval techniques [50] quantizes H component of HSV color space and gradient into certain number of bins. The texture component considers a window of 3x3 to compute the co-occurrence relationship between the pixels symmetric along diagonals. The color and textural features are effectively represented by Color Information Feature (CIF) and LBP-based feature [51]. The input image is quantized into certain number of colors in HSV color space. Texture features are extracted by using Rotated Local Binary Patterns (RLBP). Both color and texture features fused together to represent a single feature set [52]. The proposed approach in [36] describes various shape-based descriptors in CBIR domain like invariant moments, edge curvature and arc length, polygonal approximation, Fourier transform coefficient, curvature scale space and Edge Orientation Auto Correlograms (EoAC).

However, shape-based descriptors usually carry sematic information around the boundaries or specific area of interest. Furthermore, they are sensitive to image translation, rotation and scaling. There are many descriptors proposed for CBIR [53] in order to incorporate the advantages of all the low level visual features. Color, edge orientation and color difference in L∗a∗b∗ space are integrated together in Color Difference Histogram (CDH) [54]. Local Gray Gabor Pattern (LGGP) proposed in [55] combines LBP and Gabor to integrate color, shape and texture together. Both LBP and wavelet based features are considered in [56]. However, descriptors based on the combination of low-level visual features incorporate high level semantic concepts.

Low level visual features such as color, texture orientation and contrast effectively represent the characteristics of an image in CBIR domain. Single low-level visual features i.e. either color, texture orientation or contrast is not sufficient to represent high level semantics due to highly complex visual content. Existing approaches such as CH considers only color features [39]. GLCM [46], Gabor features [44], [45], MRF [47] represent spatial information of an image partially by extracting textural features only. MTH [57], RLBP [52], CIF and LBP [51], CDH [54], EDH [48], MSD [35] represent high level semantics by correlating color and edge orientations and ignore the intensity information of an image. Based on limitations of existing classic color, color and texture-based systems; there is a need of new highly correlated primary visual feature descriptor to incorporate high level semantic concepts in CBIR domain.

In this paper, a novel framework based on primary visual features named Correlated Microstructure Descriptor (CMSD) is proposed. CMSD represents high level semantics by identifying microstructure images by establishing correlations between color, texture orientation and intensity features. CMSD is a two-staged approach. In the first stage low level visual features such as color, texture orientation and intensity information are extracted from an input image and then quantized. In the second stage, microstructure images consisting of perceptual uniform regions are identified by establishing correlations between color, orientation and intensity, confirming to proximity and continuity relations. Three micro-maps i.e. micro-color, micro-orientation and micro-intensity maps are then obtained by correlating each microstructure images with quantized color, orientation and intensity features respectively. Final feature vector represent significant semantic characteristics of an image is obtained by concatenating the resultant three micro-maps. It represents only significant points lying within the perceptual uniform region and ignores the irrelevant details.

Novelty of proposed descriptor as compared to MSD [35] is described as follows:

1. Proposed method incorporates intensity information instead of using only color and orientation information however, in MSD only color and edge information is used.
2. Microstructures or regular patterns are extracted based on the correlation between two visual primary features instead of using single primary visual features as used in MSD.
3. In proposed method three separate color, edge and intensity micro maps are obtained instead of only micro color map used in MSD.
4. The proposed feature represents high level semantics by identifying microstructure image by establishing correlations between texture orientation, color and intensity features.
5. ARP and ARR of proposed descriptor outperform the MSD method on all corel1k, corel5k and corel10k datasets.

The organization for the rest of article is as follows. Section 2 explains the proposed methodology of CMSD descriptor. Results evaluation and discussion is elaborated in section 3. Section 4 concludes the proposed work.

II. CORRELATED MICROSTRUCTURE DESCRIPTOR (CMSD)

In this paper, a novel feature descriptor named as "Correlated Microstructure Descriptor (CMSD)" is proposed for image retrieval. The proposed model followed a two-staged approach for extracting features and their integration. In first stage, three primary visual features such as color, edge orientation and contrast are extracted from the input image. Then, all extracted features are quantized into some level to reduce the time and space complexity of proposed descriptor. The input image is exclusively quantized into 72 colors to extract color features in HSV color space. Multi-directional edge orientation features are obtained by quantizing the edges obtained by 4-dimensional Sobel operator into 6 levels. Local contrast information is obtained by quantizing V component of HSV into 10 levels. In the second phase, micro-color, micro-orientation and micro-intensity maps are obtained by correlating other two quantized maps, simultaneously. Correlations are established on the basis of proximity and continuity relations. Correlation is defined as the dependency between two variables. However, we can say that if there is any type of relationship exist between two variables then these two variables are correlated. For example, in order to extract micro-color map, quantized intensity and orientation map will be correlated. The correlation is measured in such a way that same regular patterns should exist in both intensity and orientation map, concurrently. So, Intensity and orientation maps are correlated. Micro-color map represents the color patterns having same underlying intensity and orientation information. Three micro-maps are then concatenated to populate a 2-D histogram. Fig 1 illustrates the proposed model for construction of micro-color, micro-orientation and micro-intensity maps.

A. COLOR QUANTIZATION IN HSV COLOR SPACE

Color features being wavelength-dependent perception is considered as a most significant visual indication for image

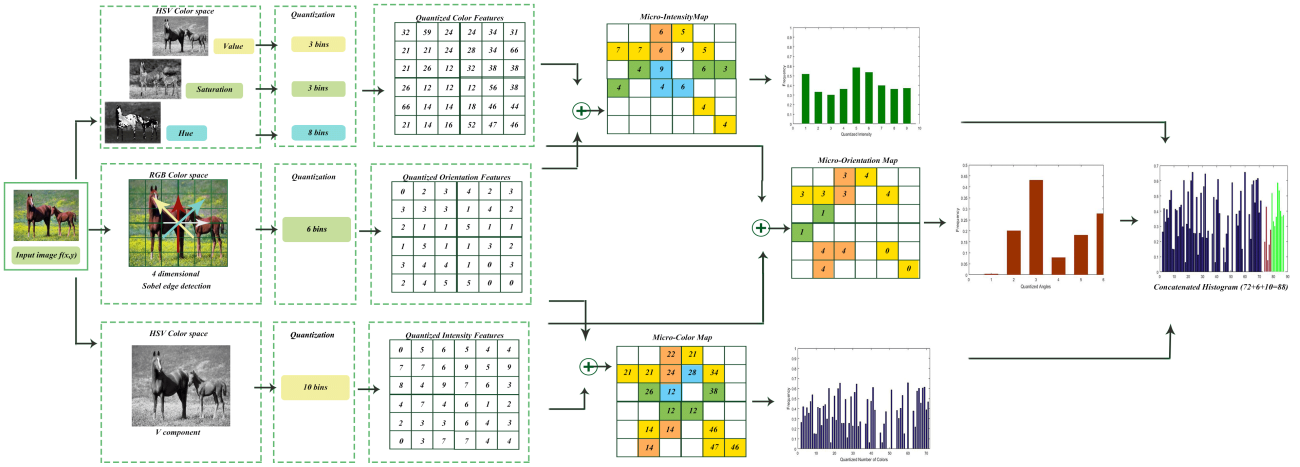

FIGURE 1. **Framework of Proposed Co-related Microstructure Descriptor (CMSD).**

retrieval and object recognition tasks [19]. However, people may have subjective nature regarding specificity of colors in different contexts. In real world objects, the appearance of color can vary by textures, lighting and viewing conditions [58]. Color channels i.e. Red (R), Green (G) and Blue (B) in RGB color space are extremely correlated and are perceptually non-uniform. Moreover, RGB is neither related to human interpretations of color, nor intrinsically related to natural color properties [59]. Therefore, an input image $f(x, y)$ is transformed into HSV color space in proposed CMSD because of its uniformity towards human color perception. Furthermore, HSV color space is invariant to illumination changes.

HSV color space consists of 3 components: Hue (h), Saturation (s) and Value (v) in cylindrical coordinate system. Hue (h) component characterizes the color type with values ranging between [0°, 360°] having red at 0°, green at 120° and blue at 240°. Saturation (s) component determines the purity of particular color with white color having values between 0 and 1. Values (v) component represents the brightness or illumination having values between 0 and 1. In order to balance the computational complexity and the perceptual uniformity, the input image $f(x, y)$ is quantized into 72 colors. According to the quantization scheme followed in CMSD as represented in (1), (2) and (3), h is quantized into B_h bins, s into B_s bins and v into B_v bins. The values of B_h, B_s and B_v are taken as 8, 3 and 3 respectively, having the total of $8*3*3 = 72$ quantization levels respectively.

$$Q_h = \{h \times (B_h/max_h)\} \quad (1)$$

$$Q_s = \{s \times (B_s/max_s)\} \quad (2)$$

$$Q_v = \{v \times (B_v/max_v)\} \quad (3)$$

Final color map CM for input image,$f(x, y)$is given by (4).

$$\mathrm{CM} = Q_h \times (B_s \times B_v) + Q_S \times B_v + Q_v \quad (4)$$

Here B_h, B_s and B_v are quantization levels and Q_h, Q_s and Q_v are quantized values for Hue (h), Saturation (s) and Values (v) respectively. The quantized color map is represented as follows:

$$Q_{CM(i)} = \{(x, y)|(x, y) \in CM = i, 0 \leq i \leq N_c - 1\} \quad (5)$$

Here x and y are the spatial co-ordinates, N_c represent the number of quantized colors and i be the resultant value obtained in quantized color map, Q_{CM}. The selected value for $N_c = 72$

B. MULTI-DIMENSIONAL TEXTURE ORIENTATION DETECTION IN RGB COLOR SPACE (MD-TOD)

Edge information incorporates strong semantic information regarding object boundaries and texture structure. The edge maps contain implicit information regarding position of pixels, their strength, and orientation and scaling. On contrary, the edge orientation maps include explicit information regarding the geometry of shape, texture pattern and edges [60]. In RGB color space, an object in an image exhibits homogenous variations within the subject area and sharper variations across the boundary. The abrupt variations in pixel intensities in gray-scale image decrease the likelihood of detecting an obvious edge [61]. Generally, first and second order edge detection operators such as Laplacian of Gaussian (LoG), Prewitt, Sobel, Robert and Average work well with gray-scale images [62]. However, the natural images are a blend of regular and irregular textures. The following problems may occur if colored images are converted to gray-scale: 1) Much of the chromatic information will be lost. 2) Most of the obvious edges produced by spectral variations will be missed. In order to overcome the aforementioned problems, MD-TOD is proposed in RGB color space because it defines the refine boundary of objects by preserving the sharper variations in intensities along boundaries in an image. Inspired by the efficiency and invariance against noise, 4-dimensional

Sobel edge detector is used to extract edge information along 4 different directions such as 0°, 45°, 90° and 135°.

Diagonal edges denoted by d_{45° and d_{135° be the edge information extracted along 45° and 135° in RGB color space by using (10). Let $\widehat{d_{45^\circ}}(R_{45^\circ}, G_{45^\circ}, B_{45^\circ})$ and $\widehat{d_{135^\circ}}(R_{135^\circ}, G_{135^\circ}, B_{135^\circ})$ be the two unit vectors in Cartesian co-ordinate system. Let R_{45°, G_{45° and B_{45° be the edges extracted along 45° by using 45° Sobel operator whereas R_{135°, G_{135° and B_{135° be the edges extracted along 135° by using 135° Sobel operator. The orientation map along diagonals is obtained as follows:

$$cos\left(\widehat{d_{45^\circ}}, \widehat{d_{135^\circ}}\right) = \frac{\widehat{d_{45^\circ}} * \widehat{d_{135^\circ}}}{|\widehat{d_{45^\circ}}||\widehat{d_{135^\circ}}|} \tag{6}$$

$$\widehat{d_{45^\circ}} * \widehat{d_{135^\circ}} = R_{45^\circ}R_{135^\circ} + G_{45^\circ}G_{135^\circ} + B_{45^\circ}B_{135^\circ} \tag{7}$$

$$|d_{45^\circ}| = (R_{45^\circ}^2 + G_{45^\circ}^2 + B_{45^\circ}^2)^{\frac{1}{2}} \tag{8}$$

$$|d_{135^\circ}| = (R_{135^\circ}^2 + G_{135^\circ}^2 + B_{135^\circ}^2)^{\frac{1}{2}} \tag{9}$$

$$Orimap_{diag} = arccos^\circ(cos\left(\widehat{d_{45^\circ}}, \widehat{d_{135^\circ}}\right)) \tag{10}$$

Similarly, compact edge information from each color channel is extracted by using 0° and 90° sobel operator along horizontal and vertical directions, respectively. Let h and v be the edge information extracted along 0° and 90° in RGB color space by using (15). Let $\hat{h}(R_h, G_h, B_h)$ and $\hat{v}(R_v, G_v, B_v)$ be the two unit vectors in Cartesian co-ordinate system. Let R_h, G_h *and* B_h be the edges extracted along horizontal direction by Sobel operator whereas R_v, G_v *and* B_v be the edges extracted along vertical direction by using Sobel operator. The orientation map along horizontal and vertical direction is obtained as follows:

$$cos\left(\hat{h}, \hat{v}\right) = \frac{\hat{h} * \hat{v}}{|\hat{h}||\hat{v}|} \tag{11}$$

$$\hat{h} * \hat{v} = R_hR_v + G_hG_v + B_hB_v \tag{12}$$

$$|\hat{h}| = (R_h^2 + G_h^2 + B_h^2)^{\frac{1}{2}} \tag{13}$$

$$|\hat{v}| = (R_v^2 + G_v^2 + B_v^2)^{\frac{1}{2}} \tag{14}$$

$$Orimap_{hv} = arccos(cos\left(\hat{h}, \hat{v}\right)) \tag{15}$$

Finally, quantized orientation map incorporating edge information extracted along all 4 directions is represented as follows:

$$Q_{Orimap} = \frac{1}{2}(Orimap_{hv} + Orimap_{diag}) \times \frac{N_o}{180} \tag{16}$$

Here N_o be the quantization levels. The value of N_o is taken as 6.

$$Q_{Orimap(j)} = \{(x, y)|x, y \in Q_{Orimap} = j, \quad 0 \leq j \leq N_o - 1\} \tag{17}$$

Here x and y be the spatial co-ordinates of orientation map computed along horizontal, vertical and diagonal directions of input image. Let j be the resultant value obtained after quantization of Q_{Orimap} into 6 levels.

C. INTENSITY MAP EXTRACTION

The contrast or brightness information is characterized by intensity or variance [63]. Contrast information complements the performance of certain primary visual features and thereby leads to enhance retrieval accuracy. Intensity information in proposed CMSD is extracted by considering the Values (v) component of HSV color space. Since, v component separated from the color information indicates the luminance or intensity of a color. Moreover, the most prominent cues for eliminating illumination are extracted by computing relationships between pixels intensities. Intensity map for input image $f(x, y)$ is constructed while using v component of HSV color space as follows:

$$IM\ (x, y) = \sum_{x=1}^{M}\sum_{y=1}^{N} \max(R(x, y), \max(G(x, y), B(x, y))) \tag{18}$$

Here x and y be the spatial co-ordinates of $f(x, y)$. M and N be the dimensions of input image $f\ (x, y)$. Since, human eye neither distinguish the small variations in pixel intensities nor identify the small variations in brightness levels [73]. Therefore, in (19) we distill the informative clues from pixels' intensities in resultant intensity map while suppressing the least useful ones by quantizing into N_I intervals. The quantized intensity map for $f\ (x, y)$ are represented by (20).

$$Q_{IM}\ (x, y) = IM\ (x, y) \times N_I \tag{19}$$

$$Q_{IM}\ (x, y)_k = \{(x, y)|(x, y) \in Q_{IM}\ (x, y) = k, \quad 0 \leq k \leq N_I - 1\} \tag{20}$$

Let k be the resultant value obtained in quantized intensity map, Q_{IM}. The selected value for N_I *is* 10.

D. CORRELATED MICROSTRUCTURE IDENTIFICATION

Since human visual system is sensitive to certain low-level visual features such as color, orientation and intensity. Orientation features are concerned with the definite boundary of object within an image. However, natural images exhibit both regular and irregular patters of objects, contributing to diverse spatial distribution of various contents, leading to resultant microstructures. However, orientation and color features alone are not sufficient to extract the entire semantic information and spatial layout of objects in an image. Therefore, three low level features i.e. color, orientation and intensity are correlated to identify microstructures. In this paper, micro-color, micro-orientation and micro-intensity maps are defined on the basis of underlying colors, texture orientations and contrast respectively. The underlying colors are obtained by correlating a group of pixels on the basis of proximity relation established due to frequently occurring similar patterns i.e. perceptual uniform regions in both orientation and intensity maps. Similarly, Micro-orientation map is defined by correlating orientation features and a microstructure image which is obtained by correlating both color and intensity

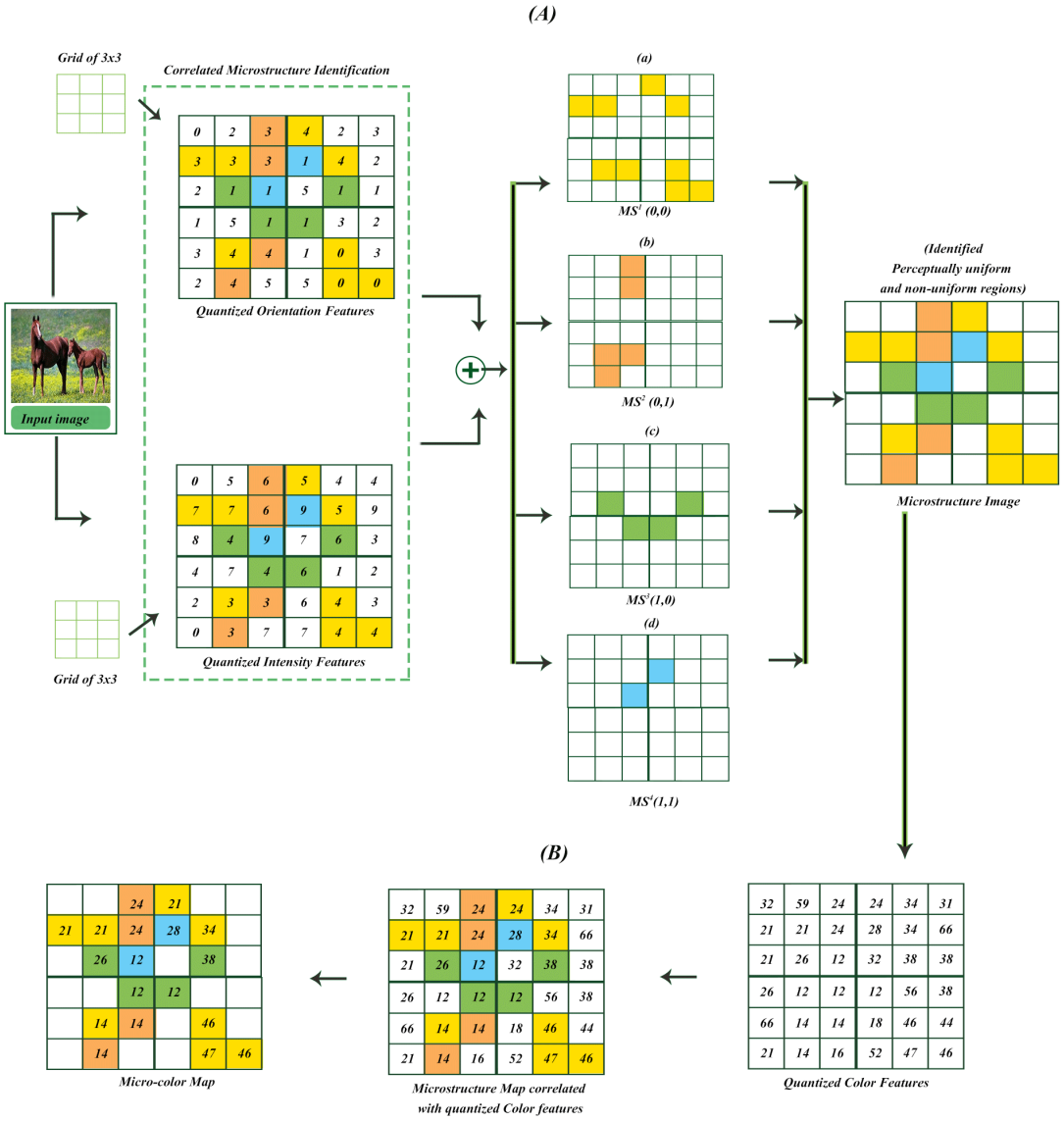

FIGURE 2. **(A) Construction of micro-structure by correlating perceptually uniform regions of orientation and intensity (B): Extraction of micro-color Map.**

features. Likewise, micro-intensity maps are obtained by correlating intensity feature and a microstructure image which is obtained by correlating both orientation and color features. Moreover, microstructures intuitively incorporate color, texture and contrast-based features, therefore, represent better feature set for image retrieval.

In order to construct the micro-color map, quantized orientation and intensity maps are considered. Therefore, the permissible range of values for Q_{Orimap} and Q_{IM} is 0 to 5 and 0 to 9, respectively. The entire Q_{Orimap} is partitioned into small grids of 3 × 3. For convenience, the step length of 3 pixels is taken along horizontal and vertical direction for each grid. Let $\hat{G}_i = \{G_{ij}, G_{i+1}, G_{i+2} \ldots, G_{pq}\}$ be the grids in which the Q_{Orimap} is partitioned where i = {1, 2, 3,. ... P} and p be the total number of grids. Similarly, $\hat{G}_j = \{G_{jj}, G_{j+1}G_{j+2} \ldots, G_{qq}\}$ be the grids in which the Q_{IM} is partitioned where j = {1, 2, 3..., q} and q be the total number of grids. In order to detect uniform perceptive regions, proximity relations are established between central pixel and the neighborhood pixels on the basis of similarity within the grid moved on both Q_{Orimap} and Q_{IM}, Simultaneously. Let G_i and G_j be the grids of step length 3 are moved on both Q_{Orimap} and Q_{IM} respectively. Let CP_O be the central pixel and k_{ii} be the neighborhood pixels of G_i in Q_{Orimap} and CP_I be the central pixel and k_{jj} be the neighborhood pixels of G_j in Q_{IM} where ii = {1,2,3,...,8} and jj = {1,2,3,..,8}. If any pixel from k_{ii} has the same value as CP_O and k_{jj} has the same value as CP_I and ii = jj while moving the grid of size 3 × 3 on both Q_{Orimap} and Q_{IM} simultaneously, then the color pattern at that particular location is retained in color map as shown in Fig 2(A). The retained region is known as perceptually uniform region

and that grid is termed as fundamental microstructure block. On contrary, if none of the values from neighborhood pixels k_{ii} and k_{jj} has same value as of central pixel CP_O and CP_I respectively, then all the values in that grid are discarded indicating perceptually non-uniform region and declaring the grid as non-fundamental microstructure block.

Different micro-structures obtained at 4 different locations are integrated to construct a single micro-structure image by correlating frequently occurring perceptually uniform regions in both Q_{Orimap} and Q_{IM}. These four fundamental locations are (0,0), (0,1), (1,0) and (1,1). Five-staged strategy is followed to construct a single microstructure in order to obtain final micro-color map.

1. At the location (0,0), a grid of size 3×3 with the step length of 3 pixels is moved on both Q_{Orimap} and Q_{IM} along left to right and top to bottom. The microstructure obtained at location (0,0) is labeled as $MS^1(x, y)$ by correlating perceptually uniform regions in both Q_{Orimap} and Q_{IM} where $0 \leq x \leq m-1, 0 \leq y \leq n-1$ and m and n be the dimensions of Q_{Orimap} as shown in Fig 2. A(a).
2. At the location (0,1), a grid of size 3×3 with the step length of 3 pixels is moved on both Q_{Orimap} and Q_{IM} along left to right and top to bottom. The microstructure obtained at location (0,1) is labeled as $MS^2(x, y)$ by correlating perceptually uniform regions in both Q_{Orimap} and Q_{IM} where $0 \leq x \leq m-1, 1 \leq y \leq n-1$ and m and n be the dimensions of Q_{Orimap} as shown in Fig 2. A(b).
3. At the location (1,0), a grid of size 3×3 with the step length of 3 pixels is moved on both Q_{Orimap} and Q_{IM} along left to right and top to bottom. The microstructure obtained at location (1,0) is labeled as $MS^3(x, y)$ by correlating perceptually uniform regions in both Q_{Orimap} and Q_{IM} where $1 \leq x \leq m-1, 0 \leq y \leq n-1$ and m and n be the dimensions of Q_{Orimap} as shown in Fig 2. A(c).
4. At the location (1,1), a grid of size 3×3 with the step length of 3 pixels is moved on both Q_{Orimap} and Q_{IM} along left to right and top to bottom. The microstructure obtained at location (1,1) is labeled as $MS^4(x, y)$ by correlating perceptually uniform regions in both Q_{Orimap} and Q_{IM} where $1 \leq x \leq m-1, 1 \leq y \leq n-1$ and m and n be the dimensions of Q_{Orimap} as shown in Fig 2A(d).
5. Finally, the consolidated micro-structure image comprising of fundamental perceptual uniform regions indicating the continuation relations is obtained as follows:

$$MS = \max(MS^1(x, y), MS^2(x, y), \times MS^3(x, y), MS^4(x, y)) \quad (21)$$

E. FEATURE REPRESENTATION

The extraction of significant yet discriminative features is the main problem in image retrieval tasks. However, to integrate visual features in identified microstructures to explore high level semantics and spatial layout of contents of image are a challenging issue. After the construction of microstructure image, next step is how to integrate the obtained microstructures with quantized visual features i.e. color, texture orientation and intensity. Based on the identified microstructures and their correlation together with quantized color features, texture orientation and intensity features, we describe CMSD as follows:

1) MICRO-COLOR MAP FEATURES

The values of microstructure image $f'(x, y)$ as shown in Fig 2A for an input image $f(x, y)$ are represented by $f'(x, y) = \alpha$ where $\alpha \in \{0, 1, 2, ...c-1\}$ and c represents the dimensions of micro-color map obtained by correlating both orientation and intensity maps. Features of microstructure image by moving grid of 3×3 are represented as follows:

$$H(\alpha) = \begin{cases} \dfrac{N_1\{f(p\alpha_0) = \alpha_0, \Delta f(p\alpha_i) = \alpha_i \, |p\alpha_i - p\alpha_0| = 1\}}{8\overline{N_1}\{f(p\alpha_0) = \alpha_0\}} \\ where \ \alpha_0 = a_i, \quad i = \{1, 2, 3, ..., 8\} \end{cases} \quad (22)$$

Let $p\alpha_0$ be the central pixel at the position (x_0, y_0) and its value be $f(p\alpha_0) = \alpha_0$. Let 8 neighbors of $p\alpha_0$ be denoted by $p\alpha_i$ at the position(x_i, y_i). Let N_1 denotes the frequency of co-occurring values of both $\alpha_0 and \alpha_i$ and $\overline{N_1}$ be the frequency of α_0. The micro-color map is obtained by correlating microstructure image $f'(x, y)$ and Q_{CM}. Final micro-color map contains only those pixel values from Q_{CM} which lies in perceptual uniform region present in $f'(x, y)$ as shown in Fig 2B. The dimensions of micro color-map features are 72.

2) MICRO-ORIENTATION MAP FEATURES

The values of microstructure image $f''(x, y)$ for an input image $f(x, y)$ are represented by $f''(x, y) = \beta$ where $\beta \in \{0, 1, 2, ...c-1\}$ and c represents the dimensions of micro-orientation map is obtained by correlating color and intensity maps. Features of microstructure image by moving grid of 3×3 on Q_{CM} and Q_{IM} are represented as follows:

$$H(\beta) = \begin{cases} \dfrac{N_2\{f(p\beta_0) = \beta_0, \Delta f(p\beta_j) = \beta_j \, |p\beta_j - p\beta_0| = 1\}}{8\overline{N_2}\{f(p\beta_0) = \beta_0\}} \\ where \ \beta_0 = \beta_j, \quad j = \{1, 2, 3, ..., 8\} \end{cases} \quad (23)$$

Let $p\beta_0$ be the central pixel at the position (x_0, y_0) and its value be $f(p\beta_0) = \beta_0$. Let 8 neighbors of $p\beta_0$ be denoted by $p\beta_j$ at the position (x_j, y_j). Let N_2 denotes the frequency of co-occurring values of both $\beta_0 and \beta_j$ and $\overline{N_2}$ be the frequency of β_0. The micro-orientation map is obtained by correlating microstructure image $f'(x, y)$ and Q_{Orimap}. Final micro-orientation map contains only those pixel values from Q_{Orimap} which lies in perceptual uniform region present in $f''(x, y)$. The dimensions of micro-orientation map features are 6.

3) MICRO-INTENSITY MAP

The values of micro-intensity map for an input image $f(x, y)$ are represented by $f''(x, y) = \gamma$ where

$\gamma \in \{0, 1, 2, ...c-1\}$ and c represents the dimensions of micro-intensity map obtained by correlating color and orientation maps. Features of microstructure image by moving grid of 3×3 on both Q_{Orimap} and Q_{CM} are represented as follows:

$$H(\gamma) = \begin{cases} \dfrac{N_3\{f\ (p\gamma_0) = \gamma_0,\ \Delta f\ (p\gamma_k) = \gamma_k\ |p\gamma_k - p\gamma_0| = 1\}}{8\overline{N_3}\{f\ (p\gamma_0) = \gamma_0\}} \\ where\ \gamma_0 = \gamma_k, \quad k = \{1, 2, 3, ..., 8\} \end{cases} \tag{24}$$

Let $p\gamma_0$ be the central pixel at the position (x_0, y_0) and its value be $f\ (p\gamma_0) = \gamma_0$. Let 8 neighbors of $p\gamma_0$ be denoted by $p\gamma_j$ at the position (x_k, y_k). Let N_3 denotes the frequency of co-occurring values of both $\gamma_0 and \gamma_k$ and $\overline{N_3}$ be frequency of γ_0. The micro-orientation map is obtained by correlating microstructure image $f'''(x, y)$ and Q_{IM}. Final micro-intensity map contains only those pixel values from Q_{IM} which lies in perceptual uniform region present in $f'''(x, y)$. The dimensions of micro-intensity map features are 10.

The final feature set associated with an input image is obtained by concatenating all the three micro-maps i.e. micro-color, micro-orientation and micro-intensity maps. The concatenated micro maps shown in Fig 1 is a part of the quantized color, orientation and intensity features. It represents only a few significant points belonging to perceptual uniform region and ignores the irrelevant details.

The CMSD can describe the different combination of color, edge orientation, intensity information and spatial distribution of the obtained microstructures by considering the dimensionality constraints.

III. RESULTS AND DISCUSSION

The preceding section illustrates the characteristics of datasets used for evaluation of proposed CMSD, evaluation matrices used for performance evaluation. It also describes the distance metrics used for measuring similarity and comparative analysis of different matrices on proposed CMSD. This section also analyses the effect of parameter variation on proposed CMSD. Performance comparison of proposed CMSD with several state-of-the-art methods such as TCM [64], MTH [57], MSD [35], SED [65], EoAC [66], EHD [48], color moment [67], CAC [39], CDH [54], MCMCM [46], SoC-GMM [40], Gabor [44], Wavelet transform [45], LBP+Wavelet [56], STH [68], MS-LSP [1] is also presented on each dataset.

A. DATASETS

CBIR systems are mostly evaluated on commonly used datasets such as Brodatz and OUTex texture datasets and Corel image datasets. Among these datasets Corel is one of the most widely used dataset for performance evaluation of image retrieval systems. There is a wide variety of images containing diverse content in Corel dataset. The performance evaluation of the proposed system is done on Corel 1k, Corel 5k and 10K datasets [69]. Corel 1k dataset contains 10 categories of semantically organized natural scenery images including bus, horse, elephant, food, African, mountain, beach, flower and dinosaur. All images in Corel 1k have same dimension of 384×256 or 256×384. Since there are 10 categories each containing 100 images therefore the total number of images in Corel 1k is 1000. Corel 5k has 50 different categories, containing a total of 5000 images of objects with varied content such as wave, pills, tree, food texture, bark wood texture, spices, stained glass and tile texture. Similarly, Corel 10k is a superset dataset containing 100 different object categories including all categories of objects of Corel 5k too. Corel 5k is basically a subset of Corel 10k both containing 100 different images in each category with dimensions 192×128.

B. EVALUATION METRICS

The performance evaluation of CBIR systems is generally measured using precision and Recall. The proposed systems employ both evaluation metrics. Precision of a query image P_{QI} can be measured using the following equation.

$$P_{QI} = \frac{I_R}{I_N} \tag{25}$$

where I_R is the representation of relevant images retrieved in response to query image from Database containing a total of I_N images. Similarly recall of a query image R_{QI} can be measured using equation:

$$R_{QI} = \frac{I_R}{I_c} \tag{26}$$

Here I_R is the representation of relevant images retrieved in response to query image from Database. Whereas I_c is the total number of images in each object category and is kept same for the same dataset and varies for different datasets.

Average Retrieval Precision (ARP) and Average Retrieval Recall (ARR) are computed by (28) and (30):

$$RP_i = \frac{1}{N_c} \sum_{i=1}^{N_c} P_{QI} \tag{27}$$

$$ARP = \frac{1}{T_c} \sum_{j=1}^{t_c} RP_j \tag{28}$$

$$RR_i = \frac{1}{N_c} \sum_{i=1}^{N_c} R_{QI} \tag{29}$$

$$ARR = \frac{1}{T_c} \sum_{j=1}^{t_c} RR_j \tag{30}$$

Here N_c stands for the total number of images in each category. The value of N_c is taken as 100 for each category of Corel 1k, 5k, and 10k. The retrieval precision and recall for each category (i) of dataset are given by RP_i and RR_i respectively. The total number of categories of each dataset is given by t_c. The value of t_c is taken as 10, 50 and 100 for Corel 1k, 5k and 10k respectively.

TABLE 1. ARR and ARP of proposed method by using different combinations of color and orientation quantization levels while intensity level is fixed to 10 on corel5k dataset in hsv color space.

Color Quantization Level	Texture orientation quantization level											
	Precision (%)						Recall (%)					
	6	12	18	24	30	36	6	12	18	24	30	36
192	62.52	62.73	62.72	62.49	62.65	62.31	7.50	7.52	7.52	7.49	7.51	7.47
128	63.35	63.34	63.28	63.23	63.29	63.03	7.60	7.60	7.59	7.58	7.59	7.56
108	61.76	62.25	62.35	62.40	62.34	62.28	7.41	7.47	7.48	7.48	7.48	7.47
72	63.14	63.36	63.45	63.31	63.62	63.31	7.57	7.60	7.61	7.59	7.63	7.59

TABLE 2. ARR and ARP of proposed method by using different combinations of color and orientation quantization levels while intensity level is fixed to 15 on corel5k dataset in hsv color space.

Color Quantization Level	Texture orientation quantization level											
	Precision (%)						Recall (%)					
	6	12	18	24	30	36	6	12	18	24	30	36
192	63.05	63.15	63.45	63.30	63.30	62.84	7.56	7.57	7.61	7.59	7.59	7.54
128	63.85	63.81	64.08	63.72	63.95	63.60	7.66	7.65	7.69	7.64	7.67	7.63
108	62.75	63.00	63.15	62.95	63.04	62.68	7.53	7.56	7.57	7.55	7.56	7.52
72	63.68	63.77	63.76	63.64	63.78	63.36	7.64	7.65	7.65	7.63	7.65	7.60

TABLE 3. ARR and ARP of proposed method by using different combinations of color and orientation quantization levels while intensity level is fixed to 20 on corel5k dataset in hsv color space.

Color Quantization Level	Texture orientation quantization level											
	Precision (%)						Recall (%)					
	6	12	18	24	30	36	6	12	18	24	30	36
192	62.89	63.14	63.26	63.10	63.08	62.81	7.54	7.57	7.59	7.57	7.57	7.53
128	63.69	64.10	64.16	64.15	64.16	63.84	7.64	7.69	7.70	7.69	7.70	7.66
108	63.04	63.09	63.45	63.26	63.26	63.00	7.56	7.57	7.61	7.59	7.59	7.56
72	63.96	63.95	63.95	63.73	63.73	63.74	7.67	7.67	7.67	7.64	7.64	7.64

C. DISTANCE METRIC

The performance of CBIR systems is mainly dependent on multiple factors such as the number of features, feature selection and their representation etc. Distance metric is also one of the significant factors, which greatly affect the performance of CBIR systems. In this regard, the choice of an appropriate distance metric is of high importance. In CMSD, Manhattan distance (L1) is used to measure distance between the query image and those which are stored in the database. Let q_i be the features extracted against a query image and s_i be the stored set of features where q, s = [1, 2, 3 . . .N_f]. L1 distance between q_i and s_i is calculated as follows:

$$D(s,q) = \sum_{i=1}^{N_f} |s_i - q_i| \tag{31}$$

Here N_f is the dimension of the feature vector obtained by proposed CMSD. Here, the value of N_f is $72 + 6 + 10 = 88$.

TABLE 4. ARR and ARP of proposed method by using different combinations of color and orientation quantization levels while intensity level is fixed to 10 on corel5k dataset in rgb color space.

Color Quantization Level	Texture orientation quantization level											
	Precision (%)						Recall (%)					
	6	12	18	24	30	36	6	12	18	24	30	36
128	58.82	58.75	59.08	58.94	59.21	58.98	7.05	7.05	7.09	7.07	7.10	7.07
64	56.55	56.59	56.77	56.60	56.89	56.73	6.78	6.79	6.81	6.79	6.82	6.80
32	51.43	51.79	52.41	52.05	52.14	51.96	6.17	6.21	6.29	6.24	6.25	6.23
16	43.77	45.08	45.54	45.68	45.91	46.10	5.25	5.41	5.46	5.48	5.51	5.53

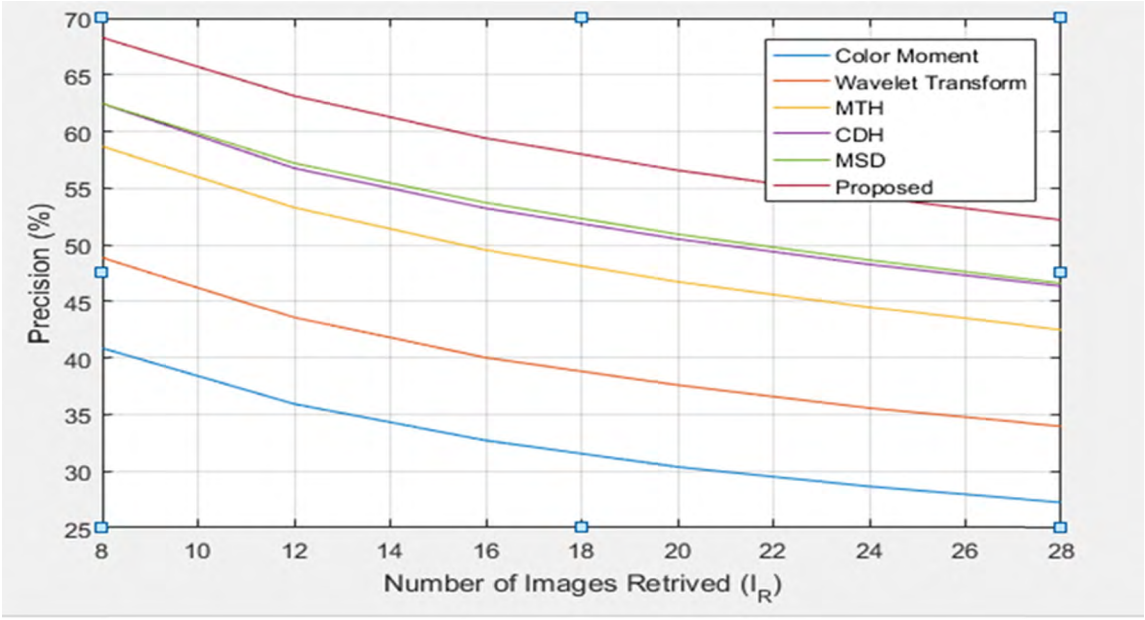

FIGURE 3. Average Retrieval precision of proposed CMSD is compared to other state-of-the-art methods on Corel 1k by varying number of retrieved images.

TABLE 5. ARR and ARP of proposed method at different similarity measures on corel5k, corel10k and corel1k.

Dataset	Performance	Distance or Similarity metrics					
		Extended-Canberra	Square-Chord	Weighted L1	Euclidean	Histogram Intersection	Manhattan
Corel-1k	Precision (%)	67.37	70.54	77.75	75.29	43.12	78.54
	Recall (%)	8.08	8.46	9.33	9.03	5.17	9.42
Corel-5k	Precision (%)	12.17	56.40	62.67	60.00	37.25	63.14
	Recall (%)	1.46	6.76	7.52	7.20	4.47	7.57
Corel-10k	Precision (%)	10.71	43.42	50.11	47.16	25.65	50.25
	Recall (%)	1.28	5.21	6.01	5.66	3.07	6.03

D. PERFORMANCE EVALUATION OF PROPOSED CMSD BY PARAMETER VARIATION

It is well known that different low-level visual features make different contributions to image retrieval. In proposed CMSD, certain low-level visual features such as color, edge orientation and intensity information are extracted from the input image. In order to evaluate the effect of integrating different low-level visual features, different quantization

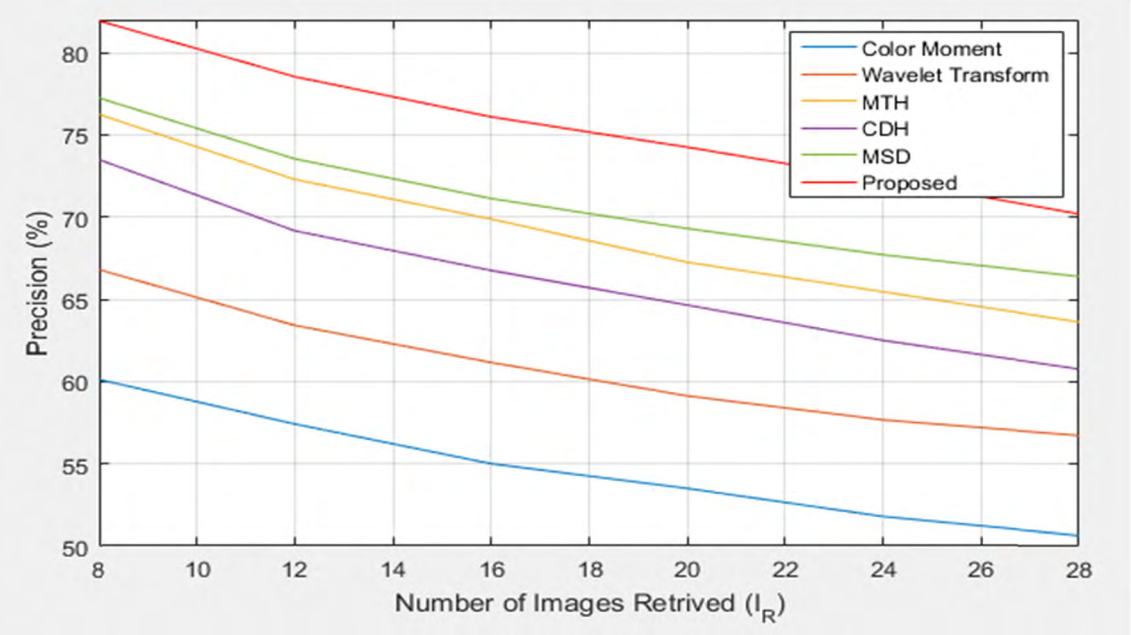

FIGURE 4. Average Retrieval precision of proposed CMSD is compared to other state-of-the-art methods on Corel 5k by varying number of retrieved images.

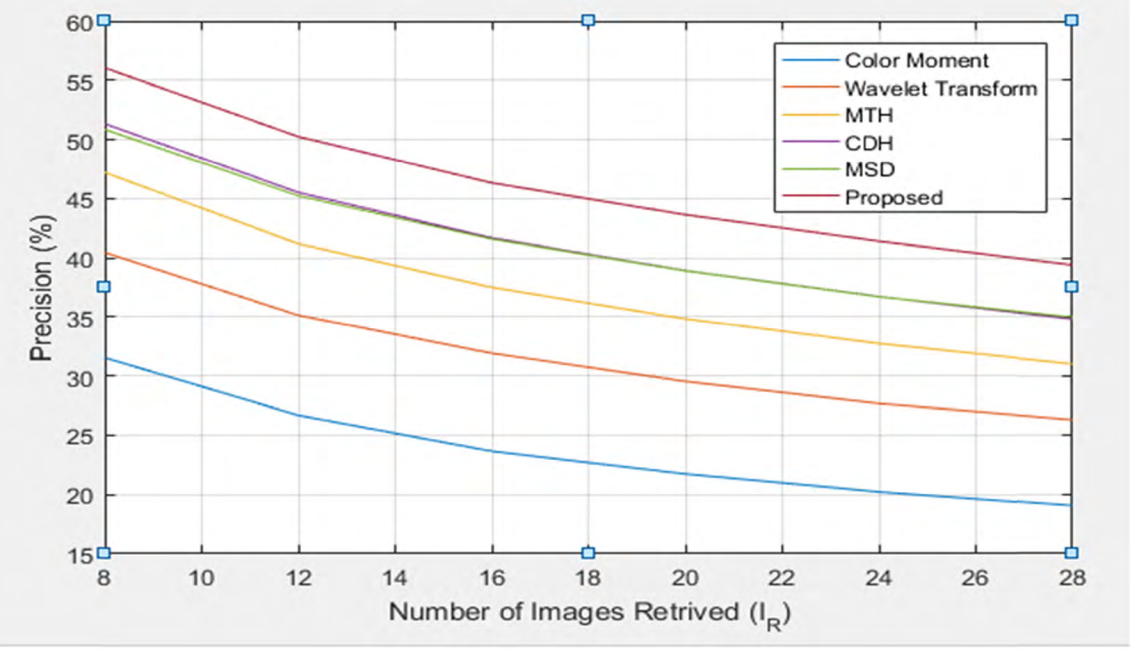

FIGURE 5. Average Retrieval precision of proposed CMSD is compared to other state-of-the-art methods on Corel 10k by varying number of retrieved images.

levels for color, orientation and intensity, several experiments are conducted on Corel 5k dataset. In the proceeding section, the performance of proposed CMSD is evaluated by varying three parameters i.e. color quantization level $N_c \in \{192, 128, 108, 72\}$, orientation quantization level $N_0 \in \{6, 12, 18, 24, 30, 36\}$ and intensity quantization level $N_I \in \{10, 15, 20$. These parameters N_c, N_0 and N_I are the critical parameters which significantly affect the ARP & ARR of proposed descriptor. Table 1-3 represents the average retrieval accuracy obtained on Corel 5k by varying N_c and N_0 while N_I is kept constant value 10, 15 and 20 in Table I, II and III, respectively. It is observed that highest precision 63.62% is obtained when $N_c = 72$ and $N_0 = 30$ in Table I, highest precision 63.95% is obtained when $N_c = 128$ and $N_0 = 30$

TABLE 6. Comparison of proposed method by using different methods used to represent intensity information on corel 1k, corel 5k and corel 10k datasets.

Dataset	Performance	Methods used for intensity information	
		V component of HSV	PWM
Corel 1k	Precision (%)	78.54	77.25
	Recall (%)	9.42	9.27
Corel 5K	Precision (%)	63.14	59.53
	Recall (%)	7.57	7.14
Corel 10k	Precision (%)	50.25	46.75
	Recall (%)	6.03	5.61

TABLE 7. Performance of proposed method on corel5k, corel10k and corel1k dataset at different combinations of color, orientation and intensity feature vectors.

Dataset	Performance	Color Feature Only	Orientation Feature Only	Intensity Feature Only	Orientation + Intensity Features	Color + Orientation Features	Color + Intensity Features	Color + Orientation + Intensity Features
Corel 1k	Precision (%)	75.70	53.83	50.79	57.54	76.00	77.16	78.54
	Recall (%)	8.92	6.32	5.95	6.88	9.12	9.24	9.42
Corel 5k	Precision (%)	58.65	24.82	29.97	37.13	60.78	61.75	63.14
	Recall (%)	7.03	2.97	3.59	4.45	7.29	7.41	7.57
Corel 10k	Precision (%)	45.77	17.50	22.77	27.98	47.88	50.12	50.25
	Recall (%)	5.49	2.10	2.73	3.35	5.74	6.01	6.03

in Table II and highest precision 64.16% is obtained when $N_c = 128$ and $N_0 = 30$ in Table III with the corresponding feature vector of sizes 112, 173 and 178, respectively. However, in order to balance the ARP and computational cost in proposed descriptor the value of N_c is set to 72, N_0 to 6 and N_I to 10 having dimensions of 88 ($72 + 6 + 10 == 88$). In Table I-III, ARP slightly decreases when N_c is set to 128, N_I is set to 10, 15 and 30 respectively, and N_0 is increased from 6 to 36. This is most likely due to perceptive inability of human visual system to discriminate simultaneously varying edge orientations. In few cases irregular behavior of ARP observed e.g. in Table 1, when $N_c = 72$ and $N_I = 10$, ARP increases for N_0 from 6 to 18 and then decreases at $N_0 = 24$. This is most likely due to noise structures being added in samples due to quantization which in turn increases the inter-class variability and decreases the representative ability of proposed CMSD. Table IV presents the ARP obtained on Corel 5k by varying N_c and N_0 in RGB color space. Here $N_c \in \{16, 32, 64, 128\}$, $N_0 \in \{6, 12, 18, 24, 30, 36\}$ and N_I is set to 10. It is observed that highest ARP is achieved 59.21% in Table IV when $N_c = 128$, $N_0 = 30$ and $N_I = 10$. It is evident from Table 1,2,3, and 4 that the results achieved in HSV color space are better than those obtained in RGB color space. Because the color channels in RGB color space are extremely correlated therefore neglects the natural color properties and is also perceptually non-uniform. However, HSV color space not only considers the natural color properties but is also strongly related to human perception. Therefore, significant improvement in average retrieval accuracy has been observed in HSV color space.

Fig 3, 4 and 5 graphically analyze the relationship between the numbers of images retrieved (I_R) and ARP of proposed CMSD as compared to state-of-the-art methods on Corel 1k, 5k and 10k respectively. It has been observed that false positive rate (FPR) increases as the number of images retrieved increases, consequently, ARP of proposed CMSD decreases. However, the performance of proposed CMSD is much better than several state-of-the-art methods because CMSD

TABLE 8. Category wise comparison of proposed method with other state of the art descriptors on corel1k dataset.

Category	Performance	MTH	MSD	CDH	SED	MCMCM	LeNET-F6 [74]	ENN [75]	Proposed
African	Precision	69.17	83.33	77.50	82.50	69.75	80	85	86.66
	Recall	8.30	10.00	9.30	9.90	6.00	8.00	9.00	10.40
Beach	Precision	61.67	43.33	56.67	28.33	54.25	60	75	42.08
	Recall	7.40	5.20	6.80	3.40	4.40	6.5	7.00	5.05
Building	Precision	45.83	63.33	47.50	47.50	63.95	75	70	81.66
	Recall	5.50	7.60	5.70	5.70	4.30	8.00	6.00	9.80
Bus	Precision	68.33	76.67	71.67	73.33	89.65	80	75	81.66
	Recall	8.20	9.20	8.60	8.80	8.00	9.00	7.00	9.80
Dinosaur	Precision	100.00	100.00	100.00	90.00	98.70	100	100	100
	Recall	12.00	12.00	12.00	10.80	11.40	12.00	12.00	12.00
Elephant	Precision	70.83	65.00	62.50	55.00	48.80	85	75	72.08
	Recall	8.50	7.80	7.50	6.60	3.35	10.00	7.00	8.65
Flower	Precision	75.00	86.67	60.83	72.50	92.30	75	80	84.16
	Recall	9.00	10.40	7.30	8.70	8.12	8.00	8.00	10.10
Horse	Precision	100.00	97.50	91.67	62.50	89.45	90	85	94.16
	Recall	12.00	11.70	11.00	7.50	8.40	11.00	9.00	11.30
Mountain	Precision	39.17	29.17	44.17	40.00	47.30	60	65	50.00
	Recall	4.70	3.50	5.30	4.80	3.90	6.5	5.00	6.00
Food	Precision	52.50	76.67	45.00	64.17	70.90	75	55	92.91
	Recall	6.30	9.20	5.40	7.70	6.15	8.00	3.00	11.15
Average	Precision	68.25	72.167	65.75	61.58	72.5	78	76.5	**78.54**
	Recall	8.19	8.66	7.89	7.39	6.40	8.7	7.3	**9.42**

TABLE 9. Comparison of proposed method with regular pattern-based state of the art descriptors on corel5k and corel10k datasets.

Dataset	Performance	Methods						
		TCM	GLCM	MTH	MSD	SED	MS-LSP	CMSD
Corel-5k	Precision (%)	27.36	38.08	49.98	55.92	56.35	56.77	63.14
	Recall (%)	3.28	1.08	6.00	6.71	21.36	23.52	7.57
Corel-10k	Precision (%)	20.42	31.95	40.87	45.62	46.48	47.13	50.25
	Recall (%)	2.45	0.63	4.91	5.48	18.36	20.22	6.03

incorporates high-level semantics and spatial layout of contents in an image by establishing correlations between low level visual features.

Table 5 presents the comparative analysis of retrieval accuracy obtained on all the three datasets by using different similarity or distance metrics of different categories. We have used two different categories of distance matrices i.e. bin-by-bin and weighted type for comparative analysis. Bin-by-bin includes Manhattan distance (L1), Euclidian distance (L2), square chord and histogram intersection. Extended Canberra and Weighted L1 belongs to weighted type. We have measured distance between trained and test features by using

TABLE 10. Comparison of proposed method with classic texture, color and shape-based state of the art descriptors on corel5k and corel10k datasets.

Dataset	Performance	Methods									
		EOAC	Gabor	EHD	Wavelet transform	LBP+ wavelet	Color moment	CDH	CAC	Soc-GMM	CMSD
Corel-5k	Precision (%)	31.23	36.22	39.46	40.22	48.44	35.82	57.23	49.05	51.80	63.14
	Recall (%)	3.74	4.35	4.74	4.82	28.87	4.29	6.87	5.89	6.22	7.57
Corel-10k	Precision (%)	23.36	29.15	32.31	31.45	35.01	26.38	45.24	40.94	47.25	50.25
	Recall (%)	2.81	3.50	3.88	3.77	18.59	3.16	5.43	4.92	5.67	6.03

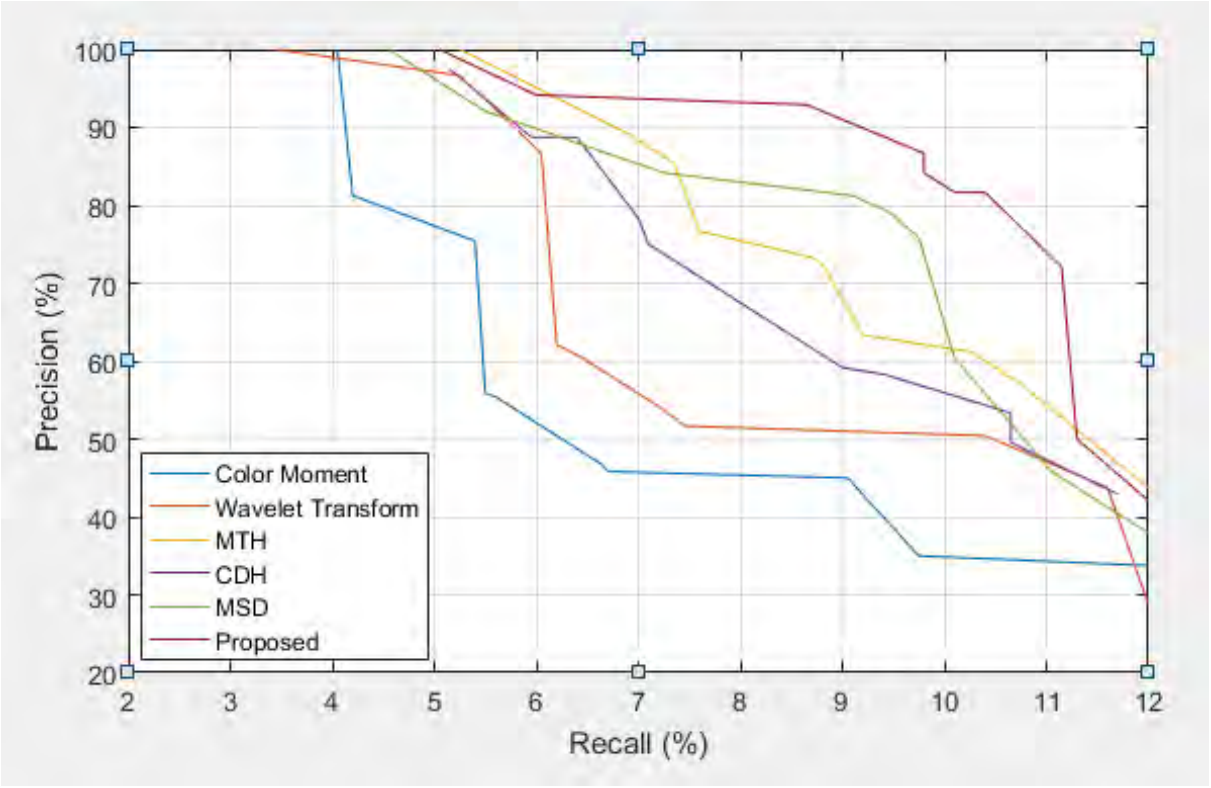

FIGURE 6. Precision-Recall graph of proposed CMSD is compared to other state-of-the-art methods on corel1k.

Manhattan distance (L1) in order to balance performance and computational cost in proposed CMSD. L1 performed better, achieved higher average retrieval accuracy i.e. 78.54% on Corel 1k, 63.14% on Corel 5k and 50.25% on Corel 10k. Manhattan distance (L1) is efficient because it does not carry out of any square and square root operations. It is most appropriate distance metric for large image databases. L1 is preferred over L2 because L2 usually fails to perform better on very small and larger distance values and square operations also increase computational cost. It has been observed from Table 5 that minimum average retrieval accuracy on Corel 5k is obtained by using histogram intersection. Histogram intersection focuses only on the peaks of histogram thereby ignoring the empty bins which can also carry some significant information required for image retrieval. On Corel 5k and 10k minimum average retrieval accuracy is obtained by using Extended Canberra because it lacks the semantic correlations between bins. Furthermore, it is sensitive to minor changes approaching to zero therefore; many false mismatches are likely to appear.

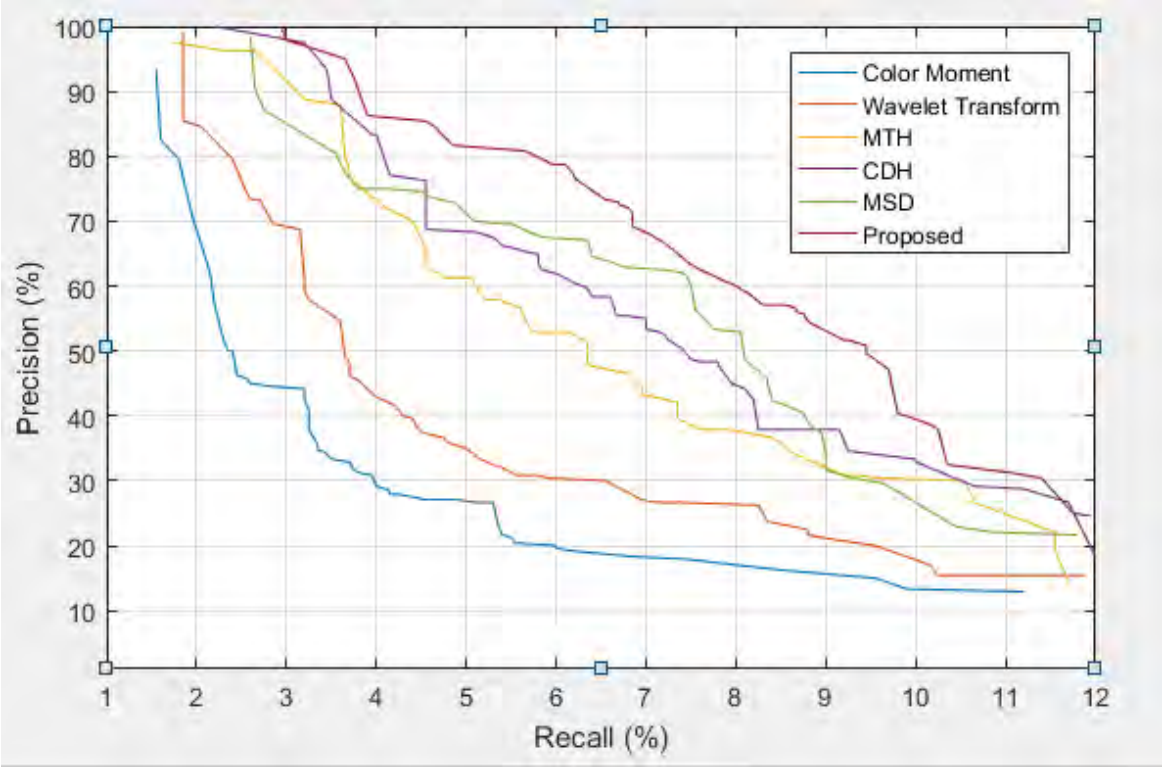

FIGURE 7. Precision-Recall graph of proposed CMSD is compared to other state-of-the-art methods on corel5k.

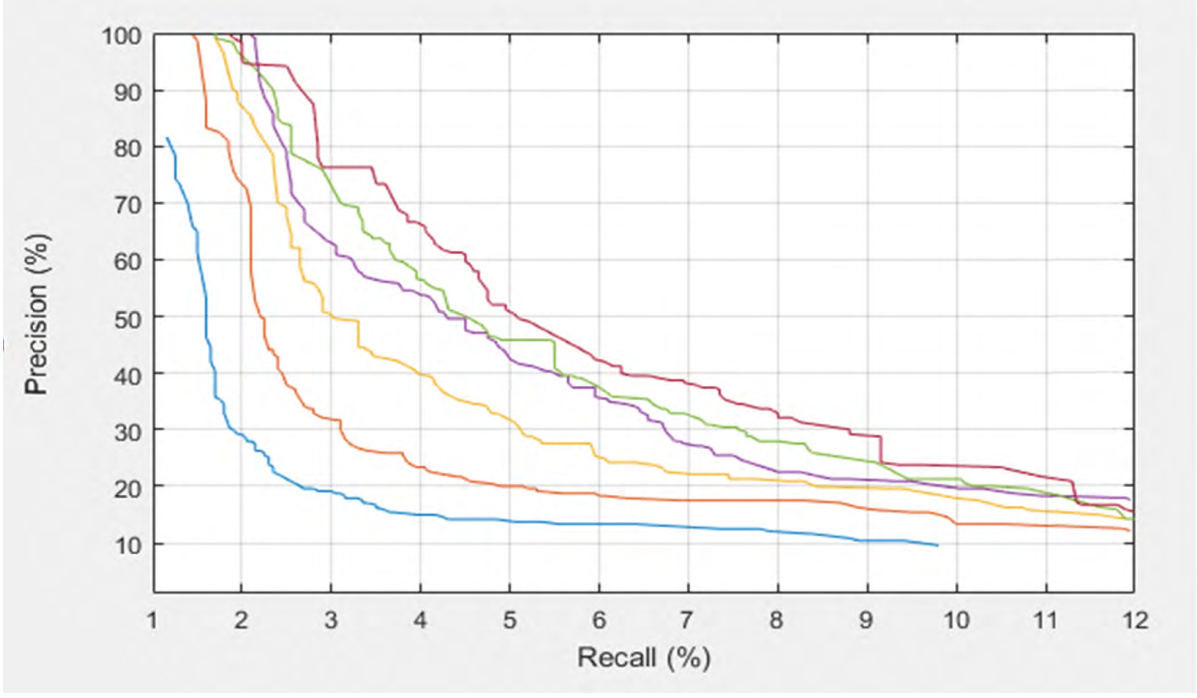

FIGURE 8. Precision-Recall graph of proposed CMSD is compared to other state-of-the-art methods on corel10k.

To analyze the performance of CMSD in presence of correlated intensity information, two different intensity extraction methods are adopted. In the first method V component of HSV is used for incorporating intensity information. In the second method Probability Weighted Moment (PWM) is used for extracting intensity information. Table 6 shows

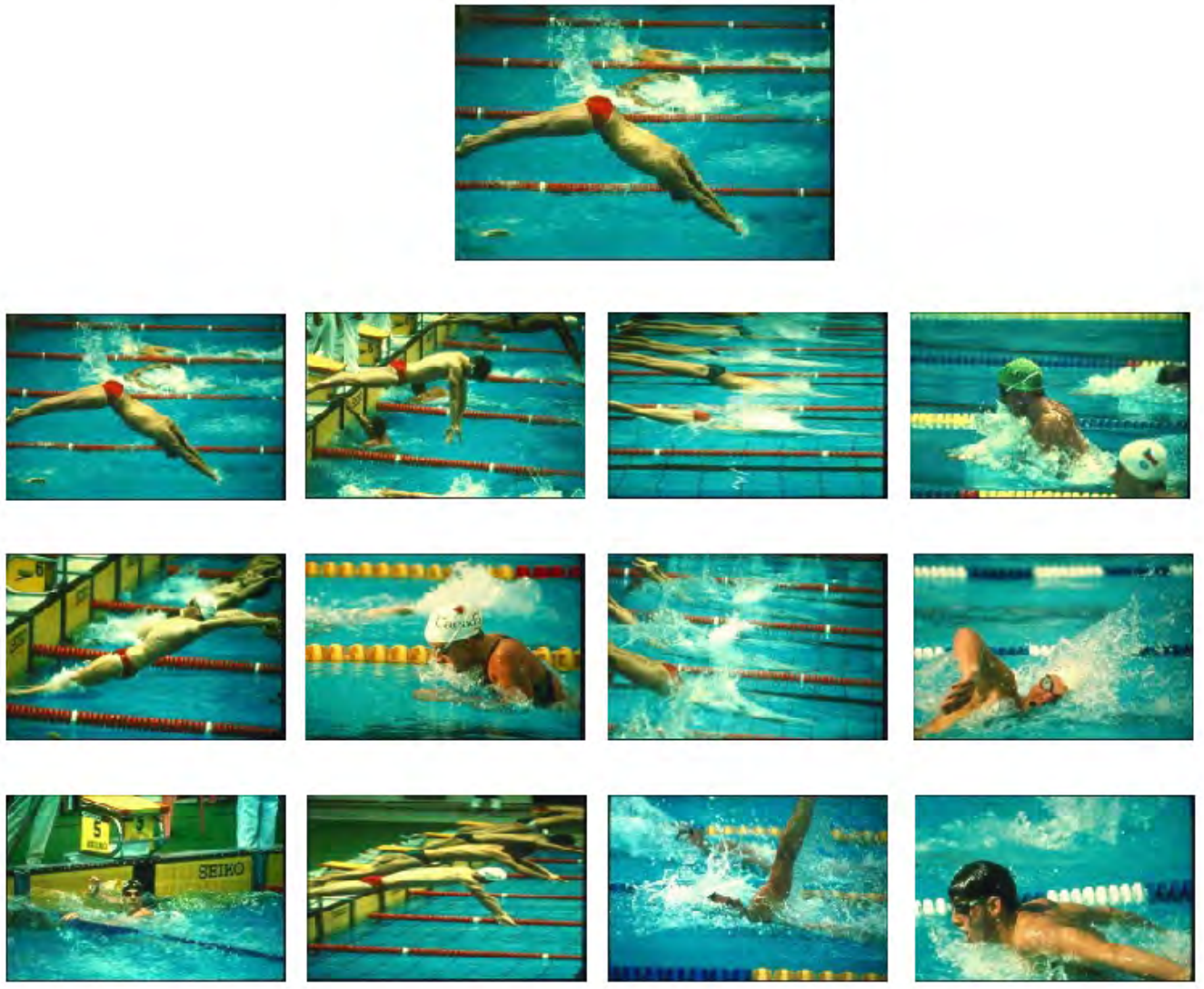

FIGURE 9. Sample image is selected from Swimmers category and top 12 relevant images are retrieved from corel5k dataset.

ARP and ARR comparison of using V component of HSV color space or PWM as intensity information using Corel 1k, Corel 5k and Corel 10k. Results evaluation shows that there is a slight difference between the ARP of PWM and Value V intensity extraction method. However, PWM is computationally extensive and also has higher time complexity. Therefore, V component of HSV is an appropriate choice for intensity information extraction based on the analysis of results in Table VI. Further experiments are conducted using V component of HSV color space.

Since, there are three different feature vectors being used for the image retrieval task in the proposed system. Therefore, we have analyzed the performance of CMSD using 7 different variants based on all three possible combinations of color, orientation and intensity feature vectors. The performance evaluation using 7 different variants shows that each combination of feature vector has a different contribution in retrieval-based tasks. The ARP and ARR of proposed descriptor using all possible combination of three feature vectors is given in Table 7. Result evaluation shows that comparatively better retrieval accuracy is obtained when combination of all three feature vectors is used. Similarly, poor results are obtained using only a single vector. It is evident from Table VII that the contribution of intensity feature is low however, this contribution is very vital. Natural Images enriched with color and contrast information. Therefore, contrast or brightness features are very helpful to distinguish between two different images having same contrast or two same images having different contrast information. However, the results of proposed descriptor are better in all the possible combinations compared to the state-of-the-art descriptors performing the same tasks on Corel 1k, 5k and 10k.

The semantic robustness of proposed system is tested on each semantic category of Corel 1k in comparison with

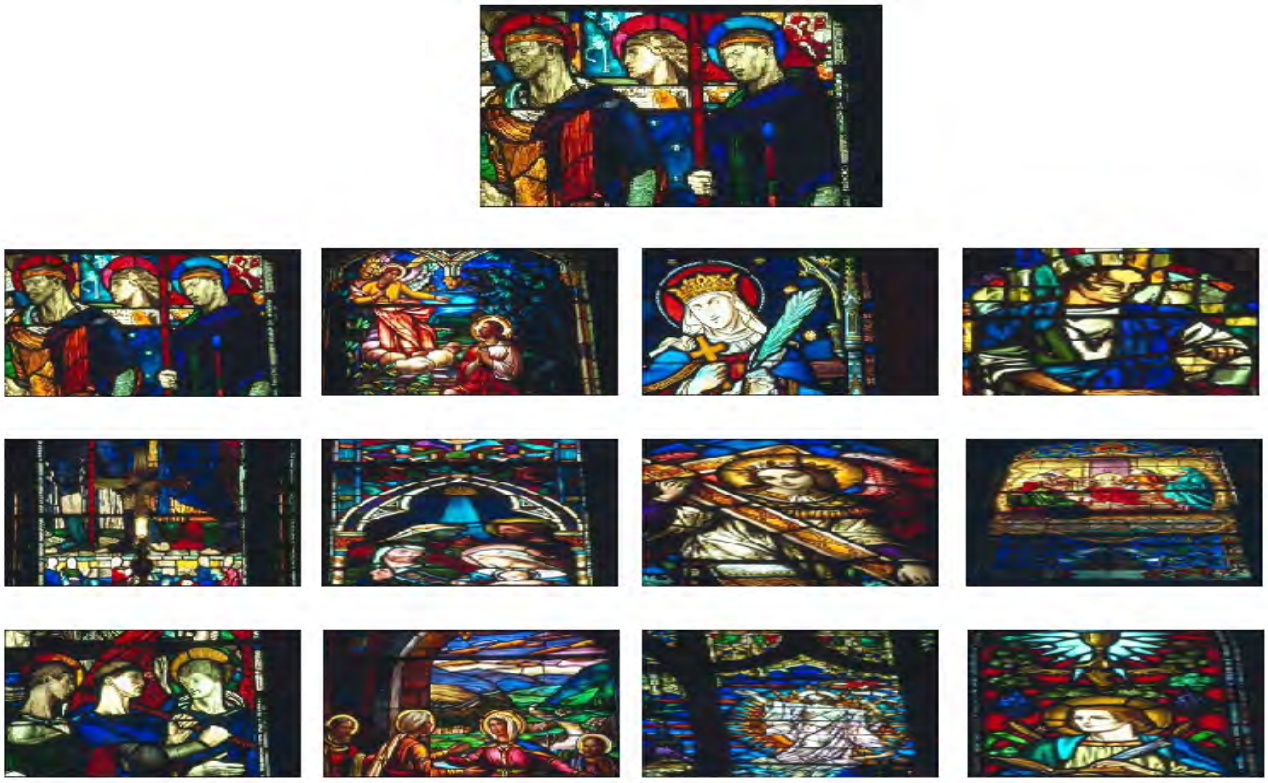

FIGURE 10. Sample image is selected from Glass category and top 12 relevant images are retrieved from corel5k dataset.

MTH [57], MSD [35], CDH [54], SED [65], MCMCM [46], LeNET-F6 [74] and ENN [75] descriptors. The results evaluation in Table 8 shows the outperformance of proposed CMSD compared to the state-of-the-art methods in most of the enriched semantic categories of Corel 1k such as African, Building, Bus, Dinosaur, Flower, Horse, Mountain and Food. In category-wise evaluation, there are more semantically enriched classes like beach, mountain and elephant are exist. The ARP is slightly less i.e. 42.08% on beach class as compared to MSD, CDH, MTH, LeNET-F6, ENN and MCMCM. Certain images of mountain category are retrieved as of beach category and vice versa, resultantly the decrease in ARP of both categories has been observed. Texture is significantly discriminative feature in images of beach and mountain category. Moreover, textural structures (overlapped backgrounds) change slightly in the images of beach category. However, human visual system cannot perceive slight textural variations. The ARP reaches the highest i.e. 100% on dinosaur's category showing strong retrieval performance of the system. The proposed CMSD extracts micro-maps consisting of perceptually uniform regions as a general visual attribute based on colors, texture orientation and intensity information. Moreover, the proposed CMSD analyses all the possible patterns to incorporate spatial distribution of contents of an image in a resultant microstructure image. Therefore, the performance of proposed CMSD is considerably better on other categories as compared to other state-of-the-arts using semantically enriched Corel 1k dataset.

E. COMPARISON OF CMSD WITH OTHER CBIR METHODS

There exist certain kinds of coarse, fine, regular and irregular patterns of textures in natural images. Therefore, in order to discriminate these images on the basis of their texture, some classic texture-based descriptors are proposed such as GLCM, Wavelet transform features, Gabor features and EHD. These texture-based descriptors showed favorably good results in presence of homogeneous and regular patterns. However, the performance is degraded on irregular texture patterns. Therefore, these descriptors can't be used where there are patterns containing both regular and irregular textures. Table 9 shows performance comparison of proposed system with regular pattern-based state of the arts descriptors including TCM [64], GLCM [46], MTH [57], MSD [35], SED [65] and MS-LSP [1] respectively. Results comparison shows an increase in retrieval accuracy of 35.78%, 25.06%, 13.16%, 7.22%, 6.79%, 6.37% compared to TCM, GLCM, MTH, MSD, SED and MS-LSP using Corel 5k. Similarly, an increase of 29.83%, 18.3%, 9.38%, 4.63%, 3.77%, and 3.12% is observed on Corel 10k on TCM, GLCM, MTH, MSD, SED and MS-LSP respectively. Table 10 shows performance comparison of proposed descriptor on classic texture, color and shape-based descrip-

FIGURE 11. Sample image is selected from Tractor category and top 12 relevant images are retrieved from corel5k dataset.

tors using Corel 5k and Corel 10k. The retrieval accuracy of proposed system in comparison to other category is 31.91%, 26.92%, 23.68%, 22.92%, 14.7%, 27.32%, 5.91%, 14.09 and 12.06% greater on EoAC [66], Gabor [44], EHD [48], wavelet transform [45], LBP+Wavelet [56], Color moment [67], CDH [54], CAC [39] and SOC-GMM [40] respectively on Corel 5k. Similarly, on Corel 10k the retrieval accuracy is 26.89%, 21.1%, 17.94%, 18.8%, 15.24%, 23.87%, 5.01%, 9.31% and 3% on EoAC, Gabor, EHD, wavelet transform, LBP+Wavelet, Color moment, CDH, CAC and SOC-GMM respectively. The increase in retrieval accuracy and recall is evidence of outperformance of proposed system compared to the classic shape, texture, color-based systems and regular patterns.

Fig 6-8 graphically represents the precision-recall curve of proposed CMSD and other state-of-the-art methods. Recall (%) is represented on x-axis and precision (%) is represented on y-axis. Generally, precision-recall curves are presented in opposite ways. For instance, if a method has higher ARP, then curve will move away from the origin of co-ordinates. A precision-recall curve will be short if a method has lower ARP. Instability in the performance of a method is indicated by multiple turning points in the curve. The part of curves will be overlapped if the performance of two methods is similar in certain image category. CMSD performs much better on all the three datasets as compared to the other state-of-the-art methods as shown in Fig 6-8. Fig 6-8 shows that the precision-recall curve of CMSD lies away from the origin of the co-ordinates. The curve of CMSD has a few numbers of turning points and it is also spanned fine for all image categories of each dataset.

A simple MATLAB GUI application is developed in which the system retrieves top 12 images in response to the input image provided by the user. Dominant information on color,

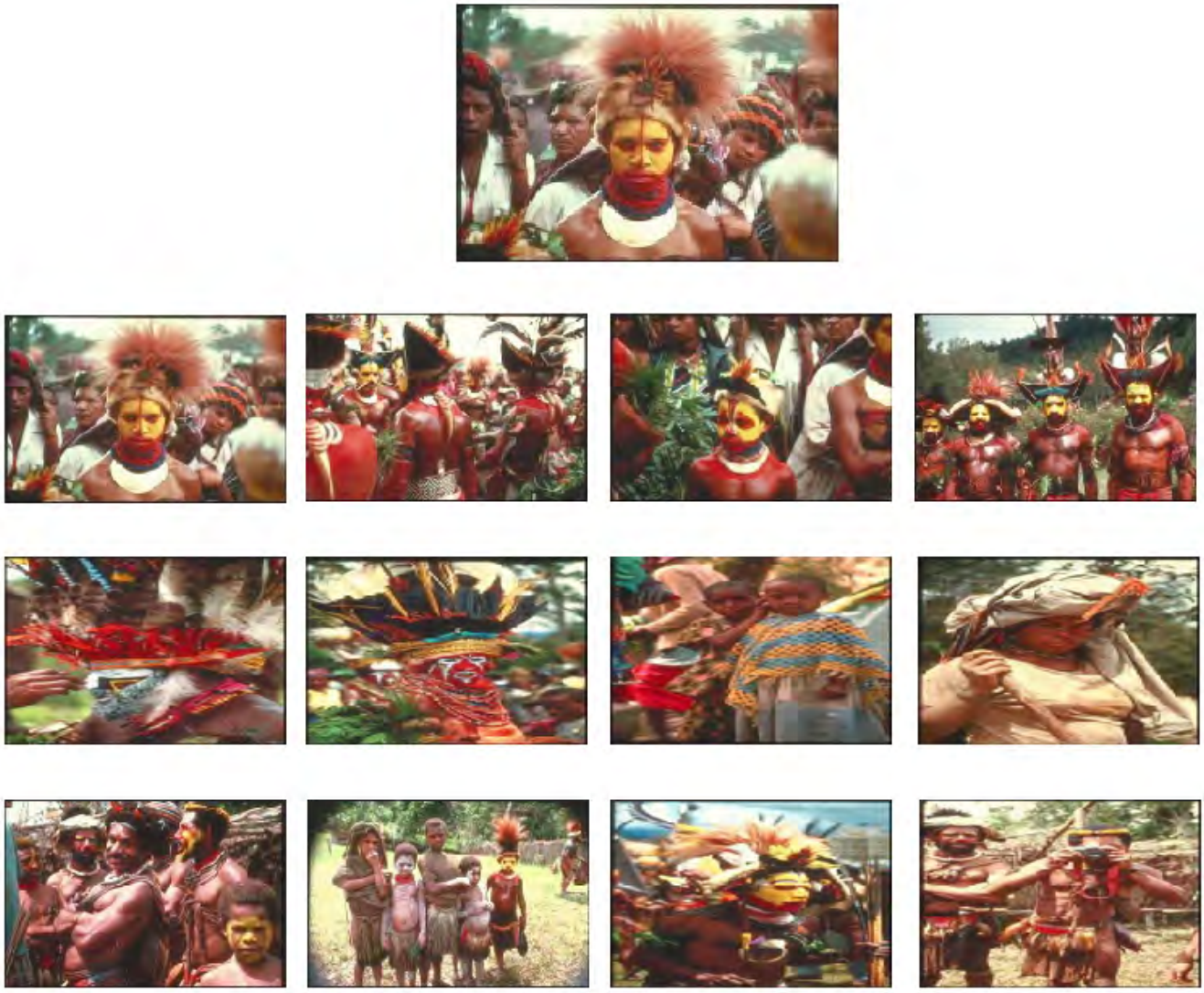

FIGURE 12. Sample image is selected from African category and top 12 relevant images are retrieved from corel5k dataset.

shape, texture (regular and irregular) contained in images is used for query to validate the performance of proposed system. Fig 9-13 shows the top 12 images retrieved against the query images taken from the swimmer, glass, tractor, African and leaf category respectively. It has been observed in Fig 9 that all the retrieved images have similar color and spatial layout of the content. Fig 10 illustrates that all the retrieved images belong to the glass category against the query image from the glass category. All the retrieved images in Fig 11 have similar color, and spatial layout of content as the query image taken from tractor category. Similarly, all the retrieved images belong to African category in Fig 12. Fig 13 shows the color and texture discriminative power of proposed CMSD in images retrieved against the query image taken from leaf category. Leaf images have richer natural color and texture content therefore all the retrieved images belong to the query image. In TCM the co-occurrence matrix is represented by using five special patterns of textons. The ultimately generated co-occurrence matrix is used to derive the energy, homogeneity, entropy and contrast features. However, the image features can be represented effectively by using only 4 features. MTH showed better results compared to TCM by the integration of co-occurrence matrix with histogram techniques. In existing descriptors like MTH, SED and STH the correlation between texture orientation and color is used for representing features of image however the intensity was missing. Since the correlations obtained on the basis of contrast information is not taken in account; the retrieval performance in MTH, SED and STH descriptors is limited. Similarly, in MS-LSP 20 local structural patterns in multiple scale and color is used for feature representation including the fine and coarse grain information of image. Although detailed information of image is incorporated however this leads to curse of dimensionality in MS-LSP.

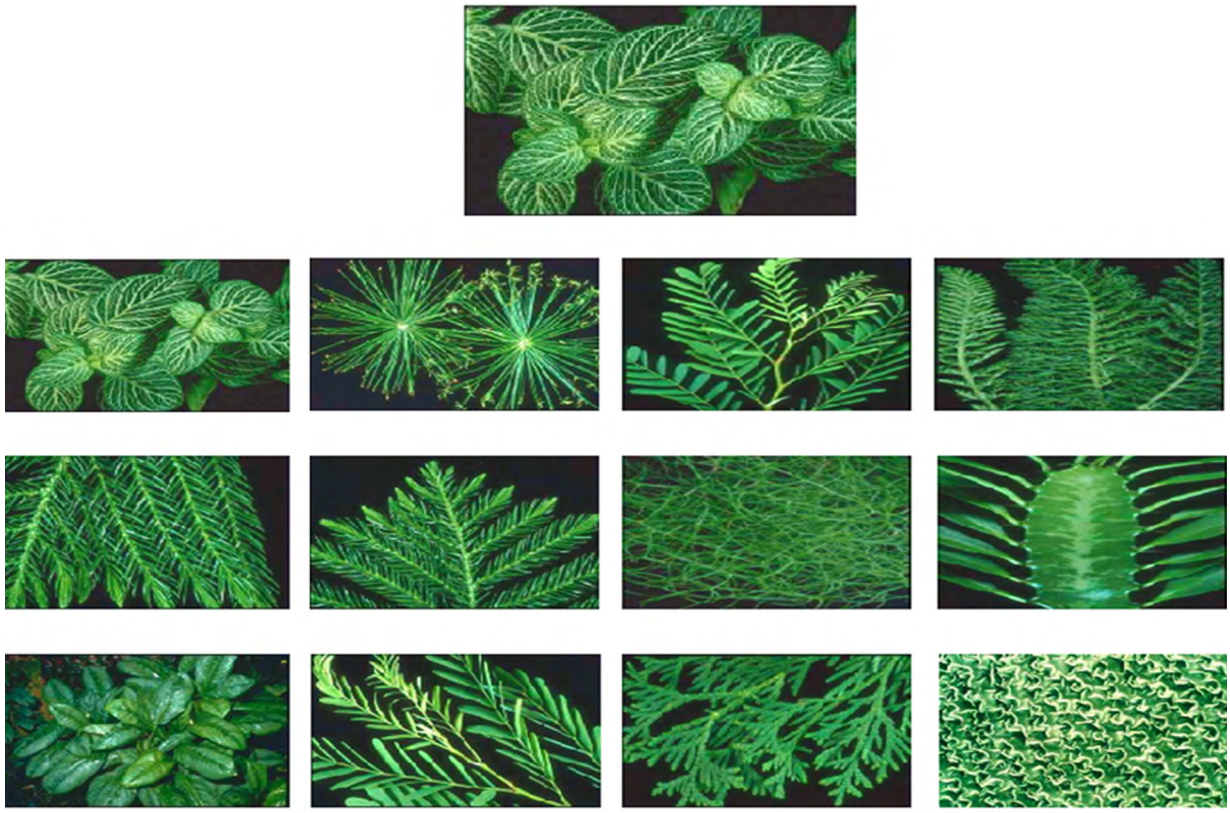

FIGURE 13. Sample image is selected from Leaf category and top 12 relevant images are retrieved from corel5k dataset.

Later, MTSD was proposed for extracting local spatial structure information from color, orientation and intensity. However, correlation was missing in MTSD which limited the retrieval accuracy of the method.

The performance of texture descriptors such as GLCM, EHD, Gabor features is degraded on natural images because texture only represents partial attributes of an image. Gabor filter is extensively used to extract textural features in image retrieval systems because of its high correspondence between its frequency and orientation representations and human visual system. EOAC does not considered textural characteristics of an image. MSD considers underlying colors identified in microstructures in correlation with similar edge orientation. However, it ignores contrast information of an image. CDH considers color differences based on only color and edge orientation information and efficiently encodes them. CDH gives each color and edge orientation equally important however, it is perceptually least significant. The Color Moment (CM) method characterizes only the color-spatial information of those pixels which are close to boundary of an image. LBP considers the textural characteristics of an image while ignoring the color features. The underlying segmentation process in SoC-GMM is sensitive to noise and illumination. The proposed CMSD method addresses the aforementioned drawbacks prevailing in existing methods by 1) finding the correlation between color, orientation and intensity information for representing image feature information. CMSD 2) represents high level semantics by identifying microstructure image by correlating color, texture orientation and intensity features. In CMSD three micro-maps i.e. micro-color, micro-orientation and micro-intensity maps are obtained by correlating each microstructure images with quantized color, orientation and intensity features respectively. Final features vector representing significant semantic characteristics of an input image is obtained by concatenating the resultant three micro-maps. It represents only significant points lying within the perceptual uniform region and ignores the irrelevant details.

IV. CONCLUSION

In this paper Correlated Microstructure Descriptor (CMSD) is presented to incorporate high level semantic concepts for image retrieval. The proposed descriptor named CMSD relies on various aspects of input image by correlating color, texture orientation and intensity information. The correlated information is mapped with microstructure to extract all fine detail within subject areas. Performance comparisons with classic color, texture and shape-based descriptors show an evident improvement of CMSD descriptor on all three datasets i.e. Core 1k, core 5k and core 10k. Richer multi-directional edge orientation map is constructed by quantizing the edges obtained by 4-dimensional Sobel operator into 6 levels.

Moreover, the local information is obtained by quantizing the v component of HSV into 10 levels. An obvious improvement comes from integration of correlated visual features on the basis of proximity and continuation relations. CMSD requires no learning and segmentation. Furthermore, the discriminative power of CMSD compared to state-of-the-art renders the descriptor a good edition to the classic color, shape and texture based CBIR systems.

ACKNOWLEDGMENT

The authors, therefore, acknowledge with thanks DSR technical and financial support.

REFERENCES

[1] J. Ahmad, M. Sajjad, S. Rho, and S. W. Baik, "Multi-scale local structure patterns histogram for describing visual contents in social image retrieval systems," *Multimedia Tools Appl.*, vol. 75, no. 20, pp. 12669–12692, 2016.
[2] J. Tang and S. T. Acton, "A decentralized image retrieval system for education," in *Proc. Syst. Inf. Eng. Design Symp.*, Apr. 2003, pp. 7–12.
[3] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, "Content-based image retrieval system for pulmonary nodules: Assisting radiologists in self-learning and diagnosis of lung cancer," *J. Digit. Imag.*, vol. 30, no. 1, pp. 63–77, 2017.
[4] D. Binu and P. Malathi, "Multi model based biometric image retrieval for enhancing security," *Indian J. Sci. Technol.*, vol. 8, no. 35, pp. 1–10, 2015.
[5] S. Deniziak and T. Michno, "Content based image retrieval using query by approximate shape," in *Proc. Federated Conf. Comput. Sci. Inf. Syst. (FedCSIS)*, vol. 8, Sep. 2016, pp. 807–816.
[6] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "An ontology approach to object-based image retrieval," in *Proc. Int. Conf. Image Process.*, vol. 3, Sep. 2003, pp. II-511–II-514.
[7] J. Ren, Y. Shen, and L. Guo, "A novel image retrieval based on representative colors," in *Proc. Image Vis. Comput.*, 2003, pp. 1–6.
[8] S. Kulkarni and B. Verma, "Fuzzy logic based texture queries for CBIR," in *Proc. 5th Int. Conf. Comput. Intell. Multimedia Appl.*, Sep. 2003, pp. 1–6.
[9] N. Maillot, M. Thonnat, and A. Boucher, "Towards ontology-based cognitive vision," *Mach. Vis. Appl.*, vol. 16, no. 1, pp. 33–40, 2004.
[10] S. Wang, J. Zhang, T. X. Han, and Z. Miao, "Sketch-based image retrieval through hypothesis-driven object boundary selection with HLR descriptor," *IEEE Trans. Multimedia.*, vol. 17, no. 7, pp. 1045–1057, Jul. 2015.
[11] N. M. Asiri, N. Alhumaidi, and N. Alosaim, "Combination of histogram of oriented gradients and hierarchical centroid for sketch-based image retrieval," in *Proc. 2nd Int. Conf. Comput. Sci. Comput. Eng. Soc. Media (CSCESM)*, Sep. 2015, pp. 149–152.
[12] S. M. Devi and C. Bhagvati, "Connected component in feature space to capture high level semantics in CBIR," in *Proc. 4th Annu. ACM Bangalore Conf. (COMPUTE)*, 2011, pp. 1–5.
[13] Y. Sun and B. Bhanu, "Image retrieval with feature selection and relevance feedback," in *Proc. 17th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 3209–3212.
[14] H. Xie, Y. Ji, and Y. Lu, "An analogy-relevance feedback CBIR method using multiple features," in *Proc. Int. Conf. Comput. Problem-Solving (ICCP)*, Oct. 2013, pp. 83–86.
[15] X. Wu and D. Fu, "Apply hybrid method of relevance feedback and EMD algorithm in a color feature extraction CBIR system," in *Proc. Int. Conf. Audio, Lang. Image Process. (ICALIP)*, Jul. 2008, pp. 163–166.
[16] W. Bian and D. Tao, "Biased discriminant Euclidean embedding for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 545–554, Feb. 2010.
[17] X. S. Zhou and T. S. Huang, "Small sample learning during multimedia retrieval using BiasMap," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Dec. 2001, pp. I-11–I-17.
[18] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "Long term learning in image retrieval systems using case based reasoning," *Eng. Appl. Artif. Intell.*, vol. 35, pp. 26–37, Oct. 2014.
[19] J. M. dos Santos, E. S. de Moura, A. S. da Silva, and R. da Silva Torres, "Color and texture applied to a signature-based bag of visual words method for image retrieval," *Multimedia Tools Appl.*, vol. 76, no. 15, pp. 16855–16872, 2017.
[20] Y. Wang *et al.*, "Separable vocabulary and feature fusion for image retrieval based on sparse representation," *Neurocomputing*, vol. 236, pp. 14–22, Jul. 2017.
[21] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[22] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2004, pp. 2–9.
[23] R. Arandjelovic and A. Zisserman, "Three things everyone should know to improve object retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Apr. 2012, pp. 2911–2918.
[24] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
[25] N. Dalal and W. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2004, vol. 5, no. 3, pp. 886–893.
[26] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Directional local extrema patterns: A new descriptor for content based image retrieval," *Int. J. Multimedia Inf. Retr.*, vol. 1, no. 3, pp. 191–203, 2012.
[27] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognit.*, vol. 43, no. 3, pp. 706–719, 2010.
[28] S. R. Dubey, S. K. Singh, and R. K. Singh, "Local bit-plane decoded pattern: A novel feature descriptor for biomedical image retrieval," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 4, pp. 1139–1147, Jul. 2016.
[29] S. R. Dubey, S. K. Singh, and R. K. Singh, "Local wavelet pattern: A new feature descriptor for image retrieval in medical CT databases," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5892–5903, Dec. 2015.
[30] S. R. Dubey, S. K. Singh, and R. K. Singh, "Multichannel decoded local binary patterns for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4018–4032, Sep. 2016.
[31] M. Verma and B. Raman, "Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval," *Multimedia Tools Appl.*, vol. 77, no. 10, pp. 11843–11866, 2018.
[32] M. Verma and B. Raman, "Local tri-directional patterns: A new texture feature descriptor for image retrieval," *Digit. Signal Process.*, vol. 51, pp. 62–72, Apr. 2016.
[33] Z. Zeng, "A novel local structure descriptor for color image retrieval," *Information*, vol. 7, no. 1, pp. 1–14, 2016.
[34] G. Deep, L. Kaur, and S. Gupta, "Local mesh ternary patterns: A new descriptor for MRI and CT biomedical image indexing and retrieval," *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.*, vol. 6, no. 2, pp. 155–169, 2018.
[35] G.-H. Liu, Z.-Y. Li, L. Zhang, and Y. Xu, "Image retrieval based on microstructure descriptor," *Pattern Recognit.*, vol. 44, no. 9, pp. 2123–2133, 2011.
[36] B. Nan, Y. Xu, Z. Mu, and L. Chen, "Content-based image retrieval using local texture-based color histogram," in *Proc. IEEE 2nd Int. Conf. Cybern. (CYBCONF)*, vol. 7, Jun. 2015, pp. 399–405.
[37] C. Zain and M. Pratama, "Multimodal image retrieval using PLSA and microstructure descriptor," *J. Math. Stat. Appl.*, vol. 1, no. 1, pp. 17–31, 2016.
[38] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.
[39] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 762–768.
[40] S. Zeng, R. Huang, H. Wang, and Z. Kang, "Image retrieval using spatiograms of colors quantized by Gaussian mixture models," *Neurocomputing*, vol. 171, pp. 673–684, Jan. 2016.
[41] X. Tian, L. Jiao, X. Liu, and X. Zhang, "Feature integration of EODH and Color-SIFT: Application to image retrieval based on codebook," *Signal Process., Image Commun.*, vol. 29, no. 4, pp. 530–545, 2014.
[42] H.-H. Tsai, B.-M. Chang, P.-S. Lo, and J.-Y. Peng, "On the design of a color image retrieval method based on combined color descriptors and features," in *Proc. 1st IEEE Int. Conf. Comput. Commun. Internet (ICCCI)*, no. 1, Oct. 2016, pp. 392–395.
[43] S. Agarwal, "Content based image retrieval using discrete wavelet transform and edge histogram descriptor," in *Proc. Int. Conf. Inf. Syst. Comput. Netw.*, Mar. 2013, pp. 19–23.

[44] J. Han and K.-K. Ma, "Rotation-invariant and scale-invariant Gabor features for texture image retrieval," *Image Vis. Comput.*, vol. 25, no. 9, pp. 1474–1481, 2007.
[45] K. Jafari-Khouzani and H. Soltanian-Zadeh, "Radon transform orientation estimation for rotation invariant texture analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 1004–1008, Jun. 2005.
[46] A. Gebejes and R. Huertas, "Texture characterization based on grey-level co-occurrence matrix," in *Proc. Conf. Inform. Manage. Sci.*, 2013, pp. 375–378.
[47] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 1, pp. 25–39, Jan. 1983.
[48] B. S. Manjunath, P. Salembier, and T. Sikora, Eds., *Introduction to MPEG-7: Multimedia Content Description Interface*. Hoboken, NJ, USA: Wiley, 2002.
[49] D. John, S. T. Tharani, and K. SreeKumar, "Content based image retrieval using HSV-color histogram and GLCM," *Int. J. Adv. Res. Comput. Sci. Manage. Stud.*, vol. 2, no. 1, pp. 246–253, 2014.
[50] A. K. Bhunia, A. Bhattacharyya, P. Banerjee, P. P. Roy, and S. Murala. (2018). "A novel feature descriptor for image retrieval by combining modified color histogram and diagonally symmetric co-occurrence texture pattern." [Online]. Available: https://arxiv.org/abs/1801.00879
[51] P. Liu, J. M. Guo, K. Chamnongthai, and H. Prasetyo, "Fusion of color histogram and LBP-based features for texture image retrieval and classification," *Inf. Sci.*, vol. 390, pp. 95–111, Jun. 2017.
[52] M. Sajjad, A. Ullah, J. Ahmad, N. Abbas, S. Rho, and S. W. Baik, "Integrating salient colors with rotational invariant texture features for image representation in retrieval systems," *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4769–4789, 2018.
[53] N. Shrivastava and V. Tyagi, "Content based image retrieval based on relative locations of multiple regions of interest using selective regions matching," *Inf. Sci.*, vol. 259, pp. 212–224, Feb. 2014.
[54] G. H. Liu and J. Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognit.*, vol. 46, no. 1, pp. 188–198, 2013.
[55] J. K. Patil and R. Kumar, "Analysis of content based image retrieval for plant leaf diseases using color, shape and texture features," *Eng. Agricult., Environ. Food*, vol. 10, pp. 69–78, Apr. 2016.
[56] P. Srivastava and A. Khare, "Integration of wavelet transform, local binary patterns and moments for content-based image retrieval," *J. Vis. Commun. Image Represent.*, vol. 42, pp. 78–103, 2017.
[57] G.-H. Liu, L. Zhang, Y.-K. Hou, Z.-Y. Li, and J.-Y. Yang, "Image retrieval based on multi-texton histogram," *Pattern Recognit.*, vol. 43, no. 7, pp. 2380–2389, 2010.
[58] J. R. Smith and S.-F. Chang, "Single color extraction and image query," in *Proc. Int. Conf. Image Process.*, vol. 3, Oct. 1995, pp. 528–531.
[59] Q. J. Qiu, Y. Liu, D. W. Cai, and J. Z. Tan, "Modified color texton histogram for image retrieval," in *Proc. Int. Conf. Comput. Sci. Appl. (CSA)*, Dec. 2013, pp. 585–589.
[60] S. Dutta and B. B. Chaudhuri, "A color edge detection algorithm in RGB color space," in *Proc. Int. Conf. Adv. Recent Technol. Commun. Comput.*, 2009, pp. 337–340.
[61] A. Raza, H. Dawood, H. Dawood, S. Shabbir, R. Mehboob, and A. Banjar, "Correlated primary visual texton histogram features for content base image retrieval," *IEEE Access*, vol. 6, pp. 46595–46616, 2018.
[62] B. O. Sadiq, S. M. Sani, and S. Garba, "Edge detection: A collection of pixel based approach for colored images," *Int. J. Comput. Appl.*, vol. 113, no. 5, pp. 29–32, 2015.
[63] J. Qiu, H. Xu, Y. Ma, and Z. Ye. (2018). "PILOT: A pixel intensity driven illuminant color estimation framework for color constancy." [Online]. Available: https://arxiv.org/abs/1806.09248
[64] G. H. Liu and J. Y. Yang, "Image retrieval based on the texton co-occurrence matrix," *Pattern Recognit.*, vol. 41, no. 12, pp. 3521–3527, 2008.
[65] X. Wang and Z. Wang, "A novel method for image retrieval based on structure elements' descriptor," *J. Vis. Commun. Image Represent.*, vol. 24, no. 1, pp. 63–74, 2013.
[66] F. Mahmoudi, J. Shanbehzadeh, A.-M. Eftekhari-Moghadam, and H. Soltanian-Zadeh, "Image retrieval based on shape similarity by edge orientation autocorrelogram," *Pattern Recognit.*, vol. 36, no. 8, pp. 1725–1736, 2003.
[67] Z.-C. Huang, P. P. K. Chan, W. W. Y. Ng, and D. S. Yeung, "Content-based image retrieval using color moment and Gabor texture feature," in *Proc. Mach. Learn.*, Jul. 2010, pp. 719–724.
[68] A. Raza, T. Nawaz, H. Dawood, and H. Dawood, "Square texton histogram features for image retrieval," *Multimedia Tools Appl.*, vol. 78, no. 3, pp. 2719–2746, 2018.
[69] L. Yu, L. Feng, H. Wang, L. Li, Y. Liu, and S. Liu, "Multi-trend binary code descriptor: A novel local texture feature descriptor for image retrieval," *Signal, Image Video Process.*, vol. 12, no. 2, pp. 247–254, 2018.
[70] A. Gordo, J. Almazan, J. Revaud, and D. Larlus, "End-to-end learning of deep visual representations for image retrieval," *Int. J. Comput. Vis.*, vol. 124, no. 2, pp. 237–254, 2017.
[71] X. Lu, X. Zheng, and X. Li, "Latent semantic minimal hashing for image retrieval," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 355–368, Jan. 2017.
[72] J. Song, L. Gao, L. Liu, X. Zhu, and N. Sebe, "Quantization-based hashing: A general framework for scalable image and video retrieval," *Pattern Recognit.*, vol. 75, pp. 175–187, Mar. 2018.
[73] R. C. Gonzales and E. R. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2007.
[74] H. Liu, B. Li, X. Lv, and Y. Huang, "Image retrieval using fused deep convolutional features," *Procedia Comput. Sci.*, vol. 107, pp. 749–754, Jan. 2017.
[75] R. Ashraf, K. Bashir, A. Irtaza, and M. Mahmood, "Content based image retrieval using embedded neural networks with bandletized regions," *Entropy*, vol. 17, no. 6, pp. 3552–3580, 2015.

HUSSAIN DAWOOD received the M.S. and Ph.D. degrees in computer application technology from Beijing Normal University, Beijing, China, in 2012 and 2015, respectively. He is currently an Assistant Professor with the Department of Computer and Network Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia. His current research interests include image processing, pattern recognition, and feature extraction.

MONAGI HASSAN ALKINANI received the master's and Ph.D. degrees from the University of Western Ontario, London, Canada, in 2011 and 2016, respectively. He is currently an Assistant Professor with the College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia. His research interests include image processing and image retrieval.

AHMAD RAZA received the B.S. degree in software engineering and the M.S. degree from the University of Engineering & Technology Taxila, Pakistan, in 2016 and 2018, respectively. His research interests include image processing, machine learning, computer vision, pattern recognition, and neural networks.

HASSAN DAWOOD received the M.S. and Ph.D. degrees in computer application technology from Beijing Normal University, Beijing, China, in 2012 and 2015, respectively. He is currently an Assistant Professor with the Department of Software Engineering, University of Engineering and Technology, Taxila, Pakistan. His research interests include image restoration, feature extraction, and image classification.

RUBAB MEHBOOB received the B.S. degree in software engineering from Fatima Jinnah Women University, Rawalpindi, Pakistan, in 2015, and the M.S. degree from the University of Engineering & Technology, Taxila, Pakistan, in 2018. Her research interests include image processing, machine learning, computer vision, pattern recognition, and neural networks.

SIDRA SHABBIR received the B.S. degree in software engineering from Fatima Jinnah Women University, Rawalpindi, Pakistan, in 2015, and the M.S. degree from the University of Engineering & Technology, Taxila, Pakistan, in 2018. Her research interests include image processing, machine learning, computer vision, pattern recognition, and neural networks.