

## Research Article

# Image Retrieval Based on a Multi-Integration Features Model

Kai Chu and Guang-Hai Liu 

*College of Computer Science and Information Technology, Guangxi Normal University, Guilin 541004, China*

Correspondence should be addressed to Guang-Hai Liu; liuguanghai009@163.com

Received 11 November 2019; Revised 22 January 2020; Accepted 29 January 2020; Published 9 March 2020

Academic Editor: Yang Li

Copyright © 2020 Kai Chu and Guang-Hai Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Feature integration theory can be regarded as a perception theory, but the extraction of visual features using such a theory within the CBIR framework is a challenging problem. To address this problem, we extract the color and edge features based on a multi-integration features model and use these for image retrieval. A novel and highly simple but efficient visual feature descriptor, namely, a **multi-integration features histogram**, is proposed for image representation and content-based image retrieval. First, a color image is converted from the RGB to the HSV color space, and the color features and color differences are extracted. Then, the color differences are calculated to extract the edge features using a set of simple integration processes. Finally, combining the color, edge, and spatial layout features allows representing the image content. Experiments show that our method produces results comparable to existing and well-known methods on three datasets that contain 25,000 natural images. The performances are significantly better than that of the BOW histogram, local binary pattern histogram, histogram of oriented gradient, and multi-tenton histogram, with performances similar to the color volume histogram.

## 1. Introduction

The CBIR technique originated in the 1990s and has become a research hotspot over the last thirty years. The CBIR has been used to describe the process of retrieving similar images from a large collection based on the image content (color, shape, and texture). The essence of the traditional CBIR is associated with similar image retrieval. Several image retrieval techniques have been developed through years of technical innovation, such as content-based and object-based methods, as well as methods based on machine or deep learning. The classic CBIR is markedly different from object recognition as most images are similar, rather than containing the same objects. Thus, training and testing sets cannot be distinguished in the classic CBIR.

Feature extraction and representation is the most important issue in CBIR and has a close relationship with human perception, where color and edge cues are among the visual search components for perception stimuli and can express meaningful characteristics in images or scenes. Feature integration theory is a perceptual and attentional theory that explains how an individual combines pieces of observable information about an object to form a complete

perception [1]. The first stage is preattention, in which we focus on one distinguishing attribute of the object. The second stage is focused attention, in which we take all the observed edge features and combine them to make a complete perception [1]. Representing image contents via extracting the color and edge features based on feature integration theory is a challenging problem. To address this problem, a novel and highly simple but efficient representation based on the multi-integration features model, namely the multi-integration features histogram (MIFH), is proposed for CBIR.

The proposed model is based on similar image retrieval rather than object-based image retrieval or object classification. Our main contributions are summarized as follows. First, a new method is proposed to extract lower-level features, which simulates the preattention stage, where color differences are used as the novel distinguishing attribute of an object. Second, we take all of the observed edge features and combine them to make a complete perception using a set of simple integration processes to provide powerful information for image representation. Third, a new image representation is proposed for image retrieval that combines color, edge, and spatial layout into an integrational feature.

These attributes improve the discriminative power for color, texture, and shape features.

The rest of this paper is organized as follows. In Section 2, we review existing visual feature description methods and image retrieval techniques. In Section 3, the multi-integration features model and MIFH method are presented. We conduct CBIR experiments in Section 4, while Section 5 concludes the paper.

## 2. Related Works

Low-level features are highly popular in pattern recognition techniques. Color, texture, and shape features are widely used attributes in saliency detection [2] and traditional content-based image retrieval [3–10]. In the 2000s, ISO developed a multimedia content description standard (MPEG-7) to represent information regarding its content. The color, texture, and shape descriptors are the core components in this standard [3]. In the early stages of CBIR development, the use of a color histogram to describe image content has been a highly popular method in CBIR, since the color histogram is invariant to the orientation and scale. This characteristic makes it powerful for image classification and CBIR [3, 5–7].

Texture is also a highly popular visual feature for CBIR as it yields information regarding the spatial arrangement of the color or intensity in an image or region. The texture descriptors can be used to characterize image textures or regions, where the local binary pattern (LBP) is a famous texture descriptor [11]. In recent years, some LBP methods have been proposed for content-based image retrieval [12–14], such as the multichannel decoded LBP method [12] and the edge-local binary pattern (edgeLBP) approach [14]. In recent years, many algorithms have been proposed that combine multivisual cues [3–10, 12, 13] and have been utilized for various tasks. Moreover, simulating the mechanism of the human cortex to extract visual features provides a new avenue for CBIR. A set of Gabor filters with different frequencies and orientations can be used to extract useful visual features from an image, such as texture and edge information. This is similar to the perception of the human visual system and has been found to be particularly appropriate for texture representation and discrimination [15–18]. Liu et al. proposed the intensity variation descriptor [19] and the gradient-structures histogram [20] to represent image content and used them for image retrieval. The proposed representation methods have the advantages of being histogram-based and having the power to discriminate spatial layout, color and edge cues.

Shape features are widely used in object recognition and image retrieval because they contain important semantic information. Humans have the ability to recognize objects through shape; however, shape can only be extracted from a set of accurate segmentation methods, which is a complicated process. In many image processing cases, shape descriptors may also be required to be invariant to translations, rotations, and scale. To avoid the requirement of accurate segmentation, some feature

descriptors exhibit strength in object recognition, such as shape context [21], scale-invariant feature transform (SIFT) descriptors [22], histogram of oriented gradients [23], and SURF descriptors [24]. The histogram of oriented gradient (HOG) is a feature descriptor that is widely used for object detection. Object-based image retrieval has become increasingly popular since 2004 when Lowe proposed the SIFT descriptors. This type of image retrieval usually uses BOW methods to represent local features. In subsequent years, various BOW-based methods have been proposed for object-based image retrieval, object recognition, and other applications [25–28]. Ahmed et al. proposed a new approach for CBIR by fashioning the spatial color information with shape and object features [29]. Pradhan et al. proposed a new method for CBIR that uses color and shape features [30, 31]. In some cases, the representation methods using dimensionality reduction [32] and structured optimal graph [33] have the power to extract the shape feature.

In recent years, deep learning techniques represented by convolutional neural networks (CNNs) have been shown to be effective at various vision tasks [17, 18, 34–39]. In early works, the approach of applying CNNs for image retrieval established the fully connected layer activation functions to be the global image descriptors [17, 35–37]. Deep learning approaches represented by CNNs have demonstrated a superior performance over hand-crafted features. For instance, Qayyum et al. proposed a framework of deep learning for CBMIR systems using a deep CNN that was trained for classification of medical imagery [38]. Tzelepi and Tefas employed a deep CNN model to obtain feature representation from the activations of convolutional layers using max pooling for CBIR [18]. Alzu’bi et al. introduced compact bilinear CNN-based architectures for several CBIR tasks using two parallel feature extractors without prior knowledge of the semantic metadata of image content [39].

Deep learning techniques can provide better performance overall in image retrieval and object recognition but require long training times. These approaches also do not typically mimic the mechanism of the human brain or the perception mechanism. Extracting visual features and representing image contents that mimic the perception mechanism within the CBIR framework requires further study.

## 3. The Multi-Integration Features Model and Representation

In digital image processing, color feature extraction, and edge detection play an important role in feature extraction and includes high-level concepts [40, 41]. In applications of CBIR and object recognition, color can provide powerful information for feature extraction and representation [2, 9]; even in the absence of shape information, combining color features with other visual features, such as textures, edge cues, and spatial attributes, is a popular technique to improve the image retrieval performance [19, 20]. Color and

edge cues are visual search components for stimuli perception that can express meaningful characteristics of images or scenes. To represent the image content by extracting color and edge features based on feature integration theory, we propose a multi-integration features model to represent image content and use it for CBIR. The architecture overview of the process is shown in Figure 1.

There are three main components of the architecture shown in Figure 1. The first component is the preattention stage which is highlighted with the red solid line, in which we focus on one distinguishing attribute of the object using color differences. The second component is the focused attention, which is highlighted with the purple dotted line, in which we take all the observed edge features and combine them to make a complete perception. Three integrations of the edge features are used with a set of simple integration processes. For the third component, we propose a novel and highly simple but efficient visual descriptor to represent an image, which is used for CBIR. The  $E_1$ ,  $E_2$ , and  $E_3$  integrations, color map, and log transformation processes occur simultaneously in the histogram.

**3.1. Color Feature Extraction and the Initial Feature Integration.** The HSV color space can mimic the human color perception well; therefore, many researchers have used it for color quantization [6, 8, 10, 19, 20]. The HSV color space is based on the cylindrical coordinate system and is defined in terms of the three components of hue ( $H$ ), saturation ( $S$ ), and value ( $V$ ), where  $S$  is the distance perpendicular to the rod,  $V$  is the central rod, and  $H$  is the angle around the rod [19, 20, 40, 41]. In the proposed MIFH method, the color feature is extracted using color quantization.

To simplify the color quantization operation in the HSV color space, each of the  $H$ ,  $S$ , and  $V$  color channels are uniformly quantized into 6, 3, and 3 bins, respectively, which results in a total of  $6 \times 3 \times 3 = 54$  color combinations. Let  $M_0(x, y)$  be the color combination or the color map as  $M_0(x, y) = w$ ,  $w \in \{0, 1, \dots, N_C - 1\}$ , where  $w$  denotes the index value of the color map in the pixel coordinate  $(x, y)$ , and  $N_C = 54$ . The  $M_0(x, y)$  is used to calculate the color features in the proposed model.

To perform the initial feature integration, we focus on one distinguishing attribute of the object using the color difference, which we calculate in the HSV color space for the integration processes. The initial feature integration involves the characteristics of neurons that it receives a set of input signals and produces output, and the activation and excitation states of the neurons are simulated by using a threshold [42]. In order to simulate such characteristics, the sigmoid function  $\text{sigmoid}(\cdot)$  is used as the active function in the initial feature integration, and it can map the input into an output in the range of  $[0, 1]$ .

We then transform the cylindrical coordinate system to Cartesian coordinates. Let  $(H, S, V)$  be a random dot in cylindrical coordinates, and  $(h, s, v)$  is the transformation of  $(H, S, V)$  in Cartesian coordinates, where  $h = S \cdot \cos(H)$ ,  $s = S \cdot \sin(H)$ , and  $v = V$ . In the input image, the central coordinates are denoted as  $(x_0, y_0)$ , and the coordinates for

the eight neighboring pixels are denoted as  $(x_i, y_i)$ ,  $i \in [1, 2, 3, \dots, 8]$ . The color difference map is defined as follows:

$$\left\{ \begin{array}{l} C(x_0, y_0) = \text{sigmoid}\left(\sum_{i=1}^8 \sqrt{(\Delta H_i)^2 + (\Delta S_i)^2 + (\Delta V_i)^2} + b\right), \\ \Delta H_i = h(x_i, x_i) - h(x_0, y_0), \\ \Delta S_i = s(x_i, x_i) - s(x_0, y_0), \\ \Delta V_i = v(x_i, x_i) - v(x_0, y_0), \end{array} \right. \quad (1)$$

where  $\Delta H_i$ ,  $\Delta S_i$ , and  $\Delta V_i$  denote the color differences between the pixel locations  $(x_0, y_0)$  and  $(x_i, y_i)$  for the  $H$ ,  $S$ , and  $V$  components. In the active function  $\text{sigmoid}(\cdot)$ , the bias parameter is  $b = 1.0$ , and the value of the bias can measure how easy it is for a neuron to generate positive (negative) excitation. The color difference for the local regions contains rich information as it can measure the degree of color differences. Figure 2 shows that the edge and important regions can be simply detected using such operations. At the same time, the information for the three-color channels was also combined into a color difference map  $C(x, y)$ .

A subsampling operation is implemented in the color difference map  $C(x, y)$  with a scale of two, and the subsampling image is denoted as  $S(x, y)$ . The size of  $S(x, y)$  is one quarter that of  $C(x, y)$ .

**3.2. First Integration of the Edge Features.** When extracting edge features for  $S(x, y)$ , we adopt the horizontal and vertical Sobel kernels as the weight values between feature maps. The Sobel operator is less sensitive to noise than other gradient operators or edge detectors while also being relatively efficient [40, 41]. These kernels are defined as follows:

$$\begin{aligned} g_x &= \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, \\ g_y &= \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}. \end{aligned} \quad (2)$$

We obtain the edge map  $G_0(x, y)$  using the Sobel operation  $\mathcal{G}(\cdot)$  which is denoted as

$$G_0(x, y) = \mathcal{G}(S(x, y)). \quad (3)$$

To extract more edge features, we need to perform the same actions  $\mathcal{G}(\cdot)$  three times to obtain the remaining three edge maps again as

$$G_i(x, y) = \begin{cases} \mathcal{G}(\mathcal{K}_i(x, y)), & i \in [1, 2, 3], \\ \text{where } \mathcal{K}_i(x, y) = S(x, y) + \sum_{k=1}^i G_{k-1}(x, y). \end{cases} \quad (4)$$

We obtain the four edge feature maps and combine them into a gradient image  $E_1(x, y)$ , which can be calculated as

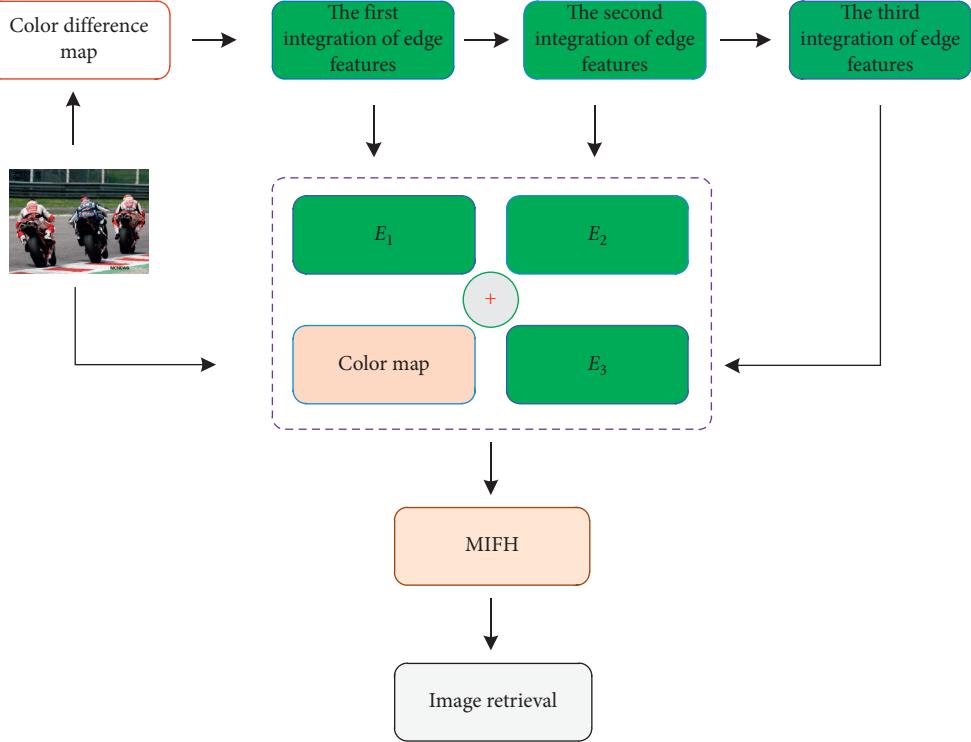
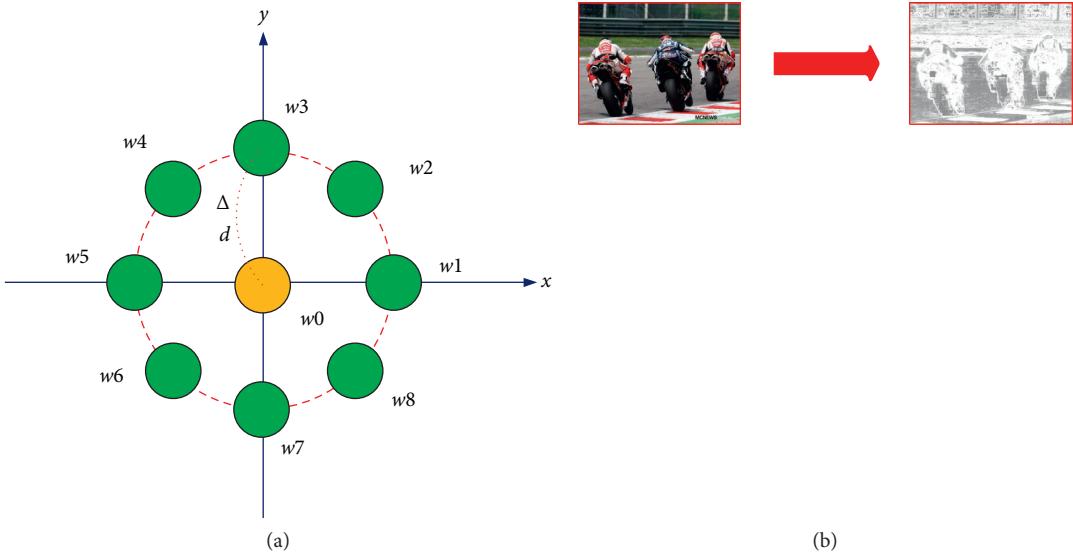


FIGURE 1: The architecture overview of the multi-integration features model.

FIGURE 2: Illustration for the computation of color differences. (a) The center point and its eight neighborhoods. (b) Color difference map derived from an input image (top right), where  $\Delta d$  denotes the distance between the center point and its neighborhood points. The coordinates of points  $w_i$  are denoted as  $(x_i, y_i)$ ,  $i = 0, 1, 2, \dots, 8$ .

$$E_1(x, y) = \sqrt{\sum_{i=0}^3 (G_i(x, y))^2}. \quad (5)$$

Then, the subsampling operation is implemented in  $G_i(x, y)$ ,  $i \in [0, 1, 2, 3]$  with a scale of two, and the subsampling feature maps are denoted as  $\mathcal{G}_i(x, y)$ ,

$i \in [0, 1, 2, 3]$ . The size of  $\mathcal{G}_i(x, y)$  is one quarter that of  $G_i(x, y)$ .

**3.3. Second Integration of the Edge Features.** For every feature map  $\mathcal{G}_i(x, y)$ ,  $i \in [0, 1, 2, 3]$ , the same actions  $\mathcal{G}(.)$  as the equations (3) and (4) need to be performed. As a result, a total of 16 edge maps  $G_{i,j}(x, y)$ ,  $i, j \in [0, 1, 2, 3]$  were

obtained. We first combine them into four gradient images  $F_i(x, y)$ ,  $i \in [0, 1, 2, 3]$  using (5), which can be calculated as

$$F_i(x, y) = \sqrt{\sum_{j=0}^3 (G_{i,j}(x, y))^2}, \quad i \in [0, 1, 2, 3]. \quad (6)$$

The four gradient images  $F_i(x, y)$ ,  $i \in [0, 1, 2, 3]$  are further combined into a single gradient image  $E_2(x, y)$ , which can be calculated as

$$E_2(x, y) = \sum_{i=0}^3 F_i(x, y). \quad (7)$$

Then, a subsampling operation is implemented for  $F_i(x, y)$ ,  $i \in [0, 1, 2, 3]$ , where the scale is two, and the subsampling feature maps are denoted as  $\mathcal{F}_i(x, y)$ ,  $i \in [0, 1, 2, 3]$ . The size of  $\mathcal{F}_i(x, y)$  is one quarter of  $F_i(x, y)$ .

**3.4. Third Integration of the Edge Features.** We combine the feature maps  $\mathcal{F}_i(x, y)$ ,  $i \in [0, 1, 2, 3]$  into a single gradient image  $E_3(x, y)$  with

$$E_3(x, y) = \sum_{i=0}^3 \mathcal{F}_i(x, y). \quad (8)$$

To represent the image content, we uniformly quantize the gradient images  $E_1(x, y)$ ,  $E_2(x, y)$ , and  $E_3(x, y)$  in different steps, where their index maps are defined as  $M_1(x, y)$ ,  $M_2(x, y)$ , and  $M_3(x, y)$ , respectively.

In the index maps  $M_k(x, y)$ ,  $k = 1, 2, 3$ , and their index values are  $M_k(x, y) = v$ ,  $v \in \{0, 1, \dots, N_v - 1\}$ . We set the number of bins to  $N_v = 16$ . In this study, the color index map  $M_0(x, y)$  and the edge index maps  $M_k(x, y)$ ,  $k = 1, 2, 3$  are used to calculate the visual features in the proposed model.

**3.5. Feature Representation.** The spatial layout information and histogram-based methods are widely used in CBIR and to describe visual features. Here, co-occurrence is adopted for image representation. Let  $(x, y)$  and  $(x + \Delta x, y + \Delta y)$  be two neighboring pixel locations in an image, where  $\Delta x$  and  $\Delta y$  denote the offsets of the  $x$ -axis and the  $y$ -axis, respectively. Two neighboring pixels have a distance of  $d = \max\{\Delta x, \Delta y\}$ . We use  $\mathcal{N}\{\cdot\}$  to denote the co-occurring number of the two-pixel values and then define the co-occurring histogram of the color map and three edge maps as follows:

$$H_k(M_k(x, y)) = \begin{cases} \mathcal{N}\{M_k(x, y) = v_1 \wedge M_k(x + \Delta x, y + \Delta y) = v_2\}, \\ \text{where } M_k(x, y) = M_k(x + \Delta x, y + \Delta y) = v_1 = v_2, \\ k = 0, 1, 2, 3. \end{cases} \quad (9)$$

Following the concatenation of  $H_k$ ,  $k \in [0, 1, 2, 3]$ , we obtain the histogram  $h_i$ ,  $i \in [0, 1, \dots, N_c + 3N_v - 1]$ . The subscript of the histogram is denoted as  $i$ ,  $i \in [0, 1, \dots, N_c + 3N_v - 1]$ . A log transformation is applied to

adjust the dynamic range in the histograms  $h_i$ . We then define the MIFH as follows:

$$H_i = \begin{cases} \log(h_i), & \log(h_i) > 0.0, \\ 0.0, & \text{else.} \end{cases} \quad (10)$$

In image processing, the log function is important as it compresses the dynamic range of images that have large variations in their pixel values [41]. The log transformation expands the values for lower-level pixels in an image while compressing the higher-level values, thus emphasizing the lower-level pixels. In the proposed method, the use of a log function can reduce the effects of errors as derived from feature quantization and contains the spatial information obtained using the co-occurring neighboring pixels. Moreover, the proposed method can discriminate color, texture, and shape features.

## 4. Experimental Results

In the experiments, we follow the traditional CBIR (similar image retrieval) rather than object-based or similar types of image retrieval. We compare the MIFH method with the color volume histograms [4], local binary pattern (LBP) [11], multi-texton histogram (MTH) [6], histogram of oriented gradients (HOG) [23], and BOW histogram [25]. Besides, we also compare the MIFH method with some extend methods which are derived from our previous works (the multi-texton histogram [6] and the micro-structure descriptor [8]), including CPV-THF [43], STH [44], and CMSD [45] methods. In the comparison experiment, the vector dimensionality for the MIFH, LBPH, BOW, MTH, and CVH methods are 102, 256, 1000, 82, and 104 bins, respectively. The vector dimensionality for the CPV-THF [43], STH [44], and CMSD [45] methods are 242, 172, and 88 bins, respectively. In the HOG descriptor, there are nine bins with a block size of three and a cell size of six. For a fair comparison, the MIFH, LBPH, BOW, MTH, and CVH methods adopt the  $L_1$  distance as the similarity measure or their original distance measures.

**4.1. Benchmark Datasets.** The Corel image dataset is the most commonly used dataset to test the performance of the CBIR. In Corel image dataset, every category in the Corel gallery contains 100 images, on which digital zoom can be used to obtain various image sizes. To measure the image retrieval performance, we select categories to build the Corel-10K and Corel-5K datasets and used them in the CBIR, where Corel-5K contains 50 categories and 5,000 images, and Corel-10K dataset contains 100 categories and 10,000 images which contain all images of Corel-5K dataset.

The third dataset is the GHIM-10K with 10,000 images. All images in this dataset are sourced from the web or a camera and were collected by *Guang-Hai Liu*. The database contains 20 categories, such as building, sunset, fish, flower, car, mountain, and tiger. In the GHIM-10K dataset, each category contains 500 images at  $400 \times 300$  or  $300 \times 400$  in JPEG format. Corel-10K, Corel-5K, and GHIM-10K datasets

can be downloaded from <http://www.ci.gxnu.edu.cn/cbir/Dataset.aspx>.

Three subsets, consisting of ten percent of each dataset were used as the query images. The performance is evaluated using the average results for each query in terms of precision and recall.

**4.2. Performance Measurements.** Selecting the metric for the performance evaluation is critical to compare CBIR experiments. In many CBIR cases [3–10], precision and recall metrics are popular ways to evaluate the performance of CBIR or other text retrieval techniques [46].

On the Corel datasets, each category contains 100 images and is denoted as  $k = 100$ , while  $k = 500$  for the GHIM-10K dataset. If there are  $n$  retrieved images and  $m$  images within the retrieved images are relevant to the query for the image retrieval, then the precision and recall metrics can be defined as follows [46]:

$$\text{Precision} = \frac{m}{n}, \quad (11)$$

$$\text{Recall} = \frac{m}{k}$$

We set  $n = 12$  in our CBIR system, i.e., there are 12 retrieved images in the image retrieval.

**4.3. Retrieval Performance.** Feature selection is a very important issue in image representation. For example, edges and colors are widely used features in content-based image retrieval. In image representation, low-level features are combined into the proposed algorithm where quantization is used to extract the features. Various feature quantizations are important factors in CBIR experiments. Feature quantization has a greater contribution to the performance of content-based image retrieval because the proposed MIFH method is a histogram-based method. Thus, the number of edge bins and color feature bins can directly influence the retrieval results.

**4.3.1. Investigation of Vector Dimensionality.** To investigate the performance of the MIFH method using different histogram bins, we set the bin numbers of the color feature to 54, 72, 108, and 192 bins, respectively. For the first, second, and third integrations, the bins of the edge maps are set to 16, 32, and 64 bins, respectively. As shown in Figure 3, in most cases, the CBIR precision reduces with an increasing number of color bins on the Corel-10K dataset.

The GHIM-10K dataset differs because in many cases where the color bins increase, the CBIR precision increases. An exception is when there are 32 and 64 edge bins, as the precision increases when more color bins are added. Moreover, varying the edge bins for the first, second, and third integrations does not clearly increase or decrease the precision of the MIFH method.

In Figure 3, the precision of the MIFH method varies from approximately 49% to 53% on the Corel-10K dataset, whereas the precision for the GHIM-10K dataset varies from

approximately 54% to 59%. The Corel-10K dataset is older and contains images with colors that are not as bright as the GHIM-10K dataset, which is relatively new. Color feature is an important factor in the proposed method, and the degree of color brightness affects the retrieval process for the proposed method. The precision change of the MIFH method does not present a clear trend for the two datasets. Thus, we reduce the vector dimensionality and set the number of color bins, the edge bins of the first integration, the edge bins of the second integration, and the edge bins of the third integration to 54, 16, 16, and 16 bins, respectively. In this case, the total vector dimensionality is  $54 + 16 + 16 + 16 = 102$  bins.

**4.3.2. Contribution of Edge and Color.** Different experiments are implemented using various combinations of visual features to investigate the contributions of edges and colors within the MIFH method, where the quantization levels of the colors and edges are 54 and 16 bins, respectively.

On the Corel-10K dataset, the precision and recall when only using the color feature are 46.48% and 5.58%, respectively. Using only the edge features of the first, second, and third integrations gives poor precision and recall. Compared with the edge features, the color features significantly increase the performance of the proposed method. If we combine these features, the performance significantly increases.

Figure 4 shows that the performances for both Corel-10K and GHIM-10K datasets when using color and edge features are better than using only the color features or edge features individually. The  $E_1$ ,  $E_2$ , and  $E_3$  denote the edge features of the first, second, third integration, respectively. With more integrations, using the edge map for only a single integration does not improve the performance.

The repetition of the Sobel operation when calculating the edge gradient is an important factor in our method. Table 1 shows that the precision of the MIFH method increases with increasing repetitions of the Sobel operation on the Corel-10K dataset. The GHIM-10K dataset differs; the repetitions of the Sobel operation have not clearly increased or decreased the precision of the MIFH method.

The best precision for the MIFH method is achieved on Corel-10K datasets when the Sobel operation is repeated four times. With more repeated Sobel operations, the number of calculations also increase; thus, we set the final repetition of the Sobel operation to four in the proposed method.

To investigate the performance of the MIFH method using different bias parameters in calculating the color difference map, sigmoid function sigmoid(.) was utilized as the active function. Various bias parameters can result in different color difference maps and may further lead to different retrieval performance. As can be seen from Table 2, the best precision for the MIFH method is achieved on the Corel-10K dataset when the bias parameter  $b = 1.0$ . On GHIM-10K dataset, the precisions of MIFH method do not have much difference between the bias parameter  $b = 1.0$  and  $b = 2.0$ . Thus, the bias parameter  $b = 1.0$  was utilized in calculating the color difference map.

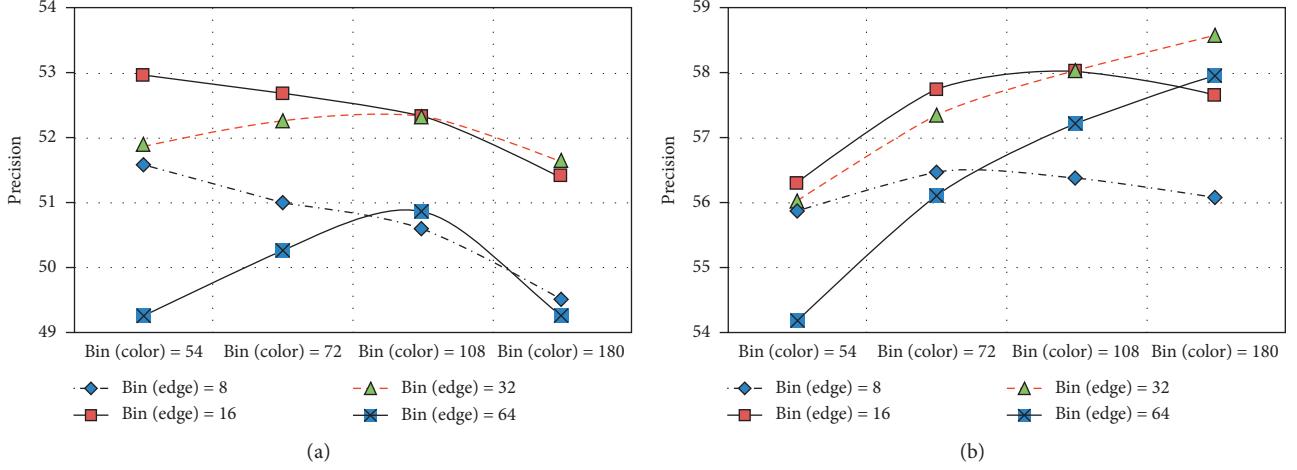


FIGURE 3: Performances of different bin numbers for the color and edge features on the (a) Corel-10K and (b) GHIM-10K datasets.

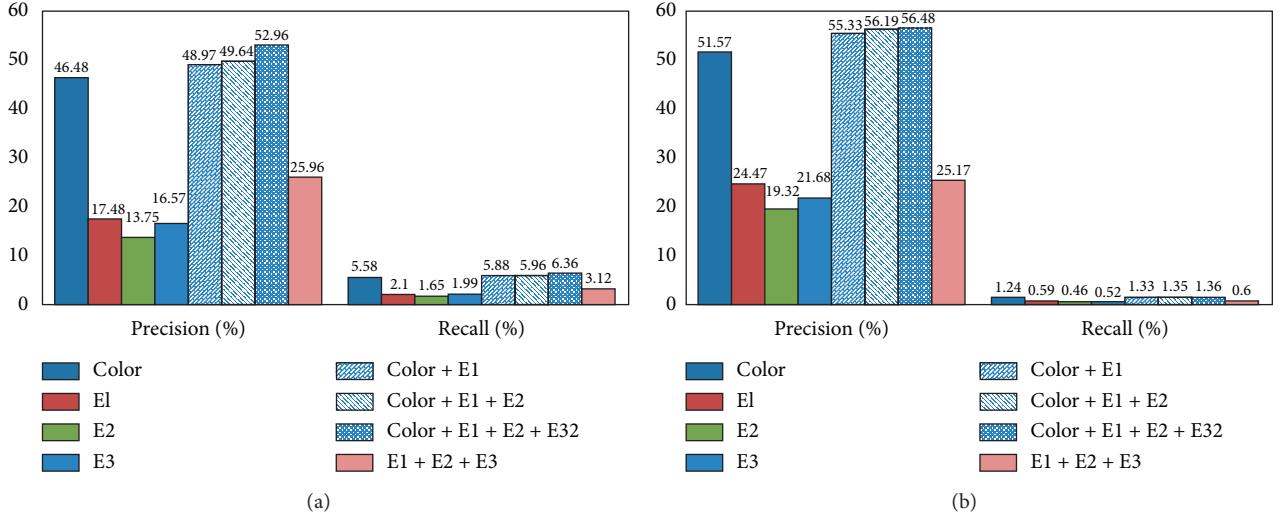
FIGURE 4: Contribution of the edge, color, and the number of integrations, where there are 54 color bins and 16 bins for each of the edge maps in the first, second, third integrations. The  $E_1$ ,  $E_2$ , and  $E_3$  denote the edge features of the first, second, and third integrations, respectively. The data are for the (a) Corel-10K and (b) GHIM-10K datasets.

TABLE 1: Repetitions of the Sobel operation when calculating the edge gradient.

Dataset	Performance (%)	Repetitions				
		1	2	3	4	5
Corel-10K	Precision	52.11	51.95	52.62	52.96	51.06
	Recall	6.25	6.23	6.31	6.36	6.13
GHIM-10K	Precision	57.71	57.10	56.23	56.48	56.64
	Recall	1.39	1.37	1.35	1.36	1.36

TABLE 2: The bias parameter in calculating the sigmoid function sigmoid (.)

Dataset	Performance (%)	The bias parameter				
		0.0	1.0	2.0	3.0	4.0
Corel-10K	Precision	48.92	52.96	51.82	51.33	50.58
	Recall	5.87	6.36	6.22	6.16	6.07
GHIM-10K	Precision	55.4	56.48	56.65	55.83	54.89
	Recall	1.33	1.36	1.37	1.34	1.32

**TABLE 3:** Performances of the compared methods on the Corel-10K, Corel-5K, and GHIM-10K datasets.

Dataset	Metric	Method								
		STH	CPV-THF	CMSD	BOW	LBPH	HOG	MTH	CVH	MIFH
Corel-10K	Precision	48.03	52.28	50.25	30.36	37.24	27.95	40.87	48.58	52.96
	Recall	5.76	6.27	6.03	3.64	4.47	3.35	4.91	5.83	6.36
GHIM-10K	Precision	—	—	—	39.67	46.89	23.96	53.13	57.48	56.48
	Recall	—	—	—	0.96	1.13	0.57	1.28	1.38	1.36
Corel-5K	Precision	60.28	63.90	63.14	—	43.82	35.75	49.98	60.13	60.16
	Recall	7.23	7.66	7.57	—	5.26	4.29	6.00	7.21	7.21

**4.3.3. Performance Comparisons.** In image representation, the color and edge features are extracted using the MIFH method, and it is natural that the MTH method can represent color and edge orientation features as well. Therefore, it stands to reason that both the MIFH and MTH methods can represent color texture features. The LBPH method is a well-known texture features descriptor that can also represent color texture features by extracting the LBP features along the R, G, and B color channels simultaneously. The HOG method is a well-known descriptor that is widely used in object recognition [23]. The technique counts the occurrences of gradient orientations in localized portions of an image. The image retrieval performances of the compared methods on the Corel-10K, Corel-5K, and GHIM-10K datasets are shown in Table 3.

The performances of the MIFH method are better than those of the BOW, MTH, HOG, and LBPH based on the experimental data. The precision of the MIFH method is higher than the STH, CPV-THF, CMSD, BOW, LBPH, HOG, MTH, and CVH methods at 4.93%, 0.68%, 2.71%, 22.60%, 15.72%, 25.01%, 12.09%, and 4.38%, respectively, on the Corel-10K dataset. On the GHIM-10K dataset, the precision of the MIFH method is higher than the BOW, LBPH, HOG, and MTH methods at 16.81%, 9.59%, 32.52%, and 3.35%, respectively. Nevertheless, the MIFH method is lower than the CVH method. On the Corel-5K dataset, the precision of the MIFH method is higher than the MTH, HOG, and LBPH methods. Considering the discrimination power of the color texture features, we confirm that the MIFH method is better than the LBPH method.

In order to determine the most suitable distance distances or similarity metrics, Canberra,  $x^2$  statistics,  $L_1$ , histogram intersection, and Chebyshev distances were adopted in the CBIR experiments. As can be seen from Table 4,  $L_1$  distance gives much better results than other metrics such as Canberra,  $x^2$  statistics, Chebyshev, Cosine, and histogram intersection. The histogram intersection gives the worst results on the two datasets.  $L_1$  distance is the simplest distance and more suitable for large scale datasets; thus, we were adopted it as the final distance in the proposed MIFH method.

## 5. Results and Discussion

Color and edge cues are among the visual search components for perception stimuli and can express meaningful characteristics in images or scenes. Combining the color and edge cues can help to describe image content. Many descriptors can be used to characterize the texture characteristics, where the local binary pattern (LBP) is a well-

known texture descriptor that can represent the local structures of image, but it cannot combine color and edge cues well. The MTH method is our prior work and was developed for the traditional CBIR, and it can represent the frequency for color and edge orientation information using a special histogram. CPV-THF, STH, and CMSD methods were derived from our previous works (the multi-texton histogram and the micro-structure descriptor), but those extend methods have not extracted the color and edge features based on feature integration theory.

The CVH method incorporates the advantages of histogram-based methods, as it takes the spatial information of neighboring pixels into account and has a performance that is similar with the MIFH method. It has the ability to discriminate color and edge features.

The BOW and HOG methods do not achieve satisfactory results on either dataset for the content-based image retrieval. The HOG method is a well-known descriptor that was developed for object recognition. Object recognition and content-based image retrieval are two different applications of pattern recognition. The task of object-based image retrieval is finding images that contain the same object or scene as the query image, whereas the classic CBIR is finding images that contain similar content as the query image. In some cases, using an object recognition-based method for content-based image retrieval cannot obtain satisfactory results compared with the application of object recognition.

The HOG method counts the occurrences of the gradient orientation to represent the image content. While the gradient orientation is a highly important cue in image representation, the discriminative power of only using the gradient orientation cue is a limiting factor in CBIR. In the BOW histogram, visual words obtained using the vector quantization of local features descriptors (e.g., Sift, SURF, ORB, and other local descriptors) require heavy computational burdens and may result in information loss.

Color is the most basic quality of visual content, so it is expected that the MIFH method has the ability to discriminate color features. In addition, edge cues are also contained in the MIFH method. This approach should have the ability to represent the general shape and texture features normally. To illustrate the representational ability of the visual features, we show two retrieval examples for the MIFH method in Figures 5(a) and 5(b). Both examples are used to show the visual effects of color and texture or other features, rather than showing whether the performances are good or not because not all images can return such good effects.

TABLE 4: The performance of the proposed method according to different distance or similarity metrics.

Dataset	Performance (%)	Canberra	$\chi^2$ statistics	Distance or similarity metrics			
				$L_1$	Histogram intersection	Cos	Chebyshev
Corel-10K	Precision	44.63	45.42	52.96	24.03	48.61	35.16
	Recall	5.35	5.45	6.36	2.88	5.83	4.22
GHIM-10K	Precision	50.33	51.4	56.48	31.01	54.87	39.98
	Recall	1.21	1.23	1.36	0.74	1.32	0.96



FIGURE 5: Two retrieval examples on the Corel-10K and GHIM-10K datasets using the MIFH method. The queries are for (a) flower and (b) motorcycle. In the two examples, the top-left image is the query image, and the similar images include the query image itself in the performance metric in terms of precision and recall.

In Figure 5, most of the returned images are correctly retrieved and ranked within the top 12 images, and all of the top retrieved images exhibit a good match for both the shape and color texture to the queries. In addition, texture is a highly popular visual feature for CBIR that yields information regarding the spatial arrangement of the color or intensity of an image or region.

## 6. Conclusions

Feature extraction and representation is an important issue in CBIR and has a close relationship with human perception. Color and edge cues are visual search components for stimuli perception that can express meaningful characteristics of images or scenes. Representing image contents via extracting color and edge features based on feature integration theory is a challenging problem. To address this problem, a novel and highly simple but efficient representation based on the multi-integration features model, called the MIFH, is proposed for CBIR using the feature integration theory.

The MIFH method provides powerful information in image representation using a set of simple integration processes. Such attributes improve the discriminative power of the MIFH method and enable a higher performance compared with the BOW, LBP, HOG, and MTH approaches.

## Data Availability

The dates and code are available at <http://www.ci.gxnu.edu.cn/cbir/Dataset>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61866005 and in part by the project of the Guangxi Natural Science Foundation of China under Grant 2018GXNSFAA138017.

## References

- [1] A. M. Treisman and G. Gelade, “A feature-integration theory of attention,” *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.
- [2] G.-H. Liu and J.-Y. Yang, “Exploiting color volume and color difference for salient region detection,” *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 6–16, 2019.
- [3] B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*, John Wiley & Sons, Hoboken, NJ, USA, 2002.
- [4] H. Ji-Zhao, G.-H. Liu, and S.-X. Song, “Content-based image retrieval using color volume histograms,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 9, Article ID 1940010, 2019.
- [5] G.-H. Liu and J.-Y. Yang, “Image retrieval based on the texton co-occurrence matrix,” *Pattern Recognition*, vol. 41, no. 12, pp. 3521–3527, 2008.
- [6] G.-H. Liu, L. Zhang, Y.-K. Hou, Z.-Y. Li, and J.-Y. Yang, “Image retrieval based on multi-texton histogram,” *Pattern Recognition*, vol. 43, no. 7, pp. 2380–2389, 2010.
- [7] J. Luo and D. Crandall, “Color object detection using spatial–color joint probability functions,” *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1443–1453, 2006.
- [8] G.-H. Liu, Z.-Y. Li, L. Zhang, and Y. Xu, “Image retrieval based on micro-structure descriptor,” *Pattern Recognition*, vol. 44, no. 9, pp. 2123–2133, 2011.
- [9] G.-H. Liu and J.-Y. Yang, “Content-based image retrieval using color difference histogram,” *Pattern Recognition*, vol. 46, no. 1, pp. 188–198, 2013.

- [10] G.-H. Liu, J.-Y. Yang, and Z. Li, "Content-based image retrieval using computational visual attention model," *Pattern Recognition*, vol. 48, no. 8, pp. 2554–2566, 2015.
- [11] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [12] S. R. Dubey, S. K. Singh, and R. K. Singh, "Multichannel decoded local binary patterns for content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4018–4032, 2016.
- [13] C. Singh, E. Walia, and K. P. Kaur, "Color texture description with novel local binary patterns for effective image retrieval," *Pattern Recognition*, vol. 76, pp. 50–68, 2018.
- [14] E. M. Thompson and S. Biasotti, "Description and retrieval of geometric patterns on surface meshes using an edge-based LBP approach," *Pattern Recognition*, vol. 82, pp. 1–15, 2018.
- [15] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America A*, vol. 2, no. 7, pp. 1160–1169, 1985.
- [16] G. Azzopardi and N. Petkov, "A CORF computational model of a simple cell that relies on LGN input outperforms the gabor function model," *Biological Cybernetics*, vol. 106, no. 3, pp. 177–189, 2012.
- [17] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval," in *European Conference on Computer Vision ECCV 2014*, Springer, Cham, Switzerland, 2014.
- [18] M. Tzelepi and A. Tefas, "Deep convolutional learning for content based image retrieval," *Neurocomputing*, vol. 275, pp. 2467–2478, 2018.
- [19] Z. Wei and G.-H. Liu, "Image retrieval using the intensity variation descriptor," *Mathematical Problems in Engineering*, vol. 2020, Article ID 6283987, 12 pages, 2020.
- [20] B.-H. Yuan and G.-H. Liu, "Image retrieval based on gradient-structures histogram," *Neural Computing and Applications*, 2020.
- [21] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–521, 2002.
- [22] D. G. Lowe, "Distinctive image features from scale-invariant key points," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [23] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp. 886–893, IEEE, San Diego, CA, USA, June 2005.
- [24] H. Bay, T.uytelaars, and L. V. Gool, "SURF: speeded up robust features," in *European Conference on Computer Vision*, vol. 1, pp. 404–417, Springer, Berlin, Germany, 2006.
- [25] J. Sivic and A. Zisserman, "Efficient visual search of videos cast as text retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 591–606, 2009.
- [26] Y. Su and F. Jurie, "Improving image classification using semantic attributes," *International Journal of Computer Vision*, vol. 100, no. 1, pp. 59–77, 2012.
- [27] L. Wang, L. Zhou, C. Shen, L. Liu, and H. Liu, "A convolutional word-merging algorithm with class separability measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 417–435, 2014.
- [28] H. Lobel, R. Vidal, and A. Soto, "Learning shared, discriminative, and compact representations for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 11, pp. 2218–2231, 2015.
- [29] K. T. Ahmed, S. Ummesafi, and A. Iqbal, "Content based image retrieval using image features information fusion," *Information Fusion*, vol. 51, no. 1, pp. 76–99, 2019.
- [30] J. Pradhan, A. K. Pal, and H. Banka, "Principal texture direction based block level image reordering and use of color edge features for application of object based image retrieval," *Multimedia Tools and Applications*, vol. 78, no. 2, pp. 1685–1717, 2019.
- [31] J. Pradhan, S. Kumar, A. K. Pal, and H. Banka, "A hierarchical CBIR framework using adaptive tetrolet transform and novel histograms from color and shape features," *Digital Signal Processing*, vol. 82, no. 1, pp. 258–281, 2018.
- [32] Z. Liu, J. Wang, G. Liu, and L. Zhang, "Discriminative low-rank preserving projection for dimensionality reduction," *Applied Soft Computing*, vol. 85, Article ID 105768, 2019.
- [33] Z. Liu, Z. Lai, W. Ou, K. Zhang, and R. Zheng, "Structured optimal graph based sparse feature extraction for semi-supervised learning," *Signal Processing*, vol. 170, Article ID 107456, 2020.
- [34] F. Radenovic, G. Tolias, and O. Chum, "Fine-tuning CNN image retrieval with no human annotation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1655–1668, 2019.
- [35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS 2012)*, NeurIPS, Lake Tahoe, NH, USA, December 2012.
- [36] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 1725–1732, IEEE, Columbus, OH, USA, June 2014.
- [37] F. Zhao, Y. Huang, L. Wang, and T. Tan, "Deep semantic ranking based hashing for multi-label image retrieval," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 1556–1564, IEEE, Boston, MA, USA, June 2015.
- [38] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, "Medical image retrieval using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 8–20, 2017.
- [39] A. Alzu'bi, A. Amira, and N. Ramzan, "Content-based image retrieval with compact deep convolutional features," *Neurocomputing*, vol. 249, pp. 95–105, 2017.
- [40] W. Burger and M. J. Burge, *Principles of Digital Image Processing: Core Algorithms*, Undergraduate Topics in Computer Science, Springer, London, UK, 2009.
- [41] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice-Hall, New Jersey, NJ, USA, 3rd edition, 2007.
- [42] G. Montavon, G. B. Orr, and K.-R. Müller, *Neural Networks: Tricks of the Trade*, Lecture Notes in Computer Science, Springer, Berlin, Germany, 2nd edition, 2012.
- [43] 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.
- [44] A. Raza, T. Nawaz, H. Dawood, and H. Dawood, "Square texton histogram features for image retrieval," *Multimedia Tools and Applications*, vol. 78, no. 3, pp. 2719–2746, 2019.
- [45] H. Dawood, M. H. Alkinani, A. Raza, H. Dawood, R. Mehboob, and S. Shabbir, "Correlated microstructure descriptor for image retrieval," *IEEE Access*, vol. 7, pp. 55206–55228, 2019.
- [46] C. J. van Rijsbergen, *Information Retrieval*, Butterworths, London, UK, 2nd edition, 1979.