

## Research Article

# Image Retrieval Using the Intensity Variation Descriptor

Zhao Wei 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](mailto:liuguanghai009@163.com)

Received 11 July 2019; Accepted 11 November 2019; Published 9 January 2020

Academic Editor: Vittorio Bianco

Copyright © 2020 Zhao Wei 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.

Variations between image pixel characteristics contain a wealth of information. Extraction of such cues can be used to describe image content. In this paper, we propose a novel descriptor, called the intensity variation descriptor (IVD), to represent variations in colour, edges, and intensity and apply it to image retrieval. The highlights of the proposed method are as follows. (1) The IVD combines the advantages of the HSV and RGB colour spaces. (2) It can simulate the lateral inhibition mechanism and orientation-selective mechanism to determine an optimal direction and spatial layout. (3) An extended weighted L1 distance metric is proposed to calculate the similarity of images. It does not require complex operations such as square or square root and leads to good performance. Comparative experiments on two Corel datasets containing 15,000 images show that the proposed method performs better than the SoC-GMM, CPV-THF, and STH methods and provides good matching of texture, colour, and shape.

## 1. Introduction

With the increasing number of images uploaded to the Internet, there is an urgent need to quickly and efficiently retrieve images from large-scale collections. This situation has triggered researchers to propose a variety of content-based image retrieval (CBIR) methods. Most of these methods first extract primary visual features such as colours, edges, and intensities. They then design feature models such as attention models or local pattern models to extract useful information, which is used to construct a feature vector histogram for image matching.

In recent years, use of the computational visual attention model to extract visual features has become a popular method for image retrieval. Liu et al. proposed a bar-shaped structure for image content analysis that combines the visual attention and orientation-selective mechanisms [1]. The bar-shaped structures emphasise the salient structural information in the primary visual features, which significantly improves the image retrieval performance. However, it only takes into account the associations between the bar-shaped structures, whose three pixels are invariant in intensity. The multitrend structure descriptor defines three trends in local structure [2], which comprise the increasing, decreasing, and invariant trends of the

three pixels according to the base of the microstructure descriptor [3]. The dLBP method [4] encodes the intensity variation of the three pixels into two bits, providing a total of eight kinds of variation. These methods have demonstrated excellent performance in image retrieval and object detection, but do not efficiently extract information on variations in colour, edges, and intensity. In order to efficiently extract these variations, we propose an intensity variation descriptor (IVD) based on dLBP [4], which can effectively combine the advantages of the HSV and RGB colour spaces and has the power to discriminate texture and spatial structure.

The proposed method has the following highlights. (1) The IVD can combine the advantages of the HSV and RGB colour spaces, which is more efficient than using a single colour space. (2) It can simulate the lateral inhibition mechanism and orientation-selective mechanism to determine the optimal direction and spatial layout. (3) We propose an extended weighted L1 distance, called W1 distance, to calculate the similarity between images and provide good CBIR performance.

The rest of this paper is organised as follows. In Section 2, we briefly review related works. The proposed method is presented in Section 3. We conduct CBIR experiments in Section 4, while Section 5 concludes the paper.

## 2. Related Works

In the field of image retrieval, various methods have been proposed to extract image features. These can be divided into two categories according to the type of features: global feature methods and local feature methods. Global feature methods extract the colour, texture, shape, and other features as essential visual features. It is simple to calculate global features with low feature dimensions. The local features are extracted from the positions of the image that are not easy to change. Therefore, it is robust to occlusion, changes in illumination and background, and geometric transformation [5, 6]. However, local feature methods use algorithms that are usually too complex and produce high-dimensional feature vectors.

The colour histogram is one of the most commonly used global feature extraction methods. It is robust to scale, rotation, and noise. However, it lacks a spatial layout and completely different images may have the same histogram distributions. Some methods have been proposed to address such problems, such as colour correlograms [7] and colour coherence vectors [8]. Texture is an image feature with a certain spatial structure that appears repeatedly. There are some well-known texture description methods, such as GLCM [9], LBP [10], Gabor [11], and wavelet [12]. Shape is also one of the most important features used to describe an image, and human beings can directly judge a category by its shape. Classical shape descriptors include Zernike moments [13], curvature scale space (CSS), and angular radial transform [14]. Among local feature methods, SIFT [15] is one of the most classical. It can achieve scale and rotation invariance by extracting certain points with invariant features in different scale spaces of an image. Due to the complexity of the algorithm and the high feature dimensions, some improved methods have been proposed, including SURF [16] and PCA-SIFT [17]. In addition, there are other local feature methods such as that of Harris and Stephens [18] and oriented FAST and rotated BRIEF (ORB) [19]. The BOW model [20] uses SIFT to extract the local features and then clusters them into visual words via a clustering algorithm. Finally, a dictionary composed of these words retrieves the corresponding results through similarity matching.

In recent years, the use of a single feature for image retrieval has failed to meet the requirements of large-scale image datasets. Accordingly, researchers began to propose image retrieval methods based on multifeature fusion. The most common fusion method is the combination of colour and texture. For example, a microstructure descriptor represents image local features by computing edge orientation similarities and underlying colours [3]. The structure element descriptor (SED) [21] combines colour and texture features using five structure elements denoting five directions. Liu and Yang proposed the colour difference histogram to integrate colour and edge orientation information based on perceptually uniform colour differences and used it for image retrieval [22]. Recently, many image retrieval methods based on LBP variants have been proposed [23–26]. Discrete wavelet transform, LBP, and grey level co-occurrence matrixes

can be used to exploit multiresolution analysis and to enhance image directional information [26–30]. Simulation of human perception and visual attention have been adopted in some multistage image retrieval frameworks [31–41]. The manifold ranking method has been used to match similar feature vectors [42, 43]. The WAS method [44] was proposed to integrate the information of adjacent images and compute their similarity. The SFW graph [35] is an efficient graph-matching method used to solve the pairwise image matching problem and reduce computational complexity.

Recently, deep learning has also been developed in the field of image retrieval, especially with image retrieval methods based on the convolutional neural networks (CNN). CNN-based methods use a pretrained or fine-tuned convolutional neural network to extract global features from the fully connected layer or local features from the intermediate layer for image retrieval [45–48]. Deep learning methods can achieve excellent retrieval results, but achieving reasonable results requires a lot of effort, including (1) knowledge and experience of network architecture, (2) huge amounts of training data, and (3) much computer time.

In this paper, we propose a simple yet efficient image retrieval method. Our method directly utilises low-level visual features to represent images and does not require complex operations such as modelling, training, clustering, and segmentation.

## 3. The Proposed Method

Color, intensity, and edges are considered as the primary visual features which are commonly used in CBIR. Here, we propose a novel descriptor, called the intensity variation descriptor (IVD), to represent variations in colour, edges, and intensity, and apply it to image retrieval. The flow diagram of the proposed method is shown in Figure 1. The IVD is built based on the intensity variations using colour, edges, and intensity in a certain direction. With the intensity variation serving as a bridge, the IVD extracts feature by simulating the lateral inhibition mechanism and orientation-selective mechanism and it effectively integrates colour, edges, and intensity for image retrieval.

In the proposed method, firstly, the input image is converted to the HSV colour space from the RGB colour space, and the H, S, and V components are uniform quantized. Then, the R, G, and B components of each pixel in the RGB colour space are taken as inputs for the IVD, and the output intensity variations are used as the feature values of the corresponding HSV-quantized colours. In addition, the edges and intensity information are extracted from the V component and used for quantization. The IVD is used to describe the intensity variation of the quantized edges and intensities in a certain direction, which is selected by the proposed local direction detection method. Finally, the intensity variation values of colours, edges, and intensities are integrated into the feature vector. After similarity matching by W1 distance, the 12 images look like the query image most are returned.

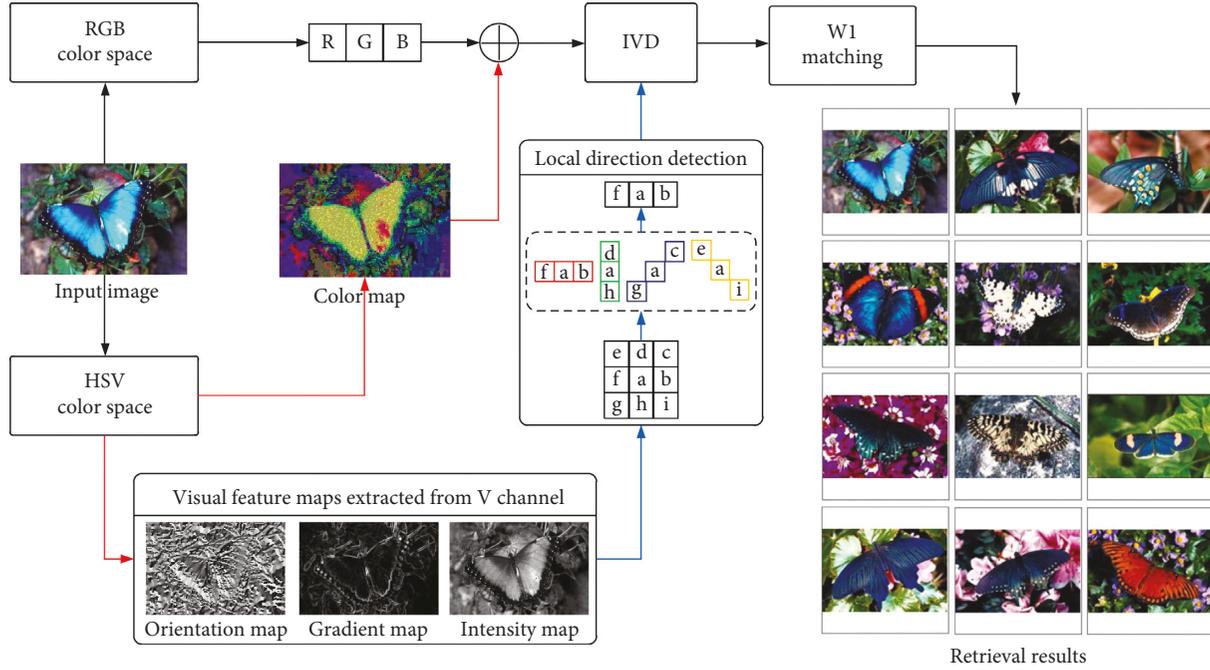


FIGURE 1: Flow diagram of the proposed method.

**3.1. The Intensity Variation Descriptor (IVD).** As mentioned before, there is a variety of pixel intensity variations. The bar-shaped structure considers the intensity invariant on three pixels [1]. The multitrend structure descriptor adds two situations, where the intensities of three pixels increase and decrease in turn from left to right based on the bar-shaped structure [1] and microstructure [3]. The LBP combines the situations of the invariant and increases into one situation, so there are four variations in LBP, as shown in Figure 2(c). The dLBP considers the intensity difference variation (IDV) based on LBP, as shown in Figure 2(d). Therefore, dLBP contains eight variations which come from four intensity variations (IVs) and two IDVs.

In order to efficiently extract more intensity variation information, an intensity variation descriptor (IVD) is proposed to extract intensity variation and intensity difference variation information based on the bar-shaped structure [1] and dLBP [4], as shown in Figure 2(e).

Let there be three intensity values,  $I_1$ ,  $I_2$ , and  $I_3$ . We define the intensity difference between them as follows:

$$\begin{aligned} d_1 &= I_1 - I_2, \\ d_2 &= I_2 - I_3. \end{aligned} \quad (1)$$

The weight of the intensity variation (IV) can be defined as follows:

$$I_v = 2^{3 \times s(d_1) + s(d_2)}, \quad (2)$$

$$s(x) = \begin{cases} 0 & x > 0 \\ 1 & x < 0 \\ 2 & x = 0. \end{cases}$$

The weight of the intensity difference variation (IDV) can be denoted as follows:

$$I_{dv} = 2^{2-s(d_1-d_2)}. \quad (3)$$

After combining the weights of the IV and IDV, we define the intensity variation value as follows:

$$IVD(I_1, I_2, I_3) = I_v \times I_{dv}. \quad (4)$$

**3.2. Extraction of Visual Features.** Human eyes are sensitive to colour and edges. Colour can provide rich information and is the most direct visual feature of images. Edges can represent the boundaries of image contents and textural structures. Therefore, colours and edges are utilised to represent image features in the IVD method.

Both the HSV and Lab colour spaces imitate human colour perception well [22, 38]. The Lab colour space is popular in calculating colour differences for image representation [22] and saliency detection [38], while the HSV colour space is more suitable for extracting colour information based on colour quantization [1, 3, 49]. Therefore, we adopted the HSV colour space for extracting colour features with the proposed method, where the H, S, and V components are uniformly quantized into 6, 3, and 3 bins, respectively. This results in a total of  $6 \times 3 \times 3 = 54$  colour bins. We define the colour bins as a colour map  $C(x, y)$ , where  $C(x, y) = s$ ,  $s \in \{0, 1, \dots, N_C - 1\}$ , and  $N_C = 54$ .

Sobel operators have good noise-suppression characteristics [50] and are simple to calculate. We used them to detect edge cues on the V component. We can obtain an edge amplitude map  $A(x, y)$  and edge orientation map  $O(x, y)$  by using uniform quantization,  $A(x, y) = t$ ,  $t \in \{0, 1, \dots, N_A - 1\}$ , where  $N_A = 9$ , and  $O(x, y) = u$ ,  $u \in \{0, 1, \dots, N_O - 1\}$ , where  $N_O = 36$ .

The V component is also utilised to represent intensity. An intensity map  $V(x, y)$  can be obtained using the above operation. We defined it as  $V(x, y) = v$ ,  $v \in \{0, 1, \dots, N_V - 1\}$ , where  $N_V = 16$ .

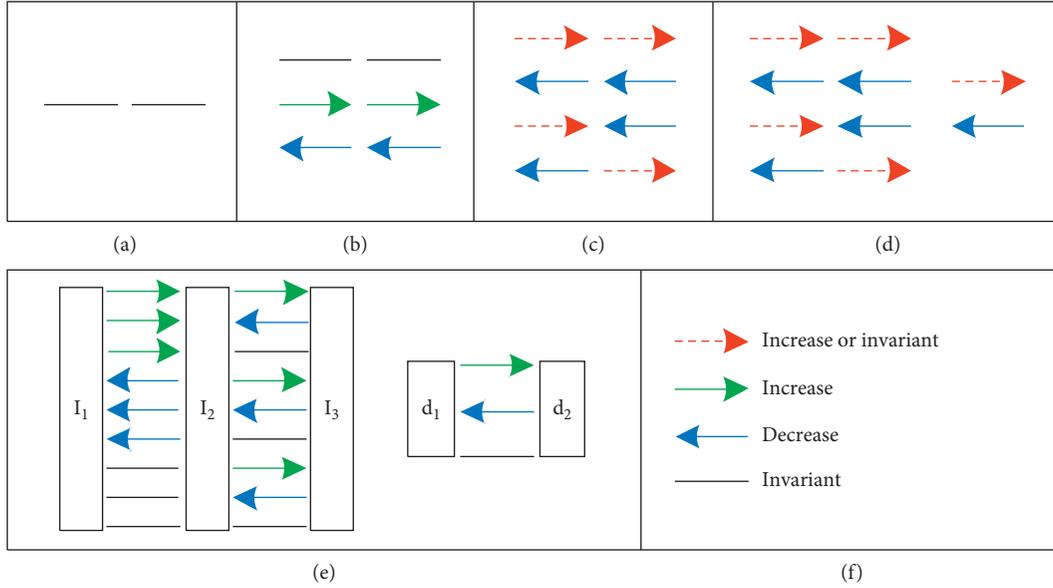


FIGURE 2: Intensity variation and intensity difference variation. (a) Bar-shaped structure, (b) multitrend structure, (c) LBP, (d) dLBP, (e) the proposed IVD, and (f) legend.

**3.3. Local Direction Detection.** In the IVD method, three intensity values are input to obtain the intensity variation. As shown in Figure 3, in a  $3 \times 3$  neighbourhood, the intensity values of three pixels in four directions can meet the requirements of IVD. Unfortunately, the calculation of IVD in four directions is too computationally intensive. Hubel and Wiesel showed that many simple cells in the visual cortex have an orientation-selective mechanism [51, 52], which makes them respond only to lines with a certain orientation [53]. Inspired by this view, we try to select three intensity values in a certain direction as the input of IVD, whose direction is the texture direction of the neighbourhood.

Let an  $m \times n$  region in an image and its grey value be  $g(x, y)$ . The grey values of pixels adjacent to  $g(x, y)$  in the  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$  directions are denoted as  $g(x+1, y)$ ,  $g(x+1, y-1)$ ,  $g(x, y+1)$ , and  $g(x+1, y+1)$ , respectively. Therefore, the average grey value difference of the local neighbourhood is denoted as follows:

$$G(\alpha) = \begin{cases} \sum_{x=0}^{m-2} \sum_{y=0}^{n-1} \frac{|g(x, y) - g(x+1, y)|}{(m-1) \times n}, & \alpha = 0^\circ, \\ \sum_{x=0}^{m-2} \sum_{y=1}^{n-1} \frac{|g(x, y) - g(x+1, y-1)|}{(m-1) \times (n-1)}, & \alpha = 45^\circ, \\ \sum_{x=0}^{m-1} \sum_{y=0}^{n-2} \frac{|g(x, y) - g(x, y+1)|}{m \times (n-1)}, & \alpha = 90^\circ, \\ \sum_{x=0}^{m-2} \sum_{y=0}^{n-2} \frac{|g(x, y) - g(x+1, y+1)|}{(m-1) \times (n-1)}, & \alpha = 135^\circ, \end{cases} \quad (5)$$

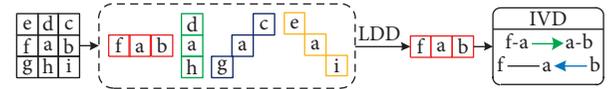


FIGURE 3: Example of IVD with a  $3 \times 3$  neighbourhood.

where  $m$  and  $n$  are the width and height of the region, respectively.  $(m-1) \times n$ ,  $(m-1) \times (n-1)$ ,  $m \times (n-1)$ , and  $(m-1) \times (n-1)$  are the numbers of pixel pairs in four directions, respectively. In this paper, we set  $m = n = 3$ . In a certain direction, we consider  $G(\alpha)$  as a numerator and the direction perpendicular to  $G(\alpha)$  as a denominator. We define the ratio  $f(\alpha)$  between them as follows:

$$f(\alpha) = \begin{cases} \frac{G(0^\circ)}{[G(90^\circ) + 1]}, & \alpha = 0^\circ, \\ \frac{G(45^\circ)}{[G(135^\circ) + 1]}, & \alpha = 45^\circ, \\ \frac{G(90^\circ)}{[G(0^\circ) + 1]}, & \alpha = 90^\circ, \\ \frac{G(135^\circ)}{[G(45^\circ) + 1]}, & \alpha = 135^\circ, \end{cases} \quad (6)$$

where adding one to the denominator prevents it becoming a zero value. The direction corresponding to the minimum of  $f(\alpha)$  is considered the direction of the  $3 \times 3$  region.

Using the above implementation, we can determine the regional direction of the edge amplitude map  $A$ , the edge orientation map  $O$ , and the intensity map  $V$ , where the intensity values of the three pixels in this direction are utilised as the input for the IVD.

**3.4. Image Representation.** In order to represent colour features efficiently, both HSV and RGB colour spaces are utilised to extract colour information. The pixels of the  $R$ ,  $G$ , and  $B$  channels are utilised as the input values of the IVD, and the output of the IVD is utilised as the histogram values. The colour features are represented as follows:

$$H_C(C(x, y)) = \sum \sum_{i=1}^8 \text{IVD}(R(x, y), G(x, y), B(x, y)). \quad (7)$$

In the representation of edges and intensity features, we determine the  $\alpha$  direction of a  $3 \times 3$  region within the edge and intensity feature map, and then use the three intensity values in the  $\alpha$  direction as the input of the IVD. Furthermore, we use the lateral inhibition formula to combine the spatial information output by the IVD [54].

$$R = \text{IVD}_0^\alpha + \sum_{i=1}^8 \gamma \times \text{IVD}_i^\alpha, \quad (8)$$

where  $\gamma$  is the lateral inhibition coefficient,  $\text{IVD}_0^\alpha$  is the output value of the IVD at the centre pixel, and  $\text{IVD}_i^\alpha$ ,  $i \in [1, 2, \dots, 8]$  is the output value of the IVD about eight pixels adjacent to the central pixel. We define  $\gamma$  as follows:

$$\gamma = \begin{cases} 1, & \text{if } \text{IVD}_0 = \text{IVD}_i, \\ 0.1, & \text{others.} \end{cases} \quad (9)$$

The histogram of the edge amplitude  $A$  is defined as follows:

$$H_A(A(x, y)) = \sum \sum_{i=1}^8 \frac{1}{9} \left( \text{IVD}_0^\alpha(A_0^{\alpha-1}, A_0^\alpha, A_0^{\alpha+1}) + \sum_{i=1}^8 \gamma \times \text{IVD}_i^\alpha(A_i^{\alpha-1}, A_i^\alpha, A_i^{\alpha+1}) \right), \quad (10)$$

where  $A_0^\alpha$  is the edge amplitude value of the centre pixel of the  $3 \times 3$  region,  $A_0^{\alpha-1}$  and  $A_0^{\alpha+1}$  are the surrounding edge amplitude values of  $A_0^\alpha$  in the  $\alpha$  direction,  $A_i^\alpha$  are the edge amplitude values of the eight pixels adjacent to the central pixel  $A_0^\alpha$ , and  $A_i^{\alpha-1}$  and  $A_i^{\alpha+1}$  are the surrounding edge amplitude values of  $A_i^\alpha$  in direction  $\alpha$ .

Similarly, the histograms of edge orientation  $O$  and intensity map  $V$  are defined as follows:

$$\begin{cases} H_O(O(x, y)) = \sum \sum_{i=1}^8 \frac{1}{9} \left( \text{IVD}_0^\alpha(O_0^{\alpha-1}, O_0^\alpha, O_0^{\alpha+1}) + \sum_{i=1}^8 \gamma \times \text{IVD}_i^\alpha(O_i^{\alpha-1}, O_i^\alpha, O_i^{\alpha+1}) \right), \\ H_V(V(x, y)) = \sum \sum_{i=1}^8 \frac{1}{9} \left( \text{IVD}_0^\alpha(V_0^{\alpha-1}, V_0^\alpha, V_0^{\alpha+1}) + \sum_{i=1}^8 \gamma \times \text{IVD}_i^\alpha(V_i^{\alpha-1}, V_i^\alpha, V_i^{\alpha+1}) \right). \end{cases} \quad (11)$$

Combining  $H_C(C(x, y))$ ,  $H_A(A(x, y))$ ,  $H_O(O(x, y))$ , and  $H_V(V(x, y))$ , the final histograms  $H$  can be obtained:

$$H = [H_C(0), H_C(1), \dots, H_C(N_C - 1), H_A(0), H_A(1), \dots, H_A(N_A - 1), H_O(0), H_O(1), \dots, H_O(N_O - 1), H_V(0), H_V(1), \dots, H_V(N_V - 1)]. \quad (12)$$

## 4. Experimental Results

In order to validate the performance of the proposed method, some state-of-the-art methods were selected for comparison: SoC-GMM [55], LDP [56], SSH [1], CPV-THF [57], and STH [58]. A distance metric and benchmark dataset were required for the comparisons. In these experiments, we propose using an improved distance to evaluate performance based on the average results of each query in terms of precision, recall, and F-measure, respectively.

**4.1. Datasets.** Two datasets were used for CBIR. (1) One was the Corel-5K dataset, which contains 5000 images with diverse content such as bark, food textures, waves, microscopic objects, and trees. It contains 50 categories, each with 100 images sized  $192 \times 128$  or  $128 \times 192$  pixels in JPEG format. (2) The second was the Corel-10K dataset containing 10,000 images of the same size as those in the Corel-5K dataset. It has various contents such as flowers, horses, fish, beaches, and mountains. It contains 100 categories, with 100 images in each category. The Corel-5K dataset is a subset of the Corel-10K dataset.

**4.2. Distance Metric.** In the CBIR experiments, we propose an improved distance formula based on CDH [22], namely, the W1 distance, to implement image matching. It can be considered as an extended weighted L1 distance. Let  $T$  and  $Q$  be the feature vectors of template and query images, respectively. Both are  $K$ -dimensional feature vectors:  $T = \{T_1, T_2, \dots, T_K\}$  and  $Q = \{Q_1, Q_2, \dots, Q_K\}$ . The W1 distance between the template and query images is simply calculated as follows:

$$D(T, Q) = \sum_{i=1}^K \frac{|T_i - Q_i|}{|T_i + Q_i + w \times u_T|}, \quad (13)$$

where  $u_T = \sum_{i=1}^K T_i / K$ . The W1 distance has the best performance when  $w = 0.01$ .

**4.3. Performance Measures.** In comparative CBIR experiments, deciding what kind of performance evaluation metric to use is important. In this paper, the precision, recall, and F-measure metrics were utilised to evaluate the effectiveness of the proposed method. They are defined as follows:

$$\begin{cases} P = \frac{I_N}{N}, \\ R = \frac{I_N}{M}, \\ F = \frac{(1 + \beta^2) \times R \times P}{(\beta^2 \times P) + R}, \end{cases} \quad (14)$$

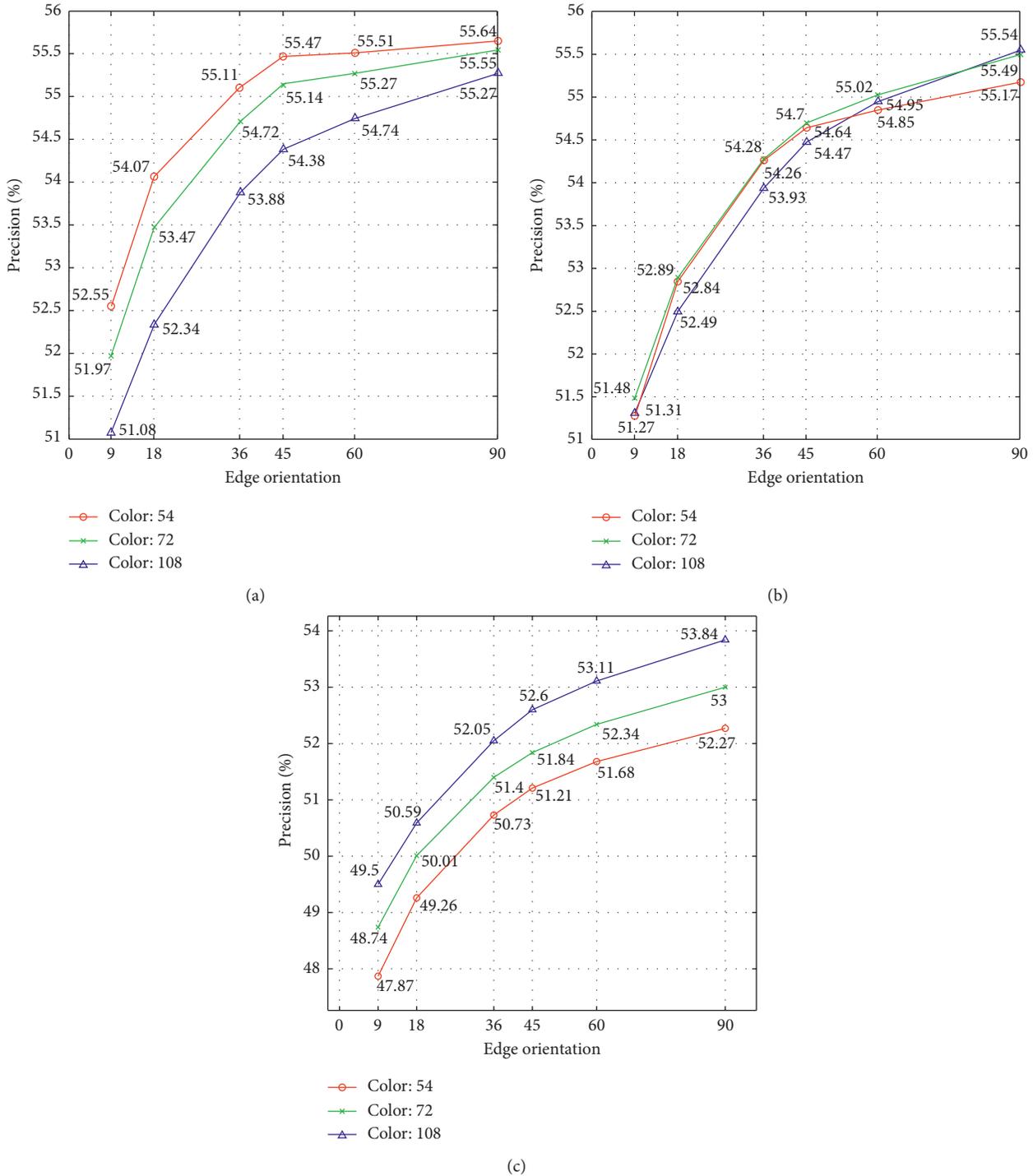


FIGURE 4: Precision using various quantization numbers of colour, edge orientation, and intensity on the Corel-10K dataset: (a) the quantization number of greyscale is 16 bins, (b) the quantization number of greyscale is 32 bins, and (c) the quantization number of greyscale is 64 bins.

where  $I_N$  is the number of images retrieved in the top  $N$  positions that are similar to the query image,  $N$  is the total number of images retrieved, and  $M$  is the total number of images in the database that are similar to the query. Parameter  $\beta$  allows one to weight either precision or recall more heavily.

In the Corel-5K and Corel-10K datasets, we set  $N = 12$ ,  $M = 100$ , and  $\beta = 1$ .

**4.4. Retrieval Performance.** The vector dimensionality has a very important impact on performance. In general, with the increase of dimensions, the retrieval performance of an algorithm will improve. However, exorbitant dimensions increase the computational burden. Therefore, choosing an appropriate vector dimensionality not only produces efficient results but also avoids excessive computational cost.

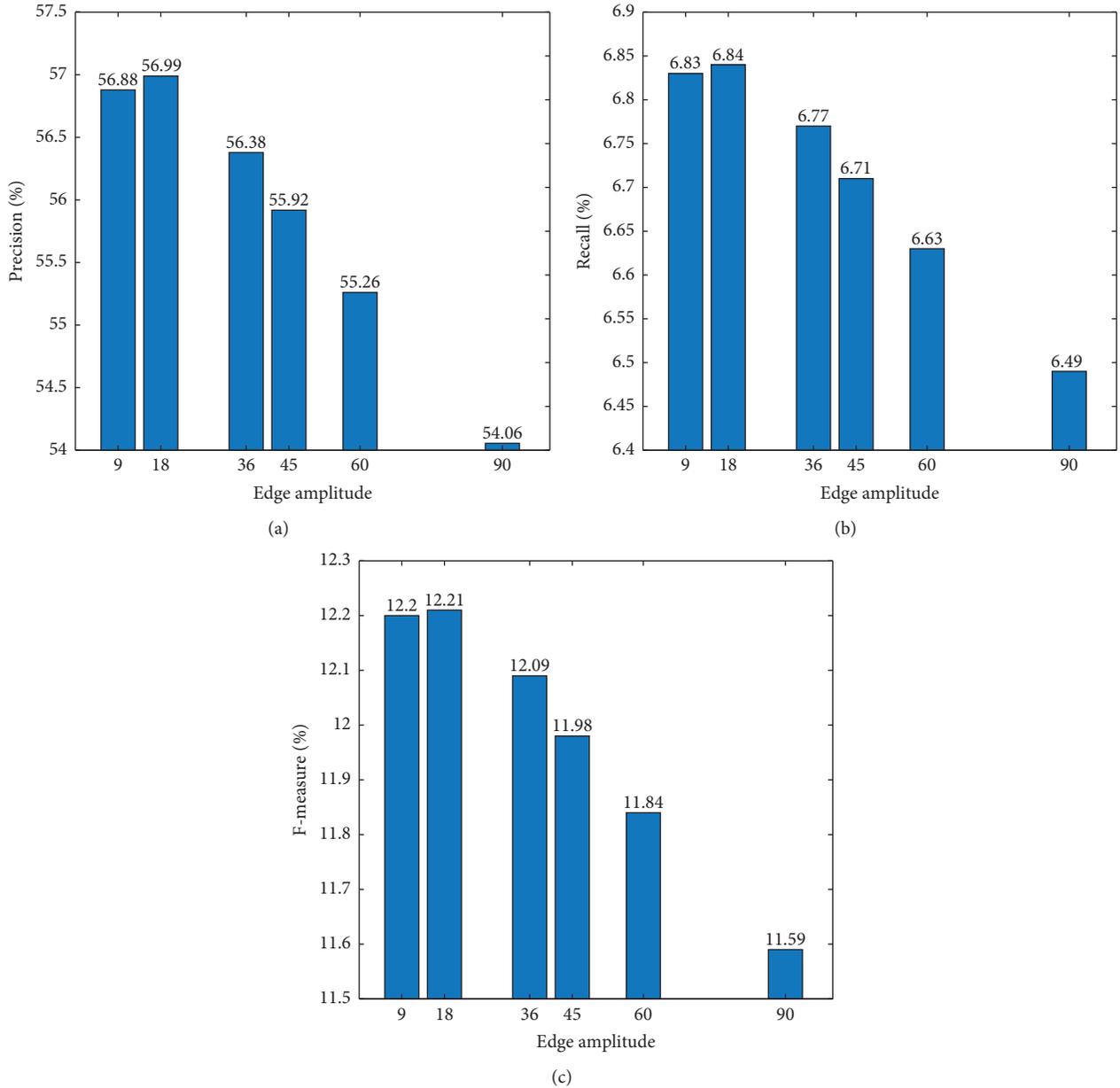


FIGURE 5: Retrieval performance using edge amplitude quantization numbers with the Corel-10K dataset: (a) precision (%), (b) recall (%), and (c) F-measure (%).

In the CBIR experiments, hue ( $H$ ) is uniformly quantified into 6, 8, and 12 bins, and Saturation ( $S$ ) and Value ( $V$ ) are both quantified as 3 bins in HSV colour space. Thus, the quantization numbers of colours are  $6 \times 3 \times 3 = 54$  bins,  $8 \times 3 \times 3 = 72$  bins, and  $12 \times 3 \times 3 = 108$  bins. At the same time, the quantization numbers of edge orientations are 9 bins, 18 bins, 36 bins, 45 bins, 60 bins, and 90 bins, and the quantization numbers of intensity are 16 bins, 32 bins, and 64 bins.

As shown in Figure 4, the precision did not change much as the quantization number of colours increased. Overall, precision declined with increases in the quantization number of intensity and increased with increases in the quantization number of edge orientation. When the

quantization number of edge orientation increased from 18 to 36, the precision increased by more than 1%. When the quantization number of edge orientation was fixed to 36 bins, the quantization number of colour and greyscale were 54 bins and 16 bins, respectively, and the vector dimensionality was the smallest and the precision was the highest. Therefore, we set the quantization numbers of colours as 54 bins and 16 bins.

We further evaluated the effects of the quantization numbers of edge amplitude. The quantization numbers of edge amplitude were 9 bins, 18 bins, 36 bins, 45 bins, 60 bins, and 90 bins. In Figure 5, the precision, recall, and F-measures decrease with increases in the quantization number of edge amplitude. According to the results, we set the

TABLE 1: Effectiveness of adding RGB colour space information and using the selected direction and lateral inhibition formula.

Dataset	Performance (%)	Default setting	Without RGB colour space information	Without lateral inhibition formula	Without selected direction			
					0°	45°	90°	135°
Corel-5K	Precision	66.91	65.50	66.50	66.39	66.08	66.54	66.09
	Recall	8.03	7.86	7.98	7.97	7.93	7.98	7.93
	F-measure	14.34	14.04	14.25	14.23	14.16	14.26	14.16
Corel-10K	Precision	56.88	55.66	56.43	56.41	56.03	56.43	56.04
	Recall	6.83	6.68	6.77	6.77	6.72	6.77	6.72
	F-measure	12.20	11.93	12.09	12.09	12.01	12.09	12.01

TABLE 2: Performance comparison of the W1 distance and other metrics.

Dataset	Performance (%)	Distance or similarity metric						
		L1	Chi-square	Canberra	D1	CDH	STH	Proposed
Corel-5K	Precision	42.44	48.57	65.00	65.09	58.20	60.27	66.91
	Recall	5.09	5.83	7.80	7.81	6.98	7.23	8.03
	F-measure	9.09	10.41	13.93	13.95	12.47	12.91	14.34
Corel-10K	Precision	33.60	39.28	54.79	54.88	48.07	50.62	56.88
	Recall	4.03	4.71	6.58	6.59	5.77	6.07	6.83
	F-measure	7.20	8.42	11.74	11.76	10.30	10.85	12.20

TABLE 3: Comparison of the proposed method with existing state-of-the-art methods.

Dataset	Performance (%)	Method					
		SoC-GMM	LDP	SSH	CPV-THF	STH	Proposed
Corel-5K	Precision	51.80	—	—	63.90	60.28	66.91
	Recall	6.22	—	—	7.66	7.23	8.03
Corel-10K	Precision	47.25	51.91	54.88	52.28	48.03	56.88
	Recall	5.67	6.23	6.58	6.27	5.76	6.83
Feature dimensions		79	134	130	242	172	115

quantization number of edge amplitude to 9 bins. Ultimately, the vector dimensionality of the proposed method is  $54 + 36 + 16 + 9 = 115$  bins.

As shown in Table 1, not using the colour information of the RGB colour space resulted in reductions in precision of at least 1% on the two datasets. If we use the lateral inhibition formula, the performance improved to a certain degree. Performance was better using the selected direction than using any single direction (0°, 45°, 90°, or 135°). Thus, using the selected direction and the lateral inhibition formula and adding the RGB colour space information, improved the performance.

In order to investigate the effectiveness of the proposed W1 distance method, we compared it with the L1 distance (Manhattan distance), the chi-square distance ( $\chi^2$  statistics), the Canberra distance, the D1 distance (weighted L1 distance), and the proposed distance in CDH methods [22]. The performances of these distances or similarity metrics are listed in Table 2. It is clear that the W1 distance performed better than the other metrics on the Corel-5K and Corel-10K datasets. Moreover, the W1 distance metric does not require complex operations such as square or square root and can be regarded as an extended weighted L1 distance [22].

**4.5. Performance Comparisons.** Here, we compare the proposed method with state-of-the-art methods such as SSH [1],

SoC-GMM [55], LDP [56], CPV-THF [57], and STH [58]. The results are listed in Table 3. The precision of the proposed method on the Corel-5K dataset was 15.11%, 3.01%, and 6.63% higher than with SoC-GMM [55], CPV-THF [57], and STH [58], respectively. The precision of the proposed method on the Corel-10K dataset was 9.63%, 4.97%, 2%, 4.6%, and 8.85% higher than with SoC-GMM [55], LDP [56], SSH [1], CPV-THF [57], and STH [58], respectively.

It has rich colours, complex textures, and various shapes within natural images. However, SoC-GMM only describes colour information. LDP, SSH, CPV-THF, and STH represent colour information in a colour space such as RGB, HSV, or Lab. However, both HSV and RGB colour spaces are considered in the proposed IVD method. It is worth mentioning that the IVD method does not increase the vector dimensionality. CPV-THF and STH can analyse texture features using texton templates. However, texton templates cannot adequately describe texture information because they only consider two pixels with the same intensity in  $2 \times 2$  square blocks. The bar-shaped structure of the SSH method only considers situations where three pixels are invariant in intensity [1]. The proposed IVD describes a total of 27 variations in the intensity of three pixels. The IVD method not only describes the intensity variation in a certain direction but can also combine the colour, edges, greyscale, and spatial information. Therefore, the proposed method

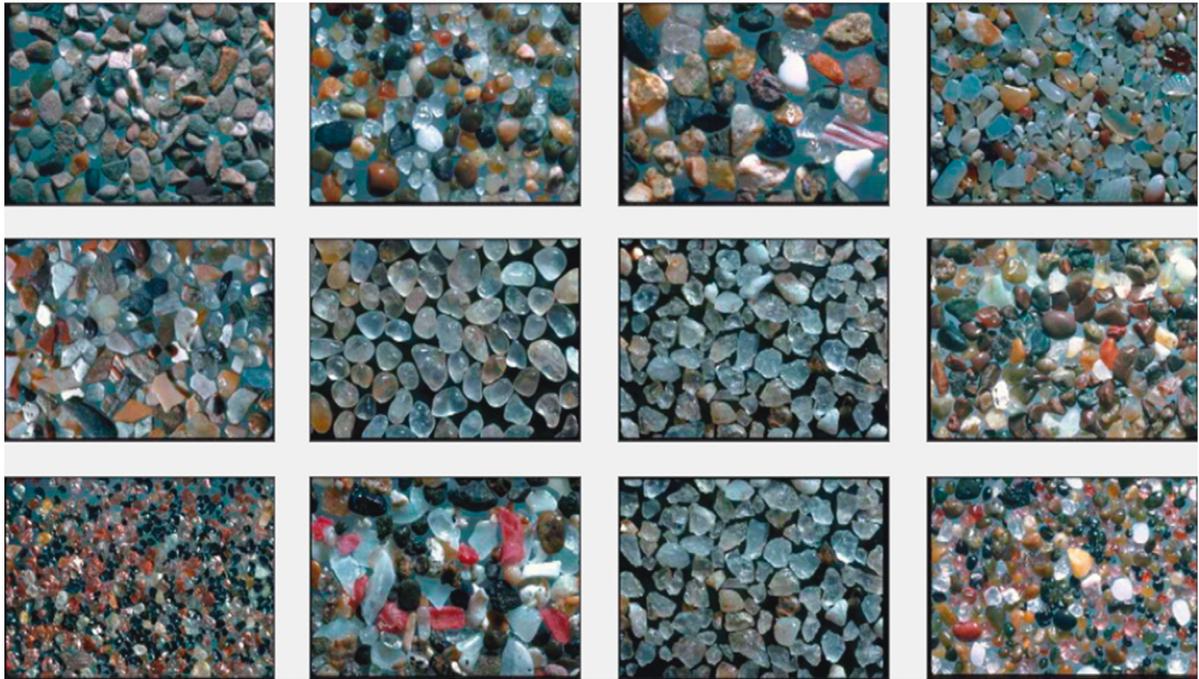


FIGURE 6: Retrieval example of a gravel image from the Corel-5K dataset using the proposed IVD method.



FIGURE 7: Retrieval example of an eagle image from the Corel-10K dataset using the proposed IVD method.

contains richer information than the SoC-GMM, LDP, SSH, CPV-THF, and STH methods.

Figures 6 and 7 show two image retrieval examples from the Corel-5K and Corel-10K datasets. Each example has 12 images, with the top-left image being the query image. In Figure 6, the query is a gravel image, which has varied colours and complex textures. In Figure 7, the query is an eagle image, which has a background of blue sky and the shape of the eagle. It can be seen that all returned images

were correctly ranked within the top 12 images. Furthermore, the colour, texture, and shape of the query image and all the returned images have certain similarities.

Figures 8 and 9 show two image retrieval examples from the Corel-5K and Corel-10K datasets using the SSH method [1], and the queries are also the gravel and an eagle. Seven returned images were incorrectly ranked within the top 12 images in Figure 8. Both the IVD and the SSH methods have discriminatory power for colour, texture, shape, and spatial

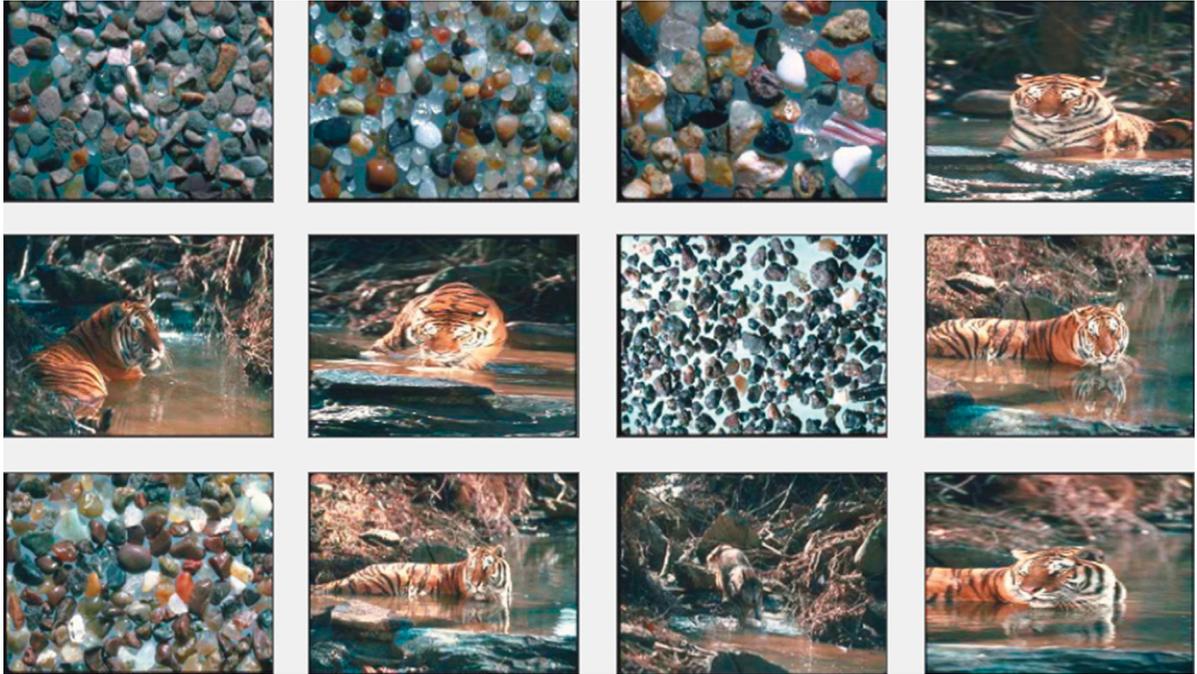


FIGURE 8: Retrieval example of a gravel image from the Corel-5K dataset using the SSH method [1].

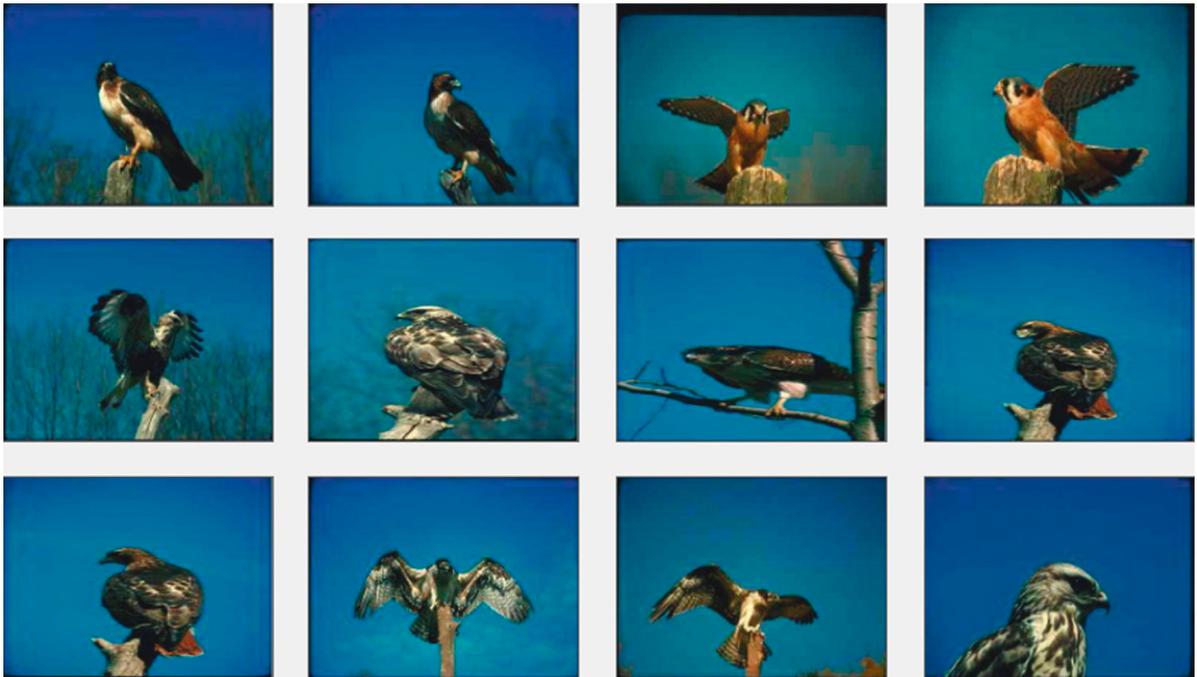


FIGURE 9: Retrieval example of an eagle image from the Corel-10K dataset using the SSH method [1].

features, but the IVD method has strong discriminatory power for texture and significantly outperforms better than that of the SSH method.

## 5. Conclusions

We have proposed a novel image representation, namely, the intensity variation descriptor, to represent image content, and applied it to image retrieval. The proposed descriptor

not only extracts the richness of colour information by combining the HSV and RGB colour spaces but effectively describes texture features by extracting information on edges and intensity variations. It still considers the direction and spatial structure information of the textures by simulating the orientation-selection mechanism and the lateral inhibition mechanism.

We have proposed an extended weighted L1 distance metric to improve the retrieval performance of the proposed

method. Experimental performance comparisons with the Corel-5K and Corel-10K datasets demonstrate that our method outperforms some state-of-the-art methods and has the power to discriminate texture and spatial structure. There are some potential applications of the proposed IVD method, and it can be applied to texture recognition, trademark image retrieval and palmprint image retrieval.

## 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] 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.
- [2] M. Zhao, H. Zhang, and J. Sun, "A novel image retrieval method based on multi-trend structure descriptor," *Journal of Visual Communication and Image Representation*, vol. 38, pp. 73–81, 2016.
- [3] 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.
- [4] J. Trefný and J. Matas, "Extended set of local binary patterns for rapid object detection," *Computer Vision Winter Workshop*, pp. 1–7, 2010.
- [5] U. Sharif, Z. Mehmood, T. Mahmood, M. A. Javid, A. Rehman, and T. Saba, "Scene analysis and search using local features and support vector machine for effective content-based image retrieval," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 901–925, 2018.
- [6] C. Singh and K. Preet Kaur, "A fast and efficient image retrieval system based on color and texture features," *Journal of Visual Communication and Image Representation*, vol. 41, pp. 225–238, 2016.
- [7] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 762–768, San Juan, Puerto Rico, June 1997.
- [8] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *Proceedings of the Fourth ACM International Conference on Multimedia-MULTIMEDIA'96*, pp. 65–73, New York, NY, USA, June 1997.
- [9] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.
- [10] T. Ojala, M. Pietikainen, and T. Maenpää, "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.
- [11] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, 1996.
- [12] W. Y. Ma and B. S. Manjunath, "A comparison of wavelet transform features for texture image annotation," in *Proceedings of the International Conference on Image Processing*, vol. 2, pp. 256–259, Washington, DC, USA, October 1995.
- [13] W. Y. Kim and Y. S. Kim, "A region-based shape descriptor using Zernike moments," *Signal Processing: Image Communication*, vol. 16, no. 1-2, pp. 95–102, 2000.
- [14] S.-F. Chang, T. Sikora, and A. Purl, "Overview of the MPEG-7 standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 688–695, 2001.
- [15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [16] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [17] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004 (CVPR 2004)*, Washington, DC, USA, June 2004.
- [18] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the Alvey Vision Conference*, Manchester, UK, September 1988.
- [19] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Proceedings of the 2011 International Conference on Computer Vision*, pp. 2564–2571, Tampa, FL, USA, December 2011.
- [20] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proceedings Ninth IEEE International Conference on Computer Vision*, Nice, France, October 2003.
- [21] X. Wang and Z. Wang, "A novel method for image retrieval based on structure elements' descriptor," *Journal of Visual Communication and Image Representation*, vol. 24, no. 1, pp. 63–74, 2013.
- [22] 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.
- [23] A. Bala and T. Kaur, "Local texton XOR patterns: a new feature descriptor for content-based image retrieval," *Engineering Science and Technology, an International Journal*, vol. 19, no. 1, pp. 101–112, 2016.
- [24] L. K. Rao, D. V. Rao, and L. P. Reddy, "Local mesh quantized extrema patterns for image retrieval," *SpringerPlus*, vol. 5, no. 1, 976 pages, 2016.
- [25] 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.
- [26] P. Srivastava and A. Khare, "Integration of wavelet transform, local binary patterns and moments for content-based image retrieval," *Journal of Visual Communication and Image Representation*, vol. 42, pp. 78–103, 2017.
- [27] M. Dey, B. Raman, and M. Verma, "A novel colour- and texture-based image retrieval technique using multi-resolution local extrema peak valley pattern and RGB colour histogram," *Pattern Analysis and Applications*, vol. 19, no. 4, pp. 1159–1179, 2016.

- [28] P. Srivastava and A. Khare, "Utilizing multiscale local binary pattern for content-based image retrieval," *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 12377–12403, 2018.
- [29] 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 and Video Processing*, vol. 12, no. 2, pp. 247–254, 2018.
- [30] M. Verma, B. Raman, and S. Murala, "Local extrema co-occurrence pattern for color and texture image retrieval," *Neurocomputing*, vol. 165, pp. 255–269, 2015.
- [31] E. Walia, S. Vesal, and A. Pal, "An effective and fast hybrid framework for color image retrieval," *Sensing and Imaging*, vol. 15, no. 1, p. 93, 2014.
- [32] N. Shrivastava and V. Tyagi, "An efficient technique for retrieval of color images in large databases," *Computers & Electrical Engineering*, vol. 46, pp. 314–327, 2015.
- [33] N. Varish, J. Pradhan, and A. K. Pal, "Image retrieval based on non-uniform bins of color histogram and dual tree complex wavelet transform," *Multimedia Tools and Applications*, vol. 76, no. 14, pp. 15885–15921, 2017.
- [34] L. K. Pavithra and T. S. Sharmila, "An efficient framework for image retrieval using color, texture and edge features," *Computers & Electrical Engineering*, vol. 70, pp. 580–593, 2018.
- [35] J. Ahmad, M. Sajjad, I. Mehmood, S. Rho, and S. W. Baik, "Saliency-weighted graphs for efficient visual content description and their applications in real-time image retrieval systems," *Journal of Real-Time Image Processing*, vol. 13, no. 3, pp. 431–447, 2017.
- [36] J. Ahmad, M. Sajjad, I. Mehmood, and S. W. Baik, "SSH: salient structures histogram for content based image retrieval," in *Proceedings of the 2015 18th International Conference on Network-Based Information Systems (NBIS)*, pp. 212–217, Taipei, Taiwan, September 2015.
- [37] 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, 2018.
- [38] 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.
- [39] G.-H. Liu, "Content-based image retrieval based on Cauchy density function histogram," in *Proceedings of the 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, pp. 506–510, Changsha, China, August 2016.
- [40] G.-H. Liu, "Content-based image retrieval based on visual attention and the conditional probability," in *Proceedings of the International Conference on Chemical, Material, and Food Engineering*, pp. 838–842, Kunming, Yunnan, China, July 2015.
- [41] J.-Z. Hua, 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.
- [42] J. Wu, L. Feng, S. Liu, and M. Sun, "Image retrieval framework based on texton uniform descriptor and modified manifold ranking," *Journal of Visual Communication and Image Representation*, vol. 49, pp. 78–88, 2017.
- [43] S. Liu, J. Wu, L. Feng et al., "Perceptual uniform descriptor and ranking on manifold for image retrieval," *Information Sciences*, vol. 424, pp. 235–249, 2018.
- [44] S. Yu, D. Niu, L. Zhang, M. Liu, and X. Zhao, "Colour image retrieval based on the hypergraph combined with a weighted adjacent structure," *IET Computer Vision*, vol. 12, no. 5, pp. 563–569, 2018.
- [45] L. Zheng, Y. Yang, and Q. Tian, "SIFT meets CNN: a decade survey of instance retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 5, pp. 1224–1244, 2018.
- [46] A. Chadha and Y. Andreopoulos, "Voronoi-based compact image descriptors: efficient region-of-interest retrieval with VLAD and deep-learning-based descriptors," *IEEE Transactions on Multimedia*, vol. 19, no. 7, pp. 1596–1608, 2017.
- [47] Y. Wei, Y. Zhao, C. Lu et al., "Cross-modal retrieval with CNN visual features: a new baseline," *IEEE Transactions on Cybernetics*, vol. 47, no. 2, pp. 449–460, 2017.
- [48] P. Liu, J.-M. Guo, C.-Y. Wu, and D. Cai, "Fusion of deep learning and compressed domain features for content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 26, no. 12, pp. 5706–5717, 2017.
- [49] B.-H. Yuan and G.-H. Liu, "Image retrieval based on gradient-structures histogram," *Neural Computing and Applications*, 2019.
- [50] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice-Hall, Upper Saddle River, NJ, USA, 3rd edition, 2007.
- [51] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of Physiology*, vol. 148, no. 3, pp. 574–591, 1959.
- [52] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of Physiology*, vol. 160, no. 1, pp. 106–154, 1962.
- [53] F. W. Campbell and J. J. Kulikowski, "Orientational selectivity of the human visual system," *The Journal of Physiology*, vol. 187, no. 2, pp. 437–445, 1966.
- [54] F. Liu, H. Duan, and Y. Deng, "A chaotic quantum-behaved particle swarm optimization based on lateral inhibition for image matching," *Optik*, vol. 123, no. 21, pp. 1955–1960, 2012.
- [55] 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, 2016.
- [56] J.-x. Zhou, X.-d. Liu, T.-w. Xu, J.-h. Gan, and W.-q. Liu, "A new fusion approach for content based image retrieval with color histogram and local directional pattern," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 4, pp. 677–689, 2018.
- [57] 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.
- [58] 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, 2018.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

