Hindawi
Computational Intelligence and Neuroscience
Volume 2020, Article ID 8876480, 10 pages
https://doi.org/10.1155/2020/8876480

Research Article

Image Retrieval Using the Fused Perceptual Color Histogram

Guang-Hai Liu and Zhao Wei

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 26 March 2020; Revised 24 September 2020; Accepted 22 October 2020; Published 25 November 2020

Academic Editor: Ezequiel López-Rubio

Extracting visual features for image retrieval by mimicking human cognition remains a challenge. Opponent color and HSV color spaces can mimic human visual perception well. In this paper, we improve and extend the CDH method using a multi-stage model to extract and represent an image in a way that mimics human perception. Our main contributions are as follows: (1) a visual feature descriptor is proposed to represent an image. It has the advantages of a histogram-based method and is consistent with visual perception factors such as spatial layout, intensity, edge orientation, and the opponent colors. (2) We improve the distance formula of CDHs; it can effectively adjust the similarity between images according to two parameters. The proposed method provides efficient performance in similar image retrieval rather than instance retrieval. Experiments with four benchmark datasets demonstrate that the proposed method can describe color, texture, and spatial features and performs significantly better than the color volume histogram, color difference histogram, local binary pattern histogram, and multi-texton histogram, and some SURF-based approaches.

1. Introduction

In the fields of image retrieval, pattern recognition, computer vision, and digital image processing, mimicking human cognition remains a challenge. In the human visual system, the perception of color begins with three types of cones in the retina called the *L*, *M*, and *S* cones. These contain pigments with different spectral sensitivities that produce trichromatic color sensations [1]. The LMS color space represents the response of the three types of cone photoreceptors in the human eye and can be translated into other color spaces or models. This allows the opponent and HSV color spaces to be calculated easily. Then, the question arises: how are opponent and HSV color spaces used to extract visual features for image retrieval? A multi-stage color model that combines the three-photoreceptors model with the opponent theory has been suggested [1]. The neural signals from the three cone photoreceptors of the eye are combined into opponent color channels at the retinal level and then transmitted to the brain. Thus, it is possible to utilize a multi-stage color model to describe and represent an image for image retrieval.

In previous work, we have proposed color difference histograms (CDHs) [2] to image retrieval based on the CLE $L^*a^*b^*$ color space [2]. However, the traditional CIE color difference formula was originally designed for simple color patches in controlled viewing conditions and is not adequate for computing image differences for spatially complex image stimuli [3]. In this paper, we improve and extend the CDH method using a multi-stage model to extract and represent an image in a way that mimics human perception.

Our main contributions are as follows: (1) a novel visual feature descriptor is proposed to represent an image. It has the advantages of a histogram-based method and is consistent with visual perception factors such as spatial layout, intensity, edge orientation, and the opponent colors. (2) We improve the distance formula of CDHs; it can effectively adjust the similarity between images according to two parameters. The proposed method provided efficient performance in similar image retrieval rather than object searching.

The rest of this paper is organized as follows. Section 2 reviews image retrieval techniques from literature published in recent decades, while Section 3 describes the fused perceptual color histogram. We describe CBIR experiments in Section 4; then Section 5 concludes the paper.

2. Related Work

The widely used attributes used to represent image content are color, texture, and shape features. In the MPEG-7 standard, different methods have been proposed to describe these features and different descriptors have been applied in CBIR. The color histogram is a widely used method to describe color features. It is invariant to orientation and scale and can give powerful image classification potential. For this reason, and for its simplicity and effectiveness, color descriptors were very popular in the early days of CBIR. The color descriptors used in the MPEG-7 standard include the dominant color descriptor, color layout descriptor, color structure descriptor, and scalable color descriptor [4]. Those color descriptors aim to provide a compact color description, capture the spatial distribution of color, and express the local color structure. Several other color features have also been proposed. Color volume and color difference have been utilized to extract color features [2, 5] and for saliency detection [6]. Varior et al. proposed the learning of invariant color features for person re-identification [7].

Texture descriptors can be used to characterize repeated geometric patterns or color regions. In the MPEG-7 standard, texture descriptors include the homogeneous texture descriptor, texture browsing descriptor, and edge histogram descriptor[4]. In recent decades, many texture analysis methods have been proposed, including Haralick's gray co-occurrence matrix (GLCM) features [8], the Markov random field (MRF) model [9], and local binary patterns (LBP) [10]. In recent years, many algorithms have been proposed for combining multiple visual cues to improve discriminative power. Liu et al. proposed the texton-based methods for image retrieval [11–13]. Singh et al. proposed a color texture descriptor based on local binary patterns for color image retrieval [14]. In order to capture the evolution of repeated geometric patterns, Thompson et al. proposed the edge-local binary pattern (edgeLBP) technique for 3D object retrieval and classification [15]. Dubey et al. proposed a multichannel decoded LBP method and utilized it for CBIR [16]. Roberto et al. proposed the orthogonal moments for texture classification [17]. A set of Gabor filters with different frequencies and orientations can mimic the perception of the human visual system (HVS). It is helpful for extracting useful visual features from an image for various applications. Based on saliency cues, bar-shaped structures, and Gabor filters, Liu et al. proposed the salient-structure histograms to image retrieval and achieved excellent performance [18].

Shape plays an important role in understanding and identifying objects; however, it is difficult to extract shape features. In many cases, shape feature extraction is often performed via accurate segmentation, which is a very difficult issue in image processing. In the MPEG-7 standard, the region-based shape descriptor, contour-based shape descriptor, and 3D shape descriptor are considered to provide good approximations of segmentation. These descriptors can describe the regions, contours, and shapes of 2D images and 3D volumes. In order to avoid accurate segmentation, some local feature descriptors, including scale-invariant feature transform (SIFT) descriptors [19], and the histograms of oriented gradients (HOG) [20], are also used in shape matching and recognition. Hong et al. proposed a novel shape descriptor that characterizes the local shape geometry based on integral kernels with respect to the size of the shape at a range of feature scales [21]. Clement et al. proposed a structural object description by learning spatial relations and shapes and utilized it for object recognition [22]. Žunić et al. introduced a disconnectedness measure for multi-component shapes [23]. Liu et al. proposed a novel structured optimal graph based sparse feature extraction method for learning the local discriminative information [24]. Malu et al. proposed a dynamic circular mesh-based shape and margin descriptor to combine the functions of structural and global contour-based descriptors and utilized it for object detection [25]. Mehmood et al. have extended the local feature descriptors for image retrieval by using the visual words model [26–29].

In the last decade, deep learning, especially by convolutional neural networks (CNNs), has been successfully applied to a variety of domains [30–38]. It requires a mass of data for training and provides useful information for various applications, including image retrieval and pattern recognition. CNN-based methods use a pre-trained or fine-tuned CNN to extract features for image retrieval and classification, for instance, by extraction of global features from its fully connected layer and extraction of local features from its intermediate layer [30–34]. Discovering how to combine deep learning with large amounts of data could provide computers with human-like image recognition capabilities. However, this field is immature and many challenges remain.

In this paper, we propose a simple yet efficient image retrieval method that simulates the dual-stage model of color vision to mimic human color perception.

3. The Fused Perceptual Color Histogram

Feature extraction has a close relationship to color space. In digital image processing, the RGB color space is very popular for representing color but has two obvious shortcomings: (1) it is not directly based on the cones in the human eye and (2) it is not uniform with respect to human color perception. In previous work, we proposed using color difference histograms (CDHs) [2] in image retrieval. The unique characteristic of CDHs is the way they count the perceptually uniform color differences between two points with different backgrounds with regard to colors and edge orientations in the CLE $L^*a^*b^*$ color space [2]. However, the traditional CIE color difference formula was originally designed for simple color patches in controlled viewing conditions and is not adequate for computing image differences for spatially complex image stimuli [3].

In this paper, we improve and extend the CDH method by utilizing a dual-stage model to extract and represent an image in a way that mimics human perception. We propose a novel visual descriptor based on fused perceptual color information by using the attributes of the opponent color and HSV color spaces. It aims to represent image content using intensity, color, and edge orientation features in the opponent color and HSV color spaces, giving it the power to describe color, texture, edge, and spatial features. Figure 1 illustrates the proposed feature extraction and discriminative representation system within the CBIR framework, which is composed of three parts: (1) RGB color space conversion into other color spaces, including XYZ, LMS, and HSV, (2) primary visual feature calculations from the HSV color space, and (3) image representation and image retrieval.

3.1. LMS and Opponent Color Spaces. In the trichromatic theory of human color vision, it is suggested that there are three kinds of cone cells (also called photoreceptors) with different spectral sensitivities. Signals produced by the three photoreceptors are sent to the central nervous system and perceived as color sensations [1]. In normal human trichromacy, the three kinds of photoreceptors having peak sensitivities in the large-, medium-, and short-wavelength portions of the visible spectrum are called the *L*, *M*, and *S* cones, respectively [3].

In order to make good use of perceptual color information, we first convert the original color image from the RGB color space to the LMS space in two steps. The first is a conversion from RGB to XYZ tristimulus values, which is a device-independent color space. This conversion can be calculated as follows [3]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.5141 & 0.3239 & 0.1604 \\ 0.2651 & 0.6702 & 0.0641 \\ 0.0241 & 0.1228 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{1}$$

In the device-independent XYZ space, we can convert the original image from the XYZ to LMS color space using the following conversion [3]:

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3897 & 0.6890 & -0.0787 \\ -0.2298 & 1.1834 & 0.0464 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \tag{2}$$

In the LMS color space, a great deal of skew is shown in the data. In order to largely eliminate this skew, we can convert the data to a logarithmic space [3]:

$$\begin{cases} L = \log L, \\ M = \log M, \\ S = \log S. \end{cases} \tag{3}$$

The large-, medium-, and short-wavelength cone signals (LMS) are combined to form a variant of the opponent color model called the AC_1C_2 opponent color space [39], which is calculated as follows:

$$\begin{bmatrix} A \\ C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 0.990 & -0.106 & -0.094 \\ -0.669 & 0.742 & -0.027 \\ -0.212 & -0.354 & 0.911 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix}. \tag{4}$$

After the above calculations or conversions, we utilize the LMS and AC_1C_2 opponent color spaces to extract visual features by using filters that approximate the contrast sensitivity functions (CSFs) of the human visual system [40].

3.2. HSV Color Space. The HSV color space can be represented as a cylindrical coordinate system in which H, S, and V are its three coordinate variables. Variable H signifies the hue, which represents the perceived colors red, yellow, green, and blue, or a combination of two of them [41]. Saturation (S) refers to the relative purity or degree to which the color is combined with white, and Value (V) indicates the brightness relative to a similarly illuminated white color [41, 42].

In the HSV cylindrical coordinate system [42], H represents the angle of rotation and ranges from 0 to 360 degrees; S represents the size of the radius (range = 0-1); and V indicates the height of the cylinder (range = 0-1), as shown in Figure 2.

3.3. Feature Quantization. In this section, color, edge orientation, and intensity maps are utilized to extract visual features via the feature quantization technique. The HSV color space is widely utilized in the field of image retrieval, is consistent with human visual perception, and can better describe the content of an image. Therefore, the color, edge orientation, and intensity features are extracted in HSV color space.

The quantized color comes from various combinations of the H, S, and V components. In our method, the H component is uniformly quantized into 6 bins, and both S and V components are uniformly quantized into 3 bins, resulting in the color map $C(x, y) = \omega, \omega \in \{0, 1, \dots, N_C - 1\}$ and $N_C = 54$.

Compared with the color and edge orientation extraction method utilized in CDHs [2], the Sobel operator is a convenient and simple edge detector. Edge orientation is first extracted from the V component using the Sobel operator and then uniformly quantized into $N_O = 36$ bins. We denote the edge orientation map as $O(x, y) = \varepsilon, \varepsilon \in \{0, 1, \dots, N_O - 1\}$.

Here, the intensity map $I(x, y)$ is obtained directly by uniformly quantizing the V component, where $I(x, y) = \tau, \tau \in \{0, 1, \dots, N_I - 1\}$ and $N_I = 16$.

3.4. Approximating the Contrast Sensitivity Functions (CSFs). The human visual system is much less sensitive to colors at high frequencies than at low ones. Hence, using contrast sensitivity functions (CSFs) to modulate frequencies that are less perceptible can better simulate the human visual system [39, 40]. In this paper, the CSFs are first used to remove information that is invisible to the human visual system in the opponent color space AC_1C_2.

The AC_1C_2 color space is spatially filtered using the CSFs by computing the difference between two points under various backgrounds in terms of colors, edge orientations, and intensity. Each channel in AC_1C_2 is spatially filtered by

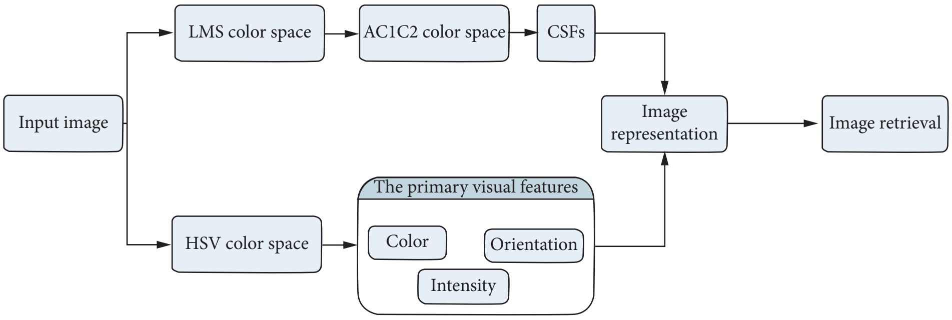


Figure 1: Flow diagram of the proposed feature extraction and discriminative representation system within the CBIR framework.

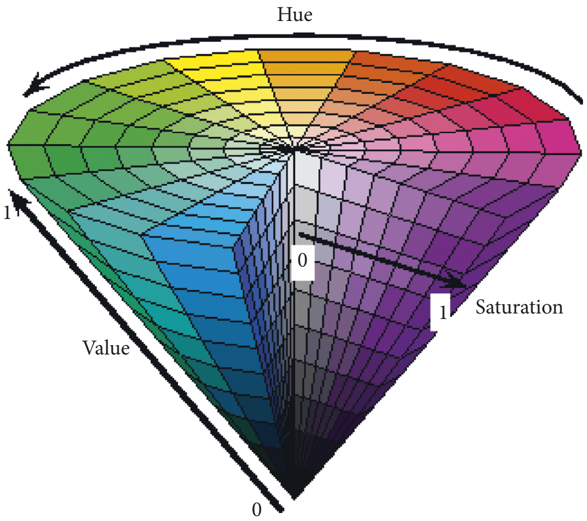


Figure 2: Illustration of the HSV color space [43].

using the CSFs to approximate the human visual system, as expressed in formulas (5) and (6) [39].

$$E_i = e^{-\left(x^2+y^2\right)/\sigma_i^2}, \tag{5}$$

$$\begin{cases} A\prime = \dfrac{\sum_{i\in[1,2,3]} w_i \left(A \otimes E_i\right)}{3}, \\ C_1' = \dfrac{\sum_{i\in[1,2]} w_i \left(C_1 \otimes E_i\right)}{2}, \\ C_2' = \dfrac{\sum_{i\in[1,2]} w_i \left(C_2 \otimes E_i\right)}{2}. \end{cases} \tag{6}$$

The weights (w_i) and spreads (σ_i) of the CSFs are listed in Table 1. The filtered opponent color space is denoted as $A'C_1'C_2'$. Here, we utilize $A'C_1'C_2'$ to represent features, together with color, edge orientation, and intensity maps.

3.5. Feature Representation. Let there be two-pixel locations (x, y) and (x', y') with d as the spacing distance. Then, the feature representation of $C(x, y)$, $O(x, y)$, and $I(x, y)$ can be expressed as follows:

$$H_c[C(x, y)] = \begin{cases} \sum_{x=1}^{M-1} \sum_{y=1}^{N-1} \sqrt{\Delta A^2 + \Delta C_1^2 + \Delta C_2^2}, \\ \text{where } C(x, y) = C(x', y'), \end{cases}$$

$$H_o[O(x, y)] = \begin{cases} \sum_{x=1}^{M-1} \sum_{y=1}^{N-1} \sqrt{\Delta A^2 + \Delta C_1^2 + \Delta C_2^2}, \\ \text{where } O(x, y) = O(x_i, y_i), \end{cases} \tag{7}$$

$$H_I[I(x, y)] = \begin{cases} \sum_{x=1}^{M-1} \sum_{y=1}^{N-1} |\Delta A|, \\ \text{where } I(x, y) = I(x', y'). \end{cases}$$

In the above formulas, M and N are the width and height of the image, respectively, $\Delta A = A'(x, y) - A'(x', y')$, $\Delta C_1 = C_1'(x, y) - C_1'(x', y')$, and $\Delta C_2 = C_2'(x, y) - C_2'(x', y')$. According to the above representations, the fused perceptual color histogram H can be obtained by concatenation of $H_c[C(x, y)]$, $H_o[O(x, y)]$, and $H_I[I(x, y)]$ as follows:

$$H = \text{CONCA}\{H_c, H_o, H_I\}, \tag{8}$$

where CONCA$\{\cdot\}$ denotes the concatenation of H_c, H_o, and H_I. The fused perceptual color histogram H contains the perceived color information related to the color, edge orientation, and intensity maps.

4. Experimental Results

In this section, we verify the effectiveness of the proposed method on four benchmark datasets containing more than 20,000 natural images, including Corel-5K, Corel-10K, Oxford buildings, and INRIA Holidays datasets. Image matching is adopted based on an improved distance formula of CDHs. For fair comparison, the proposed method will be compared with the current image retrieval methods, HOG [20], LBP [10], MTH [12], CDHs [2], CVH [5], BOW [38], and other methods [43, 44]. Most of these methods were developed for image retrieval. The details of the comparison methods are shown as follows:

(1) The codebook size for Bow was set as $K = 1000$ using the standard K-means clustering, and the cosine metric was

Table 1: Parameters of the CSFs [39].

Filters		Weight (w_i)	Spread (σ_i)
A(Achromatic)	$i = 1$	1.00327	0.0500
	$i = 2$	0.11442	0.2250
	$i = 3$	−0.11769	7.0000
C_1(Red-green)	$i = 1$	0.61673	0.0685
	$i = 2$	0.38328	0.8260
C_2(Blue-yellow)	$i = 1$	0.56789	0.0920
	$i = 2$	0.43212	0.6451

used as the baseline of the Bow method; the local features are represented by the SIFT descriptors [20]. An LBP histogram with a dimensional feature vector of 256 bins and using the average values for the three-channel LBP histogram. The histogram of oriented gradients (HOG) feature descriptor is not a global image representation method; there are nine bins with a block size of three and a cell size of six. The L1 distance was adopted as the similarity measure of LBP histogram and HOG method. (2) CDHs [2], MTH [12], and CVH [5] follow the original setting of image representation and similarity measures. (3) The results of other methods [43,44] come from their conference reports.

4.1. Datasets. In the field of image retrieval, Corel datasets are the most widely utilized. Many algorithms have been used in CBIR experiments on Corel datasets for comparison. In this paper, we implemented a CBIR experiment with two Corel datasets: Corel-5K and Corel-10K. The Corel-5K dataset is a subset of the Corel-10K dataset and contains 50 classes. Each category includes 100 images sized 192 × 128 or 128 × 192 pixels in JPEG format. The Corel-10K dataset contains richer image content in 100 categories with 100 images in each category.

For comparisons of instance retrieval methods, we also evaluate our method on the Oxford5k and Holidays datasets. The Oxford5k contains 5,062 images which have 11 different Oxford landmarks. Each landmark is represented by 5 possible queries, and it leads to a set of 55 queries, over which an object retrieval system can be evaluated. The Holidays dataset has 1,491 images which contains 500 queries and 991 corresponding relevant images.

4.2. Distance Metrics. After feature extraction, the matching of images using distance metrics is a very important part of image retrieval. In previous work with CDHs [2], a new distance formula was proposed by expanding the Canberra distance. In this paper, we improve the distance formula of CDHs. Let $T = \{T_i\}_1^K$ and $Q = \{Q_i\}_1^K$ be the K-dimensional feature vectors of a template image and query image, respectively. The distance between them is simply calculated as follows:

$$D(T, Q) = \sum_{i=1}^{K} \frac{|Q_i - T_i|}{|Q_i + w_1 \cdot T_i + w_2 \cdot \mu_T|}, \quad (9)$$

where μ_T is the mean of T and w_1 and w_2 are weight parameters used to enhance the difference between small bins and reduce the difference between large bins. In this paper, K is set to 106 bins, $w_1 = 1.4$, and $w_2 = 0.2$.

4.3. Performance Measures. All images were sampled for use as query images in each Corel dataset. Performance was evaluated using the average results of each query in terms of precision and recall. They are the most common performance evaluation criteria used in CBIR. They are defined as follows [2, 12, 13, 18]:

$$\text{Precision} = \frac{I_N}{N}, \qquad \text{Recall} = \frac{I_N}{M}, \quad (10)$$

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 dataset that are similar to the query image. Here, we set $N = 12$ and $M = 100$.

On the Oxford5k and Holidays datasets, mean average precision (mAP) is utilized to evaluate the performance of FPCH and other compared algorithms [43, 44]. We are following the original setting of query images and the corresponding relevant images.

4.4. Retrieval Performance and Discussion. In the proposed method, color, edge orientation, and intensity are utilized in representation, and their quantization number determines the vector dimensionality. Lower vector dimensionality not only is beneficial to rapid image retrieval but also requires less computation. Therefore, in the experiments, the quantization number of the above visual features needs to be determined and evaluated. We then investigate the influence of the feature quantization number.

The color quantization number consists of H, S, and V values. We set H to 6, 8, and 12, while S and V are fixed to 3. Hence, the color quantization number has 54 bins, 72 bins, and 108 bins. Furthermore, the quantization number of edge orientation has 6 bins, 12 bins, 18 bins, 24 bins, 30 bins, 36 bins, and 45 bins. The quantization number of intensity has 16 bins, 32 bins, and 64 bins. The experimental results for the Corel-10K dataset are shown in Figures 3–5. There is an evident phenomenon where the precision always increases as the edge orientation quantization number increases. When the intensity quantization number is 16 bins, the precision is inversely proportional to the color quantization number. However, the precision is proportional to the color quantization number when the intensity quantization numbers are 32 bins and 64 bins. In total, the precision decreases as the intensity quantization number increases. In order to balance the performance of the proposed algorithm with the number of feature vector dimensions, we ultimately choose a color quantization number of 54 bins, an edge orientation quantization number of 36 bins, and an intensity quantization number of 16 bins, resulting in a feature vector with a total of 106 dimensions.

In order to illustrate the validity of the distance formula proposed in this paper, we compare its performance with the typical L1, L2, Canberra, and CDHs distance formulas in experiments on the Corel-10K dataset. The experimental results are shown in Figure 6, which shows that the proposed distance formula has the best performance. The Canberra, CDHs, and proposed distance formulas can all be regarded as weighted L1 distances with different weights. This weight can reduce the influence of differences between large bins in the histograms. The Canberra distance simply utilizes the reciprocal of the bin of the template and query images as the weight. The CDHs distance formula adds the mean of the template and query images based on the Canberra distance, while the proposed distance formula utilizes two parameters to adjust the weight accurately. Thus, the proposed distance formula can achieve better results.

In order to validate the performance of the proposed FPCH method, we compare it with CDHs [2], LBP [10], MTH [12], CVH [5], HOG [20], and BOW [38]. The experimental results are shown in Table 2. It can be seen that the precision of the proposed FPCH method is higher than those of LBP, MTH, CDHs, and CVH by 19.68%, 13.52%, 6.27%, and 3.37% on the Corel-5K dataset, respectively.

On the Corel-10K dataset, the precision of the proposed FPCH method is higher than BOW, HOG, LBP, MTH, CDHs, and CVH by 22.83%, 25.24%, 15.96%, 12.32%, 7.95%, and 4.61%, respectively. The recall of the proposed FPCH method is higher than those of the above methods on both datasets.

BOW and HOG are two typical methods of image retrieval based on object recognition; LBP and MTH are two different types of texture analysis methods, while LBP focuses on describing the spatial structure of texture. MTH represents the texton attributes of images through a series of fixed-size blocks with a certain number of identical pixels. Both CDHs and CVH can simulate human color perception. The proposed FPCH method can fuse the perceptual color information of the opponent color and HSV color spaces and can represent an image through edges, spatial structure, and texture information. Experiments on the Corel-5K and Corel-10K datasets show that the proposed FPCH method is superior to the above methods.

On the Holidays and Oxford5K datasets, we compared the FPCH method with some key-point-based or local-feature-based methods, including the extension or combination of SURF, VLAD, SOP, and RootHSV [44], and the HeW method using deep Conv layer of VGG16 [43].

As can be seen from Table 3, the proposed FPCH method is superior to the 8-SURF, 64-SURF, 4-RootHSV-L1, and 4-RootHSV-L2 methods [44]. However, the mAP of the proposed FPCH method is lower than that of HeW using deep Conv layer of VGG16 [43]. According to the results of Table 4, the proposed FPCH method is completely unsuitable for object searching.

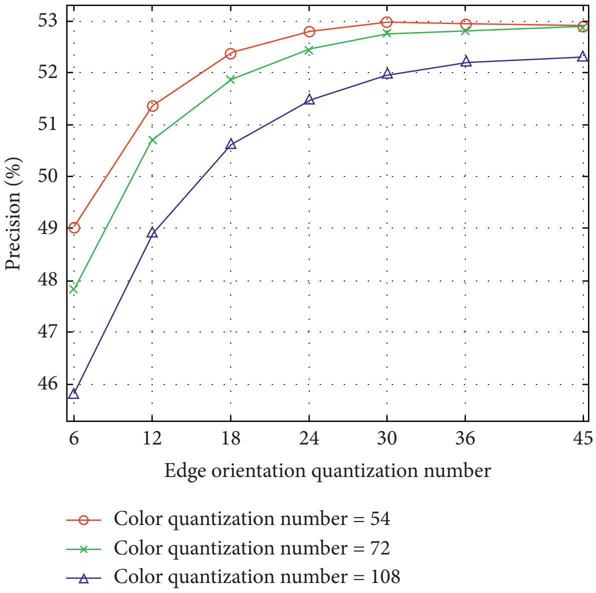


Figure 3: CBIR precision according to quantization numbers of color and edge orientation (intensity quantization number = 16).

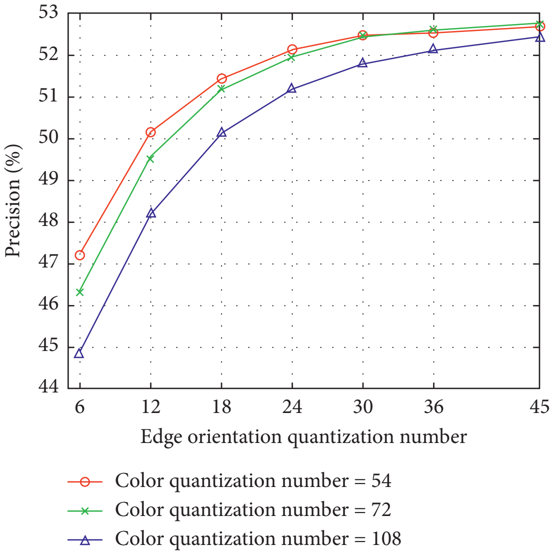


Figure 4: CBIR precision according to quantization numbers of color and edge orientation (intensity quantization number = 32).

In order to visualize the retrieval effect of the proposed FPCH method, two images from the Corel-5K and Corel-10K datasets were selected as query images. The retrieval results are shown in Figures 7(a) and 7(b), where the top-left image is the query and 12 images were retrieved. It is worth noting that images of graffiti have rich color changes and images of furniture have significant differences on both sides

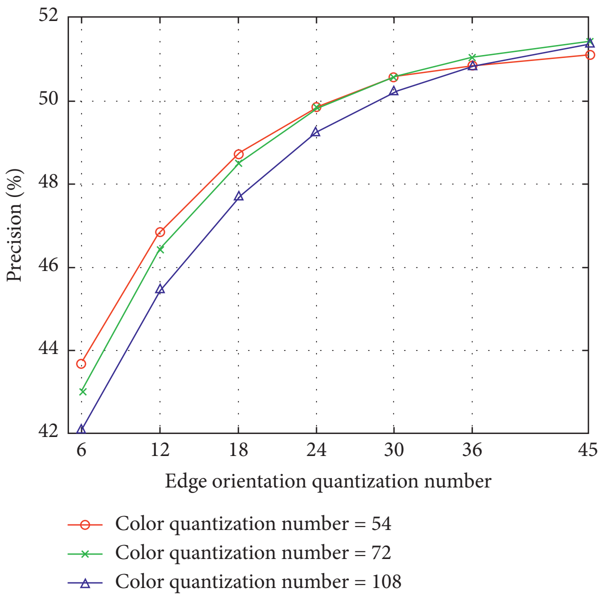
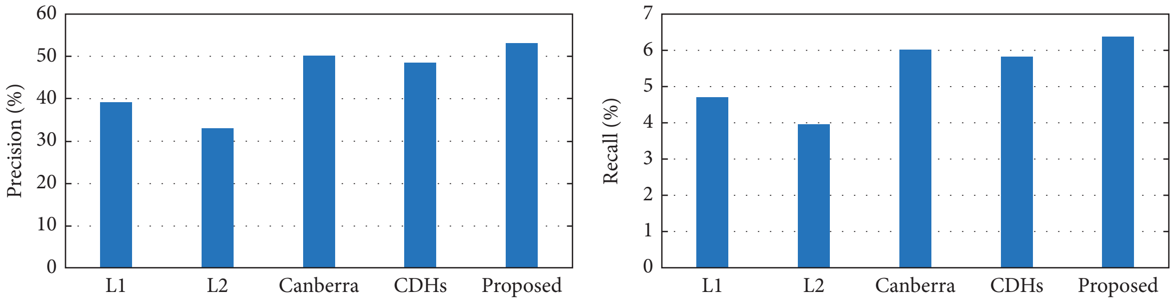


Figure 5: CBIR precision according to quantization numbers of color and edge orientation (intensity quantization number = 64).



(a)

(b)

Figure 6: Performance comparison of distance formulas in (a) precision and (b) recall.

Table 2: Performance comparison of various CBIR methods.

Dataset	Performance	Method						
		BOW	HOG	LBP histogram	MTH	CDHs	CVH	FPCH
Corel-5K	Precision (%)	—	—	43.82	49.98	57.23	60.13	63.50
	Recall (%)	—	—	5.26	6.00	6.87	7.21	7.62
Corel-10K	Precision (%)	30.36	27.95	37.23	40.87	45.24	48.58	53.19
	Recall (%)	3.64	3.35	4.47	4.91	5.43	5.83	6.38

of the objects' edges. It is clear that the proposed FPCH method can mimic human color perception and considers the differences between color, edge, and achromatic features. Therefore, 12 images similar to the query image can be correctly retrieved. Two retrieval examples are used to show the visual effects of low-level features rather than whether or not the performance is good since not all images can provide such a good effect.

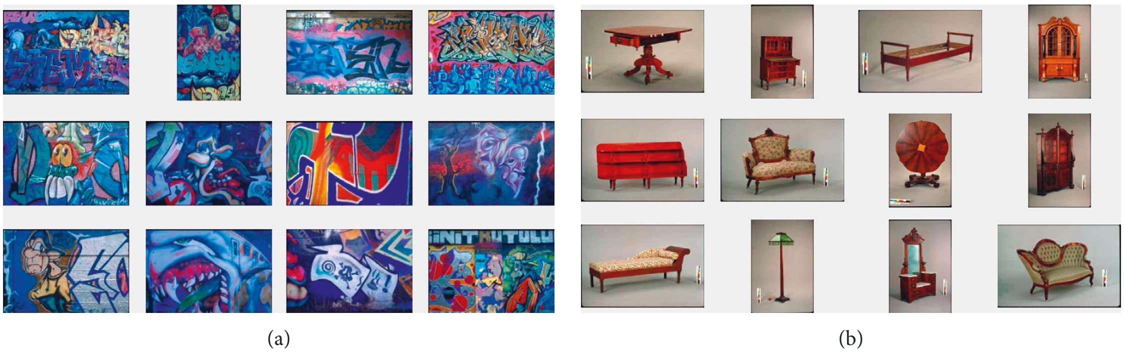
4.5. Limitations of the Proposed Method. Although the proposed method not only approximates the contrast sensitivity functions (CSFs) of the human visual system to filter out information that is invisible to humans but also utilizes color, edge orientation, and intensity to represent and describe image features, the major limitation of the proposed method is that it cannot extract the local features and the high-level features. It is clear that the proposed FPCH

Table 3: Performance comparison of various CBIR methods on the Holidays dataset.

Methods	Dimension	mAP
8-SURF [44]	128	0.463
64-SURF [44]	128	0.625
4-RootHSV-L1 [44]	128	0.657
4-RootHSV-L2 [44]	128	0.675
HeW [43]	512	0.884
FPCH	106	0.699

Table 4: Performance comparison of various CBIR methods on the Oxford5K dataset.

Methods	Dimension	mAP
VLAD + SOP + 2-step + 8-SURF	1024	0.267
VLAD + SOP + 2-step + 64-SURF	4608	0.416
HeW [43]	512	0.728
FPCH	106	0.103


(a)

(b)

Figure 7: Two query examples from the Corel-5K and Corel-10K datasets: (a) graffiti and (b) furniture.

method is entirely unsuitable for object searching according to the results of Table 4. Combining the high-level features with low-level features based on deep learning techniques will be studied in the future.

5. Conclusions

In this paper, we improve and extend the CDH method by utilizing a multi-stage model to extract and represent an image in a way that mimics human perception. We have proposed an image retrieval method that combines the attributes of the opponent color and HSV color spaces, namely, the fused perceptual color histogram. It aims to represent image content using intensity, color, and edge orientation features in the opponent color and HSV color spaces, allowing it to describe color, texture, edge, and spatial features.

In this process, we not only approximate the contrast sensitivity functions (CSFs) of the human visual system to filter out information that is invisible to humans but also utilize color, edge orientation, and intensity to represent and describe image features. The results of the experiments have shown that the proposed method can effectively describe the color, texture, and spatial structure of images and achieves better performance than existing techniques such as LBP, CDHs, CVH, 8-SURF, and 64-SURF methods on Holidays dataset.

The proposed method provides efficient performance in similar image retrieval rather than object searching.

Data Availability

The data 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] L. Russell, V. De, and K. D. Valois, “A multi-stage color model,” *Vision Research*, vol. 33, no. 8, pp. 1053–1065, 1993.

[2] G.- H Liu and J.-Y. Yang, “Content-based image retrieval using color deference histogram,” *Pattern Recognition*, vol. 46, no. 1, pp. 188–198, 2013.

[3] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.

[4] B. S. Manjunath, P. Salembier, and T. Sikora, "Introduction to MPEG-7," in *Multimedia Content Description Interface*John Wiley & Sons, Hoboken, NJ, USA, 2002.

[5] 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.

[6] 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.

[7] R. R. Varior, G. Wang, J. Lu, and T. Liu, "Learning invariant color features for person reidentification," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3395–3410, 2016.

[8] R. M. Haralick and D. Shangmugam, "Textural feature for image classification," *IEEE Transactions on System, Man, and Cybernetics SMC-*, vol. 3, no. 6, pp. 610–621, 1973.

[9] G. Cross and A. Jain, "Markov random field texture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 1, pp. 25–39, 1983.

[10] T. Ojala, M. Pietikanen, and T. Maenpaa, "Multi-resolution 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] 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.

[12] G.-H. Liu and L. Zhang, "Image retrieval based on multi-texton histogram," *Pattern Recognition*, vol. 43, no. 7, pp. 2380–2389, 2010.

[13] 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.

[14] 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, 2017.

[15] 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.

[16] 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.

[17] C. D. Ruberto, L. Putzu, and G. Rodriguez, "Fast and accurate computation of orthogonal moments for texture analysis," *Pattern Recognition*, vol. 83, pp. 498–510, 2018.

[18] G.-H. Liu, J.-Y. Yang, and Z. Y. Li, "Content-based image retrieval using computational visual attention model," *Pattern Recognition*, vol. 48, no. 8, pp. 2554–2566, 2015.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 886–893, San Diego, CL, USA, June 2005.

[21] B. Hong and S. Soatto, "Shape matching using multiscale integral invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 1, pp. 151–160, 2015.

[22] M. Clement, C. Kurtz, and L. Wendling, "Learning spatial relations and shapes for structural object description and scene recognition," *Pattern Recognition*, vol. 84, pp. 197–210, 2018.

[23] J. Žunić, P. L. Rosin, and V. Ilić, "Disconnectedness: a new moment invariant for multi-component shapes," *Pattern Recognition*, vol. 78, pp. 91–102, 2018.

[24] Z. Liu, Z. Lai, W. Ou et al., "Structured optimal graph based sparse feature extraction for semi-supervised learning," *Signal Processing*, vol. 170, 2020.

[25] G. Malu, S. Elizabeth, and S. M. Koshy, "Circular mesh-based shape and margin descriptor for object detection," *Pattern Recognition*, vol. 84, pp. 97–111, 2018.

[26] S. Jabeen, Z. Mehmood, T. Mahmood, and T. Saba, "An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model," *PLoS ONE*, vol. 13, no. 4, Article ID e0194526, 2018.

[27] Z. Mehmood, F. Abbas, T. Mahmood et al., "Content-based image retrieval based on visual words fusion versus features fusion of local and global features," *Arabian Journal for Science and Engineering43*, vol. 18, pp. 7265–7284, 2018.

[28] U. Sharif, Z. Mehmood, T. Mahmood et al., "Scene analysis and search using local features and support vector machine for effective content-based image retrieval," *Artificial Intelligence Review*, vol. 52, pp. 901–925, 2019.

[29] Z. Mehmood, "Effect of complementary visual words versus complementary features on clustering for effective content-based image search," *Journal of Intelligent & Fuzzy Systems*, vol. 35, no. 5, pp. 5421–5434, 2018.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, 2012.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the International Conference on Learning Representations*, pp. 1–14, San Diego, CA, USA, May 2015.

[32] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architec-ture for weakly supervised place recognition," *Computer Vision and Pattern Recognition*, vol. 9, pp. 5297–5307, 2016.

[33] F. Radenovi ´c, G. Tolias, and O. Chum, "CNN image retrieval learns from BoW: un- supervised fine-tuning with hard examples," in *Proceedings of the European Conference on Computer Vision*, pp. 3–20, Amsterdam, The Netherlands, October 2016.

[34] 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.

[35] J. Kim and S.-E. Yoon, "Regional attention based deep feature for image retrieval," in *Proceedings of the 29th British Machine Vision Conference*, Newcastle, UK, September 2018.

[36] S. Pang, J. Ma, J. Xue, J. Zhu, and V. Ordonez, "Deep feature aggregation and image Re-ranking with heat diffusion for image retrieval," *IEEE Transactions on Multimedia*, vol. 21, no. 6, pp. 1513–1523, 2019.

[37] F. Radenović, 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.

[38] 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.

[39] G. M. Johnson and M. D. Fairchild, "A top down description of S-CIELAB and CIEDE2000," *Color Research and Application*, vol. 28, no. 6, pp. 425–435, 2003.

[40] C. Blakemore and F. W. Campbell, "On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images," *The Journal of Physiology*, vol. 203, no. 1, pp. 237–260, 1969.

[41] W. Burger and M. J. Burge, *Principles of Digital Image Processing: Core Algorithms*, Springer, Berlin, Germany, 2009.

[42] Convert from HSV to RGB Color Space: https://ww2.mathworks.cn/help/images/convert-from-hsv-to-rgb-color-space.html.

[43] S. Kan, Y. Cen, Z. He, Z. Zhang, L. Zhang, and Y. Wang, "Supervised deep feature embedding with handcrafted feature," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5809–5823, 2019.

[44] S.-C. Kan, Y.-G. Cen, Y. Cen et al., "SURF binarization and fast codebook construction for image retrieval," *Journal of Visual Communication and Image Representation*, vol. 49, pp. 104–114, 2017.