IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 23, 2020, accepted September 30, 2020, date of publication October 12, 2020,
date of current version October 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3030056

Color Image Retrieval Algorithm Fusing Color and Principal Curvatures Information

HAOXIANG ZHANG[1], MAN JIANG[1], AND QIQI KOU[2]
[1]School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China
[2]School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China

Corresponding author: Qiqi Kou (kouqiqi@cumt.edu.cn)

This work was supported by the Fundamental Research Funds for the Central Universities under Grant 2020QN49.

ABSTRACT Aiming at the problems of high dimensional features, poor viewpoint robustness and long retrieval time of existing algorithms in the image retrieval system, this paper presents a new image retrieval algorithm by integrating image color information and surface geometry principal curvatures information. In the proposed method, the color image is first quantized and counted to obtain its color histogram. Simultaneously, the Hessian matrix is used to extract the texture information and the joint histogram of oriented gradient with mix-sampling and multi-scale is constructed. And then, the obtained color histogram and histogram of oriented gradient are fused to obtain the final joint histogram. Experiments are performed on public datasets, and comparison and analysis with representative algorithms based on a single visual feature or set of visual features to verify the performance of our algorithm. The experimental results show that the proposed method has the advantages of low dimensionality, fastness, strong viewpoint robustness and high precision, and can realize image retrieval efficiently.

INDEX TERMS Image retrieval, color, principal curvatures, texture, joint histogram.

I. INTRODUCTION

With the application of computer vision and digital devices, image processing is more and more widely used in medical images [1], [2], image super-resolution [3], [4], remote sensing images [5], and other fields such as social applications [6], [7]. Nowadays, benefiting from the rapid development of image processing, people can get access to image resources quickly. However, the problem that people urgently need to solve in the information age is how to effectively organize and query those massive visual data. So, image retrieval plays a dynamic role in those applications. For example, images are closely related to the study of geographers. Therefore, by using image retrieval technology, the geographic image information captured by the satellite can be found easily. what's more, image retrieval technology can also be used for skin detection of pedestrians with high accuracy [8]. Thence, image retrieval technology has been attracting the attention of numerous researchers. Image retrieval technology originated in the 1970s, and at that time, the text-based image retrieval(TBIR) [9]–[12] was the main technology. However, the disadvantage is that it requires manual annotation. Thus, methods based TBIR are very time-consuming. So the CBIR [13]–[17] is beginning to attract the attention of researchers. CBIR is based on some features, such as color, texture. The approach extracts the objective attributes of the image itself, so it does not require the assistance of external knowledge [18]. All the time, researchers have conducted a number of studies and many methods have been proposed for CBIR, such as histogram of oriented gradients (HOG) [19]. Due to the outstanding performance in representing objects, HOG has been widely used in human detection and image retrieval. Agarwal *et al.* [20] applied HOG in CBIR and verified that HOG-based retrieval system can improve precision. Min *et al.* [21] developed a content-based image retrieval method which combines gradient magnitude distribution and statistics of gradient orientation and had relatively good performance. Hu *et al.* [22] proposed Gradient Field HOG (GF-HOG) as image feature descriptor which improved accuracy. Fu *et al.* [23] proposed BHoG descriptor for sketch-based image retrieval. The technology encoded principal gradient orientation into binary codes, which is proved to be discriminative. Based on HOG, Wu *et al.* [24] presented a new method called gradient histogram markov stationary features(GHMSF) by combining the HOG descriptor and MSF(markov stationary feature), which make the

The associate editor coordinating the review of this manuscript and approving it for publication was Joewono Widjaja.

image retrieval more effectively. Joolee *et al.* [25] proposed Ternary Histogram of Oriented Gradient (THOG) descriptor for video retrieval and produced competitive accuracy. As can be seen, HOG-based descriptors has been successfully in this retrieval area. Besides, color features are of vital importance in visual features and widely used in image retrieval. Although color features ignore local spatial information, it still has strong stability. Due to the advantage of color features, many algorithms often combine color features with other features. Walia *et al.* [26] presented a retrieval approach by combination of color, shape and texture information which utilized two-stage strategy to identify the closest matches. Tajeripour *et al* [27] proposed a CBIR approach based on fusion of color and texture features. The algorithm first used object cropping algorithm to preprocess the image. And then, by taking color local binary patterns and local variance into consideration as operators, the algorithm has further discriminant information. Dubey *et al.* [28] introduced multi-channel LBP that is extracted from color images, and then, multi-channel LBP combined with two modes of adder and decoder. By using all components of color space Hor *et al.* [29] presented a new image retrieval method which is generated by two different texture descriptors and used the two descriptors to extract texture information with high accuracy. By combining local and global features, Bani and Fekri-Ershad [30] proposed a new algorithm. In this approach, the quantized color histogram was employed to extract the global color information in spatial domain. In the frequency domain, the texture information is extracted using the Gabor filter bank. In paper [31], Alsmadi used neutrosophic clustering algorithm and canny edge method to extract shape features. And then, discrete wavelet transform and canny edge histogram were employed to extract color features and gray-level co-occurrence matrix(GLCM) was applied to extract texture features. By combing above features, the image retrieval framework performance is increased. In [32], GLCM pixel difference and texture features were extracted as retrieval features for target retrieval. By establishing the consolidation of color and texture features, paper [33] used color quantization information, spatial texture information and shape information as image retrieval features, and this algorithm has been proven to achieve good image retrieval results. Ahmed *et al.* [34] proposed a new method which combined the local image features, spatial information in BoW architecture for image retrieval. This approach used keypoints detection strategy to improve precision. However, there still remain unsolved problems: 1) the robustness of the algorithm to viewpoint variance needs to be improved. 2) Making a tradeoff between accuracy and efficiency is still challenge. 3) High feature dimensionality. Thus, this paper proposed a new image retrieval algorithm called F-MHOGs that fuses the color information and geometric structure with mix-sampling and multi-scale principal curvatures(PCs) of the image. The key contributions of this paper as follows:

- Based on the PCs [35], [36] information, a descriptor called PC-WJHOG is proposed which holds macro- and micro-structure texture information.
- A new measure of conversion is proposed by analyzing the geometric curvature information of the image surface.
- In order to improve robustness of the algorithm, this paper further combines the color features and put forward mix-sampling and multi-scale algorithm.
- We present a moderate feature dimensionality which makes a tradeoff between accuracy and efficiency.

II. RELATED WORK

A. BRIEF REVIEW OF PCS

About the spatial structure characteristics of the PCs information, papers [37]–[39] have made in-depth research and have made progress. In three-dimensional Euclidean space, given a point p_0, its normal vector n and tangent plane are unique at a differentiable surface,and there is an infinite number of normal planes which contain normal vector n at point p_0. In all normal planes, there is a pair of normal planes that can make curvatures at the intersection between them and surface maximum or minimum in all the curvatures. These two curvature properties are the PCs which represent the extremum of normal curvature. Therefore, no matter the image rotates at any angle, PCs will remain unchanged [40], [41]. Besides, by using PCs, significant structures can be preserved while removing non-significant. As shown in Fig. 1, the maximum curvature and the minimum curvature are expressed by $\lambda_{\max}$ and $\lambda_{\min}$, respectively.

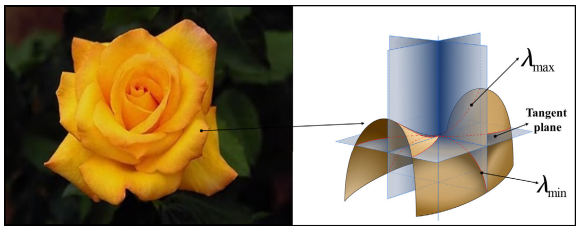

FIGURE 1. Structure diagram of principal curvatures on the image surface.

The PCs information of pixels in the image can be obtained by calculating the eigenvalues of the Hessian matrix[42]. Before calculating the PCs information, the determinant and trace of each pixel of color images are first computed by using the Hessian matrix. And then the curvature information of the surface geometric structure is extracted according to the eigenvalues. Finally, the characteristics of the image surface curvature information is converted.

For a gray image $I(x, y)$, the Hessian matrix H of a point on $I(x, y)$ is defined as follows:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} = \begin{bmatrix} (p \otimes I) \otimes d_2 & (p \otimes I') \otimes d_1 \\ (p \otimes I') \otimes d_1 & (d_2 \otimes I) \otimes p \end{bmatrix} \quad (1)$$

The $\otimes$ represents convolution. p, d_1, and d_2 are provide by paper[43]. Taking D_{xx} as an example, the $(p \otimes I) \otimes d_2$ indicates that a 1D convolution down the columns using interpolator p, and then, a 1D convolution along the rows using filter d_2. Similarly, $I' = (d_1 \otimes I) \otimes p$ indicates that I' can be obtain by a 1D convolution down the columns using d_1, along the rows using p. To acquire D_{xy}, our method first compute I', and use p and d_1 to convolve with I'. Fig. 2 shows the Hessian matrix at the point Q.

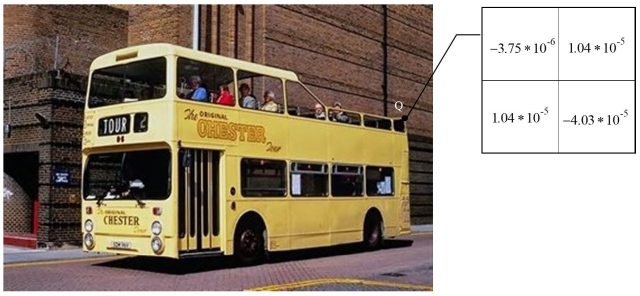

FIGURE 2. The hessian matrix of point Q.

The product of matrix eigenvalues is equal to the determinant of the matrix, and the sum of the matrix eigenvalues is equal to the trace of the matrix[44]. So two eigenvalues can be obtained by the determinant(det) and the trace(tr) of H:

$$\begin{aligned} trH &= \lambda_1 + \lambda_2 = D_{xx} + D_{yy} \\ \det H &= \lambda_1 \lambda_2 = D_{xx} D_{yy} - (D_{xy})^2 \end{aligned} \tag{2}$$

To find out eigenvalues, the operation is simplified by constructing $(\lambda - \lambda_1)(\lambda - \lambda_2) = 0$. According to the quadratic equation of one unknown:

$$\lambda_{1,2} = \frac{(\lambda_1 + \lambda_2) \pm \sqrt{(\lambda_1 + \lambda_2)^2 - 4\lambda_1 \lambda_2}}{2} \tag{3}$$

Setting the larger result indicates $\lambda_{\max}$ and another is set as $\lambda_{\min}$. By putting Eq. 2 into Eq. 3, the two eigenvalues can be obtained, which are shown in the Eq. 4:

$$\begin{aligned} \lambda_{\max} &= \frac{(D_{xx} + D_{yy}) + \sqrt{(D_{xx} - D_{yy})^2 + 4D_{xy}^2}}{2} \\ \lambda_{\min} &= \frac{(D_{xx} + D_{yy}) - \sqrt{(D_{xx} - D_{yy})^2 + 4D_{xy}^2}}{2} \end{aligned} \tag{4}$$

$\lambda_{\max}$ and $\lambda_{\min}$ describe the local extremum of the texture feature and reflect the macro and micro texture information of the image. Fig. 3 shows the results of the image represented by PCs with different Gaussian standard deviations.

In the process of extracting the Hessian matrix, it can be found that the matrix is composed of decimals. So some images are selected from the database to study, and most of these decimals hover between $-3.76{*}10^{-4}$ and $3.46{*}10^{-4}$, which are relatively small. Therefore, our method does not directly use those values. Inspired by the paper [37], this paper uses the Eq. 5 to transform PCs.

$$\xi = \frac{\lambda_{\max}}{\lambda_{\min}} \tag{5}$$

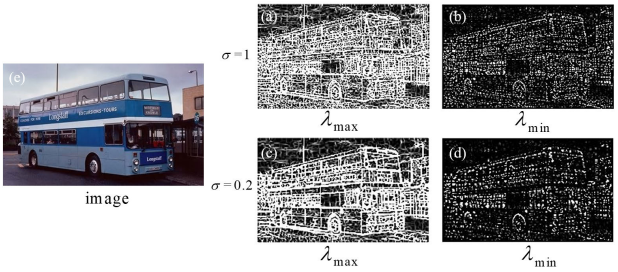

FIGURE 3. The results represented by principal curvatures.(a)(c) maximum curvature; (b)(d) minimal curvature; (e) original image.

However, ξ is still relatively small and fluctuates in a small range above and below zero. In order to improve the discrimination, the Eq. 6 is used to convert ξ:

$$\eta = \frac{2}{1 + \exp(-2{*}\xi)} - 1 \tag{6}$$

B. HOG

Since the HOG feature descriptor was proposed, it has been widely used in pedestrian detection area. HOG has good invariance to geometric and optical deformation of the image. HOG descriptor can be constructed by calculating and counting the gradient direction histograms of local areas of the image. The descriptor first divides the image into small cells, and calculates the gradient and of the pixels in each cell. And then, HOG combines multiple cells into a pixel block. The horizontal and vertical gradients of each pixel on gray image $I(x, y)$ can be calculated according to those following formulas:

$$G_x(x, y) = I(x + 1, y) - I(x - 1, y) \tag{7}$$

$$G_y(x, y) = I(x, y + 1) - I(x, y - 1) \tag{8}$$

where $G_x(x, y)$, $G_y(x, y)$ represent the gradients in the horizontal and vertical directions.

Next the gradient magnitude and gradient direction are calculated by Eq. 9 and Eq. 10:

$$\nabla G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{9}$$

$$\theta(x, y) = \tan^{-1}\left(\frac{G_y(x, y)}{G_x(x, y)}\right) \tag{10}$$

where $\nabla G(x, y)$, $\theta(x, y)$ represent gradient magnitude and gradient direction, respectively.

Following the previous methods, the gradient information of all pixels in cells can be extracted. According to a certain rule or shape, the cells are formed into larger blocks, and the descriptors of the cells in each block are concatenated to obtain the descriptor of the corresponding blocks. Finally, the feature descriptors of all blocks are connected to form HOG feature descriptor.

III. THE PROPOSED METHOD

The framework of the algorithm is shown in Fig. 4. As can be seen from Fig. 4, the F-MHOGs algorithm contains two parts:

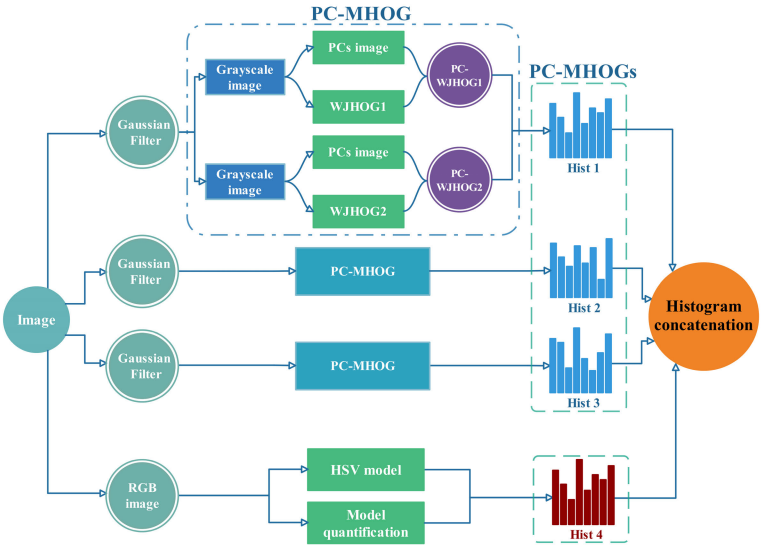

FIGURE 4. Framework of the proposed F-MHOGs algorithm.

histogram of color information, histogram of directional gradients based on PCs.

A. CONSTRUCTION OF COLOR HISTOGRAM

Before constructing the color histogram of the target image, image needs to be converted from RGB to HSV color space. And then, HSV model is quantized by Eq. 11 to present the distributions of the image:

$$h = \begin{cases} 0, & H \in [0, 20] \\ 0, & H \in (315, 360] \\ 1, & H \in (20, 40] \\ 2, & H \in (40, 75] \\ 3, & H \in (75, 155] \\ 4, & H \in (155, 190] \\ 5, & H \in (190, 270] \\ 6, & H \in (270, 295] \\ 7, & H \in (295, 315] \end{cases}, \quad s = \begin{cases} 0, & S \in [0, 0.2] \\ 1, & S \in (0.2, 0.7] \\ 2, & S \in (0.7, 1] \end{cases}$$

$$v = \begin{cases} 0, & V \in [0, 0.2] \\ 1, & V \in (0.2, 0.7] \\ 2, & V \in (0.7, 1] \end{cases} \tag{11}$$

where h, s and v are the quantization levels of each channel, respectively. After constructing color vectors for HSV, the following formula is used to construct color histogram $\mathbf{T}_{hsv}$:

$$\mathbf{T}_{hsv} = w_h * h + w_s * s + w_v * v \tag{12}$$

where w_h, w_s and w_v are weights of each channel which are 9, 3 and 1, respectively. So $\mathrm{T}_{hsv} \in [0, 1, 2, \cdots, 71]$. In fact, the last component of $\mathbf{T}_{hsv}$ is always zero, so our algorithms finally extracts 71 bins color features.

B. PC-MHOGS OPERATOR

As shown in Fig. 4, the PC-MHOGs descriptor extracts grayscale images of different scales through three Gaussian filters. In each scale space, the images are processed by using different sampling windows, and the PC-MHOG can be obtained in this way. Our operator uses 8*8 and 32*32 windows as sampling window in each scale space. Thus, the PC-WJHOGs can be generated in two sampling modes, called PC-WJHOG1 and PC-WJHOG2.

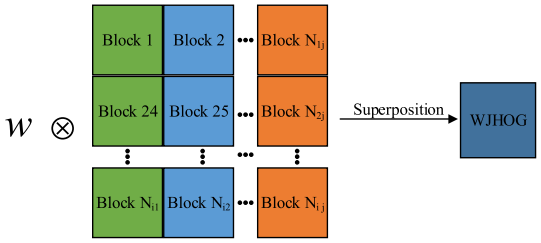

FIGURE 5. Generation of the WJHOG.

The calculation of PC-WJHOG based on PCs is very simple. In chapter II, the image of PCs information has been extracted, so the WJHOG can be used to extract the gradient and direction information of PCs for generating PC-WJHOG. the calculation method of WJHOG is as following. The Fig. 5 shows the generation of the WJHOG. In fact, WJHOG(WJHOG1 and WJHOG2) divides the image into small cells, each cell has 8*8 or 32*32 pixels. In the process of extracting features, WJHOG1 uses 8*8 sampling mode and WJHOG2 uses 32*32. No matter which mode is used, a cell always provides a n bins feature vector and a pixel block contains 2*2 cells. So each pixel block can obtain a 2*2*n bins feature vector. Different from HOG, WJHOG does not concatenate those pixel blocks information, but superimposes all pixel blocks into a pixel block, which contains all block's information. So it can be used to describe image features. This way makes the feature vector of the WJHOG descriptor still is 2*2*n bins. Additionally, the number of pixel blocks in horizontal and vertical directions can be got by the following

formula:

$$block_{hor} = \left[floor(\frac{image_{width} - blocksize}{stride}) + 1 \right]$$
$$block_{ver} = \left[floor(\frac{image_{height} - blocksize}{stride}) + 1 \right] \quad (13)$$

where $image_{width}$ and $image_{height}$ refer to the image width or height. $blocksize$ refers to pixel block's size, which contains 2*2 cells, stride is the step, and $floor$ indicates rounding down. $block_{hor}$ and $block_{ver}$ represent blocks of pixels in the horizontal and vertical directions, respectively. So the pixel blocks in a image($block_{num}$) can be obtained by this formula:

$$block_{num} = block_{hor} * block_{ver} \quad (14)$$

Now the PC-WJHOG1 and PC-WJHOG2 have been got. And next, we take a linear combination of the two sampling modes. In the experiment section, the reasons for choosing the sampling method will be given. Finally, the PC-MHOGs is structured by concatenating PC-MHOG in three scale space. In the following chapters, use $\mathbf{P}_{pc}$ to indicate PC-MHOGs descriptor.

C. FEATURE FUSION

Color information is a global feature, which reflects the surface properties of the image area and has strong stability. However, it cannot reflect the local information changes of the image[34]. The PCs information on the surface of the image belongs to the texture information of the image, Therefore, the histogram based on the color information ($\mathbf{T}_{hsv}$) and the histogram based on the texture information($\mathbf{P}_{pc}$) calculated by the PCs are concatenated to obtain the final image retrieval descriptor, called F-MHOGs.

IV. EXPERIMENT AND DISCUSSION

A. EXPERIMENTAL SETUP

In order to test and evaluate the performance of the image retrieval algorithm proposed, the algorithm is tested on two classic databases. Consistent with other works [32], [34], the top-20 image retrieval test are iterated randomly for 10 times. Eleven algorithms are compared on the corresponding databases and similarity measurement method is Euclidean distance which is shown in Eq. 16. All algorithms are evaluated from several aspects including AP, Recall, SFET, RT, mAP and Dimension. AP is average accuracy which is obtained for every category by calculating the average of 10 iterations in which each iteration contains top-20 retrieved images. mAP is mean Average Precision which is computed by all categories of AP. Recall is recall rate. SFET and RT are time cost for single feature extraction and time cost for retrieval, respectively.

$$AP = \frac{\sum_{i=1}^{n} P}{n} \quad (15)$$

where P is the precision for each category and n is the number of iterations which is 10 in our case. It means that the experiment is performed for 10 different test images for each category.

FIGURE 6. Example of each type of images in Corel 1-k dataset.

$$d_{F-MHOGs} = w_1 \sqrt{\sum_{i=1}^{Pc} (x_{pc}^i - y_{pc}^i)} + w_2 \sqrt{\sum_{j=1}^{T} (x_{hsv}^j - y_{hsv}^j)} \quad (16)$$

where the superscripted i and j represent feature vector of $\mathbf{P}_{pc}$ and $\mathbf{T}_{hsv}$, respectively. Pc and T represent feature dimensionality. x_{pc}^i and y_{pc}^i represent query image and test image in database, respectively. When calculating $d_{F-MHOGs}$, we set $w_1 = 0.45$ and $w_2 = 0.55$.

B. DATABASE

1) COREL-1K

In the Corel-1k [45] database, there are 10 kinds of color pictures, namely Africa, Sea, Bus, Building, Dinosaur, Elephant, Flower, Horse, Mountain, Food, as shown in Fig. 6. Each category of picture contains 100 samples, so there are 1000 pictures in the Corel-1k database. All the images are the same size of either 256 * 384 or 384 * 256 in this dataset. The imaging conditions of the database are complex, and there are obvious differences in the same category. What's more, there are strong similarities in the different categories. For these reasons, the database is usually used to test the retrieval accuracy and efficiency of algorithms.

2) COIL-100

The COIL-100[34] contains 100 images of different categories, and each image includes the imaging results of the same item under different viewing angles. The angle interval of each item imaging is 5°, so in the range of 0-360°, each item has a total of 72 imaging results under different viewing angles. The imaging results of the same object at different viewing angles are also different. In other words, there are viewpoint variance in this database. Due to this variance, it is challenging to perform accurate image retrieval on the COIL-100 database. Fig. 7 shows the imaging situation under two different viewpoints in 12 categories, namely, Tomato, Cat, White Cup, Frog, Pink Cup, Mud Pot, Jug, Fancy Feast, Rolaids, Car, Soft Edri, Statue, Herb-0x.

C. PARAMETRIC ANALYSIS

1) EXPERIMENTAL TESTS FOR SAMPLING PARAMETERS AND GAUSSIAN STANDARD DEVIATION

In this section, the experiment will focus on the selected parameters for sampling and Gaussian standard deviation.

TABLE 1. the parameter analysis at different standard deviations on corel-1k.

Sampling	Performance	Space1 $\sigma 1 = 0.2, 0.5, 1$			Space2 $\sigma 2 = 0.5, 1, 2$		
		9bins	12bins	15bins	9bins	12bins	15bins
(16,32)	mAP(%)	81.98	83.75	84.03	82.05	83.70	83.83
	SFET(s)	0.0483	0.0486	0.0487	0.0484	0.0487	0.0489
(8,16)	mAP(%)	82.38	83.55	84.00	82.15	83.50	83.70
	SFET(s)	0.0496	0.0497	0.0506	0.0509	0.0514	0.0518
(4,8)	mAP(%)	82.35	83.18	83.70	82.28	83.15	83.55
	SFET(s)	0.0520	0.0546	0.0549	0.0512	0.0538	0.0550
(8,32)	mAP(%)	82.28	83.70	**84.05**	82.20	83.55	83.83
	SFET(s)	0.0489	0.0493	0.0507	0.0494	0.0491	0.0493
(4,32)	mAP(%)	82.43	83.73	83.75	82.275	83.15	83.55
	SFET(s)	0.0519	0.0542	0.0545	0.0505	0.0534	0.0545

FIGURE 7. Example of each type of images in COIL-100 dataset.

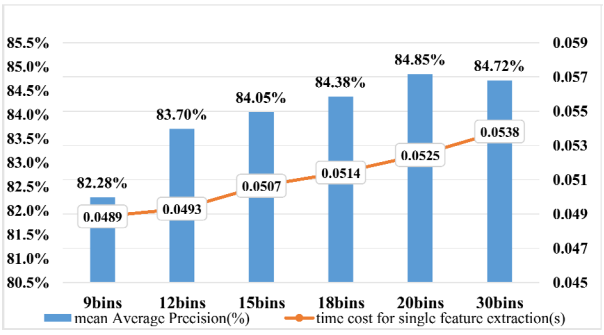

FIGURE 8. Mean Average Precision and time cost for single feature extraction of the top-20 under the sampling mode (8,32).

Specifically, the $\sigma 1 = 0.2, 0.5, 1$ and $\sigma 2 = 0.5, 1, 2$ are chosen as the scaling factors of scale space in our experimental test, respectively. Our algorithm is tested under scale space1 and scale space2. And 9bins, 12bins, and 15bins are extracted from each cell, respectively. The selected parameters are evaluated in terms of mAP and SFET. In addition, the five sampling modes are shown in Table 1. As the table shows, the mAP in space1 are generally better than mAP in space2. At the same time, the sampling mode(8,32) has the best performance in all modes, whose mAP is 84.05% in space1 and 83.83% in space2. If sampling radius is smaller, the extracted features will be more complete, but it will cost more time, such as (4,8). In addition, the mAP is related to the vector length, so the vector length of specified sampling mode will be further analyzed in next section. In fact, the time cost difference between each model is not significant. And it fluctuates between 0.048s and 0.054s.

2) THE ANALYSIS OF THE VECTOR LENGTH

In the last section, the results show that the sampling mode (8,32) has the best performance in listed sampling mode. Based on sampling mode (8,32), 9, 12, 15, 18, 20, and 30 that are divisor of 180 are selected as the experimental data. As shown in Fig. 8, the bar graph shows mAP of each vector length, and line chart shows time cost of extracting single image's features. It can be found that, as the length increases, time cost gets longer. Although the SFET is only 0.0489s and the vector length is 9, its mAP is not very high. Meanwhile, when the length is 20, the mAP is 84.85%, which is highest and the SFET is 0.0525s. The difference between them is about 4 milliseconds. However, after 20, the mAP starts to fall and the SFET is 0.0538s, which takes the longest time for feature extraction. Considering dimension and SFET, the 15bins and 20bins are selected as our main parameters for next sections.

D. COMPARISON WITH OTHER ALGORITHMS ON COREL-1K

After analyzing and determining the parameters of our algorithm, the proposed algorithm is compared with others. All algorithms are tested from six aspects, AP, Recall, SFET, RT, mAP and Dimension to evaluate performance. Each algorithm retrieved the top-20 image in the Corel-1k dataset. The experimental results are shown in Table 2 to Table 4.

1) RESULTS AND DISCUSSION FOR RETRIEVAL ACCURACY AND RECALL (%)

As can be seen from Table 2 and Table 3, compared with the algorithm proposed by Ahmed, our algorithm has promoted a lot in some test results. In the category of Bus, our AP and Recall are 99% and 19.8%, respectively. The work[34] is 75% and 15%, respectively. In the category of Flower, ours is 97.8% and Ahmed's algorithm is 90%. Compared with the method proposed by Pavithra, F-MHOGs has better retrieval results in most categories.

It is worth mentioning that in the category of Africa, the retrieval accuracy of the F-MHOGs are 93.8% which is

TABLE 2. Retrieval accuracy (%) of different algorithms on corel-1k dataset.

Methods	Africa	Sea	Building	Bus	Dinosaur	Elephant	Flower	Horse	Mountain	Food
Pavithra[33]	81	66	78.8	96.3	100	**70.8**	95.8	98.8	67.8	77.3
Niu [46]	83.2	46.2	72.2	97.8	99	63.4	95.4	92.4	47.8	89.5
Ahmed[34]	90	60	**90**	75	100	70	90	**100**	70	**90**
Sun[47]	33	22	40	23.3	71	27.5	50	59.2	26.7	65
Kundu[48]	44	32	52	62	40	80	57	75	57	56
Somnug[32]	67.6	59.8	58	94	99.8	58	88.6	93.8	47.8	49.2
Dubey[28]	75	55	67	95	97	63	93	89	45	70
Xiao[49]	67	60	56	96	98	53	93	82	46	58
Walia[26]	73	39.2	46.3	82.5	98	59.3	86	89.8	41.8	53
Badu[50]	56	53	61	89	98	57	89	78	51	69
Ours$^{1\,(15)}$	**93.8**	65.5	67.8	98.5	**100**	65	97.5	96.3	69.8	86.5
Ours$^{2\,(20)}$	**92.8**	**66**	70.5	**99**	**100**	67.5	**97.8**	96.3	**72**	87

TABLE 3. Recall rate (%) of different algorithms on corel-1k dataset.

Methods	Africa	Sea	Building	Bus	Dinosaur	Elephant	Flower	Horse	Mountain	Food
Pavithra[33]	16.2	13.2	15.8	19.3	20.0	**14.2**	19.2	19.8	13.6	15.5
Niu [46]	16.6	9.2	14.4	19.6	19.8	12.7	19.1	18.5	9.6	17.9
Ahmed[34]	18.0	12.0	**18.0**	15.0	20.0	14.0	18.0	**20.0**	14.0	**18.0**
Sun[47]	6.6	4.4	8.0	4.7	14.2	5.5	10.0	11.8	5.3	13.0
Kundu[48]	8.8	6.4	10.4	12.4	8.0	16.0	11.4	15.0	11.4	11.2
Somnug[32]	13.5	12.0	11.6	18.8	20.0	11.6	17.7	18.8	9.6	9.8
Dubey[28]	15.0	11.0	13.4	19.0	19.4	12.6	18.6	17.8	9.0	14.0
Xiao[49]	13.4	12.0	11.2	19.2	19.6	10.6	18.6	16.4	9.2	11.6
Walia[26]	14.6	7.8	9.3	16.5	19.6	11.9	17.2	18.0	8.4	10.6
Badu[50]	11.2	10.6	12.2	17.8	19.6	11.4	17.8	15.6	10.2	13.8
Ours$^{1\,(15)}$	**18.8**	13.1	13.6	19.7	20.0	13.0	19.5	19.3	14.0	17.3
Ours$^{2\,(20)}$	18.6	**13.2**	14.1	**19.8**	20.0	13.5	**19.6**	19.25	**14.4**	17.4

nearly 4% higher than the best current algorithm. Remember that HOG descriptor is an excellent detection algorithm, and it can represent objects appropriately [20], leading to great success in human detection. Moreover, our descriptor is HOG-based algorithm which also maintains the advantage of HOG. This is why our algorithm also performs with high precision in the category of Bus. Because the category of Dinosaur is dominated by regular texture information, our algorithm achieves remarkable performance in this category. Similarly, color features also play an important role in image retrieval. In the category of Flower, the color information has provided distinctive features, so some algorithms proposed by papers [33], [34], [46] all have good accuracy and recall in the category of Flower. Even so, in this category, our proposed descriptor still outperforms any descriptor. The reason is that PCs information captures the macro- and micro-structure texture information which can present discriminant features. In the category of Elephant, the performance of our descriptor do not have desirable accuracy slightly. The probable reason is that some images are created under exposure and algorithms are unable to extract persuasive image information under the imaging condition which leads to a failure to perform better results. But it can't ignore that the accuracy of our algorithm outperforms other algorithms in more than half of the categories. The striking performance of our descriptor demonstrates that the proposed approach is indeed an outstanding image retrieval algorithm.

2) DISCUSSIONS FOR PERFORMANCE OF DIFFERENT ALGORITHMS

Table 4 presents RT, SFET and dimension of various algorithms. As before, statistical computations for many times and then take the average. Remarkably, the total runtime in work ^{1}(**Ours$^{1\,(15)}$**) is only 0.093s. And whether our work^1 or work^2(**Ours$^{2\,(20)}$**), both are better than others in SFET and RT. Moreover, the total runtime of work^1 and work^2 does not beyond 0.1s. Although paper [34] has a competitive SFET, the retrieval stage is still time-consuming, because each keypoint needs to be matched. In addition, paper [26] utilized two-stage strategy to find out the closest matches which causes plenty of time to be consumed. Due to high feature dimensionality, the algorithm [28] and others spend more time to retrieve.

In chapter III, all pixel blocks are superimposed into a pixel block to compress the feature dimensionality when

TABLE 4. Comparison of the performance of different algorithms.

Methods	SFET	RT	Total	Dimension
Pavithra[33]	0.671	1.108	1.779	>768
Niu [46]	--	--	--	205
Ahmed[34]	0.142	1.190	1.332	--
Sun[47]	9.15	1.027	10.177	900
Kundu[48]	0.4	--	--	99
Somnug[32]	--	--	--	>4096
Dubey[28]	102.4	16.49	118.89	1024
Xiao[49]	--	--	--	>1240
Walia[26]	2.797	4.096	6.893	260
Badu[50]	--	--	--	**63**
Ours$^{1\ (15)}$	**0.051**	**0.042**	0.093	$251_{(71+3*60)}$
Ours$^{2\ (20)}$	0.052	0.045	0.097	$311_{(71+3*80)}$

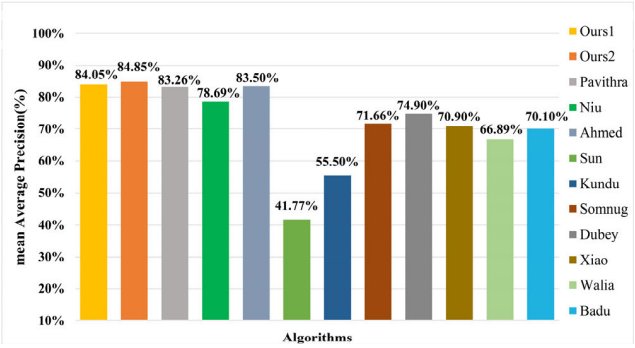

FIGURE 9. Mean average precision of different algorithms in corel-1k.

constructing PC-WJHOG, and the color information is quantified to keep feature dimensionality modest. As a consequence, our descriptor presents a moderate feature dimensionality which is 251 bins in total. In addition, profiting by template convolution, the efficiency of texture feature extraction is enhanced. Simultaneously, employing quantified color information can also reduce computation complexity. So, compared with the other algorithms, our method has the outstanding performance in total runtime and dimension. Another experimental results are shown in Fig. 9. As can be intuitively seen from the figure, our algorithm surpasses all algorithms in mAP which means there is more balanced performance in this dataset than other descriptors. Even in some categories that are difficult to retrieve, the proposed approach still achieves remarkable performance. Therefore, it can be concluded that our F-MHOGs has an excellent performance.

E. COMPARISON OF OUR ALGORITHM AND MAINSTREAM ALGORITHMS ON COIL

On the COIL-100 dataset, the F-MHOGs algorithm is compared with SIFT, HOG and other mainstream algorithms. The AP and mAP are shown in Table 5 and Fig. 10, respectively. The results show that the proposed F-MHOGs has the highest retrieval accuracy in most categories. Surprisingly, in the categories of Cat, Statue, Pink Cup, White, our method achieves 100%. Although algorithm[34] proposed by Ahmed *et al.* also combines color and texture features, keep in mind that there is viewpoint variance in this database. The features extracted by this algorithm is not against viewpoint

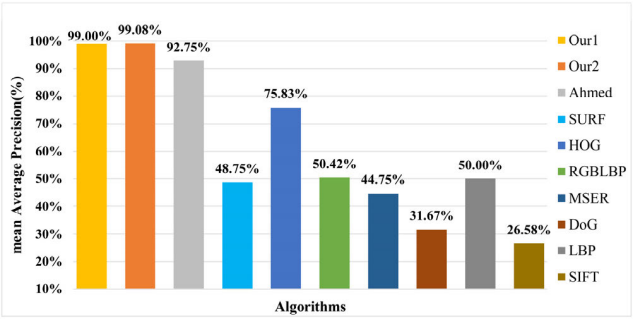

FIGURE 10. Mean average precision of different algorithms in COIL.

variance very well. By contrast, the PCs preserves notable texture features well. So by combing PCs, our algorithm can resist viewpoint variance effectively and has more accurate retrieval results. Notice that our method and HOG descriptor both have significant accuracy in the categories of White Cup, Rolaids and Fancy Feast. This verifies that HOG-based descriptors can make full use of gradient information to represent objects. However, in the category of Pink Cup, the proposed F-MHOGs performs better than the original HOG. The reason is that Pink Cup and White Cup have strong geometric similarity in some angles, HOG only focuses on the gradient information. Our algorithm not only focuses on the gradient information, but also employs PCs information and color features simultaneously. Similarly, even if Mud pot and statue have strong color similarity, our method can still distinguish them well. At the same time, the accuracy of SURF, MSER, etc. which are based on single feature are lower than method based on multi-feature such as our approach. It proves that retrieval algorithm based on multi-feature fusion has stronger discriminability.

However, it can be observed that images are not retrieved well in the category of Car. Seen from Fig. 10, there is non-rigid deformation in this category. The F-MHOGs is a descriptor proposed against viewpoint variance and it has limited capability to address non-rigid deformation. But the comparison of mAP in Fig. 10 still clearly shows that our algorithm has a more significant precision, which also indicates that proposed descriptor has more stable performance than others in the tested categories.

F. COMPUTATIONAL COMPLEXITY

In this paper, HSV color histogram and PC-WJHOG descriptor based on PCs are combined. First, RGB color space is converted to HSV color space. And then according to the color distribution of the image, each channel of the HSV model is quantized to reduce the complexity of the model. After quantization, the color feature dimensionality is only 71 bins. In the process of extracting PC-MHOGs, the eigenvalues of the Hessian matrix are calculated for PCs. Considering cost-effectiveness, convolution templates are used to speed up the calculation process. And all pixel blocks are superimposed into a pixel block to compress dimensionality. So, the dimensionality of PC-MHOGs is 180 bins. Hence, the

TABLE 5. Retrieval accuracy (%) of different algorithms on coil-100 dataset.

Methods	**Our**[1(15)]	Ahmed	SURF	HOG	RGBLBP	MSER	DoG	LBP	SIFT
Tomato	**98**	93	75	45	20	15	60	35	15
Car	93	**98**	65	65	55	22	15	60	22
Pink Cup	**100**	88	50	80	50	35	30	60	40
Cat	**100**	90	45	45	25	55	10	40	32
White Cup	**100**	94	40	100	50	60	25	25	45
Statue	**100**	100	30	65	55	45	10	25	35
Rolaids	**100**	65	60	100	85	40	50	65	20
Mud Pot	**100**	100	45	100	50	90	50	70	20
Herb-0x	**100**	90	45	100	65	25	20	40	25
Frog	**98**	95	65	65	45	45	45	55	20
Fancy Feast	**100**	100	20	45	45	35	40	60	25
Jug	99	**100**	45	100	60	70	25	65	20

final feature dimensionality of F-MHOGs is 251 bins (180 + 71). Furthermore, suppose a dataset contains N images and the size of each image is $w * h$. Our method uses the quantized color information, and the time complexity is $O(D * N * w * h)$. D is 3, which corresponds to number of color image channels. To capture the PCs information of images, the Hessian matrix of each pixel on N images are extracted. So the time complexity is $O(4 * N * w * h)$ in total. Next, the PC-WJHOG is obtained by computing the gradient magnitude and gradient direction of each image. The time complexity is $O(2 * N * w * h)$. And then, the PC-WJHOG is computed six times, and the time complexity is $O(6 * 2 * N * w * h)$. So the final time complexity of our approach is $O(N * w * h)$, which demonstrates that the retrieval time of the algorithm proposed in this article is much shorter than other existing methods.

V. CONCLUSION AND FUTURE WORK

In this paper, a new color image retrieval algorithm F-MHOGs is presented, which fuses color histogram and the weighted joint histogram of oriented gradient with mix-sampling and multi-scale. The algorithm is performed on two public databases, Corel-1k and COIL-100. Compared with the current popular algorithms, the proposed method has a lower feature dimensionality, leading to lower computational time. Furthermore, through the analysis of results on the COIL-100, the F-MHOGs has better robustness to the viewpoint variance, which performs more excellent accuracy than other works. As expected, the proposed method not only has high precision and strong robustness of viewpoint variance, but also is efficient.

For future work, we plan to explore the underlying relationship between images and extracted features to improve the robustness to overexposure and non-rigid deformation. Meanwhile, we will also study the texture and geometric features of the proposed descriptor and believe it can help us to probe the skin detection area.

REFERENCES

[1] A. Ahmed, "Implementing relevance feedback for content-based medical image retrieval," *IEEE Access*, vol. 8, pp. 79969–79976, 2020.

[2] Y. Cai, Y. Li, C. Qiu, J. Ma, and X. Gao, "Medical image retrieval based on convolutional neural network and supervised hashing," *IEEE Access*, vol. 7, pp. 51877–51885, 2019.

[3] L. Chen, Q. Kou, D. Cheng, and J. Yao, "Content-guided deep residual network for single image super-resolution," *Optik*, vol. 202, Feb. 2020, Art. no. 163678.

[4] Z. San-You, C. De-Qiang, J. Dai-Hong, K. Qi-Qi, and M. Lu, "Adaptive diagonal total-variation generative adversarial network for super-resolution imaging," *IEEE Access*, vol. 8, pp. 57517–57526, 2020.

[5] W. Jing, D. Zhang, and H. Song, "An application of ternary hash retrieval method for remote sensing images in panoramic video," *IEEE Access*, vol. 8, pp. 140822–140830, 2020.

[6] Z. Zhang, F. Zhou, S. Qin, Q. Jia, and Z. Xu, "Privacy-preserving image retrieval and sharing in social multimedia applications," *IEEE Access*, vol. 8, pp. 66828–66838, 2020.

[7] H. Wang, Z. Xia, J. Fei, and F. Xiao, "An AES-based secure image retrieval scheme using random mapping and BOW in cloud computing," *IEEE Access*, vol. 8, pp. 61138–61147, 2020.

[8] S. Fekri Ershad, "An innovative skin detection approach using color based image retrieval technique," *Int. J. Multimedia Appl.*, vol. 4, no. 3, pp. 57–65, Jun. 2012.

[9] S. Unar, X. Wang, C. Zhang, and C. Wang, "Detected text-based image retrieval approach for textual images," *IET Image Process.*, vol. 13, no. 3, pp. 515–521, Feb. 2019.

[10] N. Samet, S. Hicsonmez, and F. Sener, "Creating image tags for text based image retrieval using additional corpora," in *Proc. 24th Signal Process. Commun. Appl. Conf. (SIU)*, May 2016, pp. 1321–1324.

[11] T. Ho and N. Ly, "A scene text-based image retrieval system," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2012, pp. 79–84.

[12] M. Zeng, B. Yao, Z. J. Wang, Y. Shen, F. Li, J. Zhang, H. Lin, and M. Guo, "CATIRI: An efficient method for content-and-text based image retrieval," *J. Comput. Sci. Technol.*, vol. 34, no. 2, pp. 287–304, Mar. 2019.

[13] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2874–2886, May 2012.

[14] J.-S. Li, I.-H. Liu, C.-J. Tsai, Z.-Y. Su, C.-F. Li, and C.-G. Liu, "Secure content-based image retrieval in the cloud with key confidentiality," *IEEE Access*, vol. 8, pp. 114940–114952, 2020.

[15] 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.

[16] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, "Content-based brain tumor retrieval for MR images using transfer learning," *IEEE Access*, vol. 7, pp. 17809–17822, 2019.

[17] M. Garg and G. Dhiman, "A novel content based image retrieval approach for classification using GLCM features and texture fused LBP variants," *Neural Comput. Appl.*, Jun. 2020, doi: 10.1007/s00521-020-05017-z.
[18] A. Savakis, R. Sharma, and M. Kumar, "Efficient eye detection using HOG-PCA descriptor," *Proc. SPIE*, vol. 9027, Mar. 2014, Art. no. 90270J.
[19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 886–893.
[20] M. Agarwal and R. P. Maheshwari, "HOG feature and vocabulary tree for content-based image retrieval," *Int. J. Signal Imag. Syst. Eng.*, vol. 3, no. 4, p. 246, 2010.
[21] J. Hyeong-Min and K. Jong-Hyo, "Content-based image retrieval using a HOG (Histograms of Oriented Gradients) based similarity scoring," in *Proc. Korean Electron. Eng. Assoc. Conf.*, 2011, pp. 876–879.
[22] R. Hu, M. Barnard, and J. Collomosse, "Gradient field descriptor for sketch based retrieval and localization," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 1025–1028.
[23] H. Fu, H. Zhao, X. Kong, and X. Zhang, "BHoG: Binary descriptor for sketch-based image retrieval," *Multimedia Syst.*, vol. 22, no. 1, pp. 127–136, Feb. 2016.
[24] Q. Wu, H. Li, J. Niu, and Y. Wang, "Gradient histogram Markov stationary features for image retrieval," in *Proc. IEEE Conf. Anthol.*, Jan. 2013, pp. 1–5.
[25] J. Joolee and Y.-K. Lee, "Video retrieval based on image queries using THOG for augmented reality environments," in *Proc. IEEE Int. Conf. Big Data Smart Comput.*, Jan. 2018, pp. 557–560.
[26] E. Walia, S. Vesal, and A. Pal, "An effective and fast hybrid framework for color image retrieval," *Sens. Imag.*, vol. 15, no. 1, p. 93, Nov. 2014.
[27] F. Tajeripour, M. Saberi, and S. F. Ershad, "Developing a novel approach for content based image retrieval using modified local binary patterns and morphological transform," *Int. Arab J. Inf. Technol.*, vol. 12, no. 6, pp. 574–581, 2015.
[28] S. R. Dubey, S. K. Singh, and R. K. Singh, "Multichannel decoded local binary patterns for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4018–4032, Sep. 2016.
[29] N. Hor and S. F. Ershad, "Image retrieval approach based on local texture information derived from predefined patterns and spatial domain information," 2019, *arXiv:1912.12978*. [Online]. Available: https://arxiv.org/abs/1912.12978
[30] N. Tadi Bani and S. Fekri-Ershad, "Content-based image retrieval based on combination of texture and colour information extracted in spatial and frequency domains," *Electron. Library*, vol. 37, no. 4, pp. 650–666, Aug. 2019.
[31] M. K. Alsmadi, "Content-based image retrieval using color, shape and texture descriptors and features," *Arabian J. Sci. Eng.*, vol. 45, no. 4, pp. 3317–3330, Apr. 2020.
[32] S. Somnugpong and K. Khiewwan, "Content-based image retrieval using a combination of color correlograms and edge direction histogram," in *Proc. 13th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, Jul. 2016, pp. 1–5.
[33] L. K. Pavithra and T. S. Sharmila, "An efficient framework for image retrieval using color, texture and edge features," *Comput. Electr. Eng.*, vol. 70, pp. 580–593, Aug. 2018.
[34] K. T. Ahmed, S. Ummesafi, and A. Iqbal, "Content based image retrieval using image features information fusion," *Inf. Fusion*, vol. 51, pp. 76–99, Nov. 2019.
[35] M. P. Do Carmo, *Differential Geometry of Curves and Surfaces*. Upper Saddle River, NJ, USA: Prentice-Hall, 1976.
[36] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scales," *Image Vis. Comput.*, vol. 10, no. 8, pp. 557–564, Oct. 1992.
[37] Q. Kou, D. Cheng, H. Zhuang, and R. Gao, "Cross-complementary local binary pattern for robust texture classification," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 129–133, Jan. 2019.
[38] Q. Kou, D. Cheng, L. Chen, and Y. Zhuang, "Principal curvatures based local binary pattern for rotation invariant texture classification," *Optik*, vol. 193, Sep. 2019, Art. no. 162999.
[39] Q. Kou, D. Cheng, L. Chen, and K. Zhao, "A multiresolution gray-scale and rotation invariant descriptor for texture classification," *IEEE Access*, vol. 6, pp. 30691–30701, 2018.
[40] J. Zhang, H. Zhao, and J. Liang, "Continuous rotation invariant local descriptors for texton dictionary-based texture classification," *Comput. Vis. Image Understand.*, vol. 117, no. 1, pp. 56–75, Jan. 2013.
[41] M. Mellor, B.-W. Hong, and M. Brady, "Locally rotation, contrast, and scale invariant descriptors for texture analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 1, pp. 52–61, Jan. 2008.
[42] R. Lakemond, S. Sridharan, and C. Fookes, "Hessian-based affine adaptation of salient local image features," *J. Math. Imag. Vis.*, vol. 44, no. 2, pp. 150–167, Oct. 2012.
[43] H. Farid and E. P. Simoncelli, "Differentiation of discrete multidimensional signals," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 496–508, Apr. 2004.
[44] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. Philadelphia, PA, USA: SIAM, 1994.
[45] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 9, pp. 947–963, Sep. 2001.
[46] D. Niu, X. Zhao, X. Lin, and C. Zhang, "A novel image retrieval method based on multi-features fusion," *Signal Process., Image Commun.*, vol. 87, Sep. 2020, Art. no. 115911.
[47] S. Qiping, "Deep learning based image retrieval," *J. Jing De Zhen Univ.*, vol. 33, no. 3, pp. 21–24, 2018.
[48] M. K. Kundu, M. Chowdhury, and S. R. Bulò, "A graph-based relevance feedback mechanism in content-based image retrieval," *Knowl.-Based Syst.*, vol. 73, pp. 254–264, Jan. 2015.
[49] Y. Xiao, J. Wu, and J. Yuan, "MCENTRIST: A multi-channel feature generation mechanism for scene categorization," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 823–836, Feb. 2014.
[50] M. B. Rao, B. P. Rao, and A. Govardhan, "CTDCIRS: Content based image retrieval system based on dominant color and texture features," *Int. J. Comput. Appl.*, vol. 18, no. 6, pp. 40–46, Mar. 2011.

HAOXIANG ZHANG is currently pursuing the master's degree with the China University of Mining and Technology. His research interests include image processing, computer vision, and pattern recognition.

MAN JIANG is currently pursuing the master's degree with the China University of Mining and Technology. Her research interests include computer vision and pattern recognition.

QIQI KOU received the B.S. and M.S. degrees from the Anhui University of Science and Technology, in 2012 and 2015, respectively, and the Ph.D. degree from the School of Information and Control Engineering, China University of Mining and Technology, in 2019. He is currently a Lecturer with the School of Computer Science and Technology, China University of Mining and Technology. His research interests include image processing, computer vision, and pattern recognition.