ORIGINAL ARTICLE

Content-based image retrieval using a fusion of global and local features

Hee Hyung Bu[1] | Nam Chul Kim[2] | Sung Ho Kim[1]

[1]School of Computer Science and Engineering, Kyungpook National University, Daegu, Republic of Korea

[2]School of Electronics Engineering, Kyungpook National University, Daegu, Republic of Korea

Correspondence
Hee Hyung Bu, School of Computer Science and Engineering, Kyungpook National University. Daegu, Republic of Korea.
Email: hhbu@knu.ac.kr

Abstract

Color, texture, and shape act as important information for images in human recognition. For content-based image retrieval, many studies have combined color, texture, and shape features to improve the retrieval performance. However, there have not been many powerful methods for combining all color, texture, and shape features. This study proposes a content-based image retrieval method that uses the combined local and global features of color, texture, and shape. The color features are extracted from the color autocorrelogram; the texture features are extracted from the magnitude of a complete local binary pattern and the Gabor local correlation revealing local image characteristics; and the shape features are extracted from singular value decomposition that reflects global image characteristics. In this work, an experiment is performed to compare the proposed method with those that use our partial features and some existing techniques. The results show an average precision that is 19.60% higher than those of existing methods and 9.09% higher than those of recent ones. In conclusion, our proposed method is superior over other methods in terms of retrieval performance.

KEYWORDS
CLBP magnitude, color autocorrelogram, content-based image retrieval, Gabor local correlation, SVD

1 | INTRODUCTION

Recently, the quantity of multimedia has dramatically increased with the evolution of computer systems and networks. Multimedia is often used as a tool for increasing the communication efficiency. Image utilization is particularly increasing due to the development of mobile phone cameras and image-editing technologies. The retrieval of similar images in a large amount of images is provided as an internet service to users. It can also be used in online art galleries and shopping malls and digital studios.

Content-based image retrieval is the process of extracting objective feature information representing images, computing the similarity between the features extracted from a query image and an image in a database, and retrieving images in order of high similarity. A content-based image retrieval system is more useful when retrieving images without specific keywords, those outside the radius of our knowledge, or those that exist

on foreign-language websites that use languages we do not know.

The features used in content-based image retrieval are usually classified as texture, color, and shape features. These are low-level features; hence, they may be needed not only in image retrieval but also in various fields, such as image recognition and classification, computer vision, and deep learning. For content-based image retrieval, early studies have used the features extracted from a single class. Lately, however, a combination of features extracted from two or three classes has been applied.

The methods used for the color feature extraction include the extract histogram [1], color autocorrelogram [2], color structure descriptor (CSD) [3], scalable color descriptor (SCD) [3], color moment [4], and dominant color descriptor (DCD) [3]. The histogram [1] is global statistics that possess information of the global color distribution and the scale- and rotation-invariant characteristics. However, it does not exhibit a high retrieval performance due to lack of local information. In space, the color autocorrelogram [2] contains the color distance information, while the CSD [3] has the local color structure information, enabling them to complement the disadvantage of the histogram. The SCD [3] uses Haar transform coefficients and is a compact descriptor composed of 63 bits. Color moments [4] are scale and rotation invariant. The two color moments are an image's average and standard deviation. The DCD [3] uses the dominant colors of an image in the hue max min diff (HMMD) color space.

The methods used for the texture feature extraction include the edge histogram [3], gray level co-occurrence matrix (GLCM) [5], Gabor transformation [6], wavelet transformation [7], multi-resolution multi-direction (MRMD) transformation [8], local binary pattern (LBP) [9], rotation-invariant uniform LBP (RULBP) [9], completed LBP (CLBP) [10], and so on. The edge histogram has a local-edge distribution in an image, where the edges are captured at the vertical, horizontal, and diagonal directions ($45°$ and $135°$), and is non-direction specific. Meanwhile, the GLCM [5] considers the spatial relationship of pixels used to extract statistics features like contrast, correlation, energy, and homogeneity. Gabor transformation [6] has multi-resolution and -direction characteristics. It performs an optimal measurement of the local spatial frequencies and is usually employed for image analysis, texture classification, image processing, and face recognition. Wavelet transformation [7] is used to extract local spectral and temporal information as decomposing a signal by using wavelets of varying scale and location. In two dimensions, discrete wavelet transformation is used to extract the geometric moment of normalized coefficients. We previously proposed the

MRMD [8] method that is useful for the MRMD texture analysis. The LBP [9] considers the subtractions of the center pixel from the circularly symmetric neighborhood. It derives an RULBP [9] having at most two bitwise transitions. The CLBP [10] is used to generalize the LBP, such that a pixel is expressed as the three binary maps of CLBP-Center, CLBP-Sign, and CLBP-Magnitude.

The methods used for the shape feature extraction include the Zernike moments [11] and moment invariants [12, 13] based on area and the Fourier descriptor [14] and chain code [15] based on contour. These are especially useful in images with shapes (e.g., logo, symbol, and icon). Selvan and Ramakrishnan [16] presented a new trial using singular value decomposition (SVD) to obtain shape or structural features. They considered singular values to have a probability density function (PDF) and estimated the PDF slope. However, we regard singular values herein to have the form of deterministic exponential functions in terms of the index order and measure the slopes of their exponential functions related to the global structure.

Recent methods include multichannel decoded LBPs [17], directional local extrema patterns [18], and combined modified color histogram and diagonally symmetric co-occurrence texture pattern [19], among many others. Multichannel decoded local binary patterns [17] are introduced as two adder- and decoder-based schemas for the LBP combination from more than one channel. The joint information of each channel is captured in each of the output channels of the adder and the decoder before the histogram computation. Directional local extrema patterns [18] consider encoding the spatial relation between any pair of neighbors in a local region for the directional edge information in the $0°$, $45°$, $90°$, and $135°$ directions in an image. The combination of modified color and texture information [19] is proposed as a texture descriptor using a diagonally symmetric local binary co-occurrence pattern and a color descriptor considering the inter-channel relationship between the H and S channels of an image. The method proposed in Ahmed and Mohamed [20] conducts strategies of early and late fusion. In early fusion, the method employs a combination of 18 color and 12 texture features. In late fusion, it employs three of the most common distance measures. The methods proposed in Anibou and others [7] and Dubey and others [21] are new methods of the local image feature description based on the local binary patterns extracted from multiple filtered images.

Many of the recent studies that used machine learning have been referred to Hameed and others [22] and Rout and others [23]. The method proposed in Kumar and Kumar [24] investigated the performance of the extreme gradient boosting classifier for two-dimensional

(2D) object recognition. That in Chhabra and others [25] considered three classifiers, namely, decision tree, random forest, and multilayer perceptrons. Meanwhile, the approach proposed in Hameed and others [26] uses a support vector machine for classification.

So far, the problem with image retrieval is that only a few methods have been proposed for combining well-matched color, texture, and shape features, such that the retrieval performance is at a desired level.

Our proposed feature extraction method is a combination of color, texture, and shape features. As an advantage, it uses low-level features with a high retrieval performance, which do not need a training process unlike deep learning. Low-level features may be used in deep learning. Furthermore, the extracted feature vector has a low dimension, thereby saving storage.

The contribution of our proposed content-based image retrieval is a fusion of color, texture, and shape features. The color feature is extracted from the color autocorrelogram. The texture feature is extracted from the uniform-magnitude local binary pattern (UMLBP) and Gabor local correlation. Lastly, the shape feature is extracted from SVD, in which the SVD feature belongs to the global feature, and others belong to the local feature. The proposed method has multi-resolution and multi-direction characteristics, while the SVD, UMLBP, and color autocorrelogram have rotation-invariant characteristics. The proposed method combines well-coordinated, low-level features and has excellent aspects of high retrieval performance and low dimension.

The remainder of this paper is structured as follows: Section 2 provides the preliminaries; Section 3 describes the entire content-based image retrieval system and the feature extraction method proposed herein; Section 4 discusses the experimental procedure and results; Section 5 provides a discussion; and Section 6 concludes this study.

2 | PRELIMINARIES

2.1 | Multi-resolution, multi-direction filter

We chose the MRMD filter set used in the previous paper [8] as a tool for extracting the MRMD features in this work. Given an image I, the general MRMD filter operation is expressed as follows [8]:

$$y_{r,\theta}(p) = \sum_{q \in Q} h_{r,\theta}(p) \cdot I(p-q), \qquad (1)$$

where $h_{r,\theta}(p)$ is a high-pass filter defined as

$$h_{r,\theta}(p) = \begin{cases} -1, & q=(0,\ 0) \\ 1, & q=2^{r-1}\cdot(\cos\theta,\ \sin\theta) \end{cases} \qquad (2)$$

and p is the pixel position. Symbols r and θ stand for resolution level and direction, respectively. Q stands for the support region of filtering. The filter $h_{r,\theta}$ consists of two tabs corresponding to the center pixel and the pixel of position $2^{r-1}\cdot(\cos\theta, \sin\theta)$. Resolution and direction are those of $r \in \{1, 2, \cdots, M\}$ and $\theta = 2\cdot\pi\cdot n/N$, $n \in \{0, 1, 2, \cdots, N-1\}$, respectively, where M and N are the maximum of the resolution levels and the number of directions, respectively.

2.2 | Singular value decomposition

SVD [16, 27] is widely used in image processing and compression, computer vision, among others. Singular values stand for the diagonal elements of the diagonal matrix of SVD. SVD is explained below [28, 29].

Any $m \times n$ matrix $\mathbf{A}$ can be factored as

$$\mathbf{A} = U \cdot \Sigma \cdot V^{\mathrm{T}}, \qquad (3)$$

where U is $m \times m$ and orthogonal. V is $n \times n$ and orthogonal. U and V comprise singular vectors. Σ is $m \times n$, of which the principal diagonal elements are singular values of the A matrix in a descending order.

2.3 | Local binary pattern

The output of the LBP operator is obtained from a binary pattern of a local neighborhood with a threshold at the gray value of the center pixel that characterizes the spatial structure of the local image texture [9, 30].

$$\mathrm{LBP}_{P,R} = \sum_{p=0}^{P-1} s\left(g_p - g_c\right) \cdot 2^p, \ s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}, \qquad (4)$$

where P is the total number of equally spaced pixels on a circle and R is the neighborhood radius. Notations g_c and g_p are the gray values of the center pixel and the circularly symmetric neighborhood, respectively. If the g_c coordinate is $(0, 0)$, then the g_p coordinates are given by $(-R \cdot \sin(2\cdot\pi\cdot p/p), R \cdot \cos(2\cdot\pi\cdot p/p))$.

Operation U corresponds to the following number of spatial transitions (bitwise 0/1 changes) in the pattern [9]:

$$U(\text{LBP}_{P,R}) = \left| s\left(g_{p-1} - g_c\right) - s(g_0 - g_c)\right|$$
$$+ \sum_{p=1}^{P-1} \left| s\left(g_p - g_c\right) - s\left(g_{p-1} - g_c\right)\right|. \tag{5}$$

The following operator is for the uniform LBP (ULBP) with a U value of at most 2:

$$\text{LBP}_{P,R}^{\text{riu2}} = \begin{cases} \sum_{P=0}^{P-1} s(g_p - g_c), & \text{if } U(\text{LBP}_{P,\ R}) \le 2 \\ P+1, & \text{otherwise} \end{cases} \tag{6}$$

where superscript$^{\text{riu2}}$ of the LBP means the rotation-invariant "uniform" patterns with a U value of at most 2.

2.4 | Gabor wavelet

A 2D Gabor wavelet is a Gaussian function of a sine wave with frequency and direction. The following expression of the 2D Gabor wavelet is used herein [6, 31]:

$$g(x, y) = \frac{1}{2 \cdot \pi \cdot \sigma_x \cdot \sigma_y} \cdot \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) - 2 \cdot \pi \cdot j \cdot W \cdot x \right], \tag{7}$$

where $j = \sqrt{-1}$; x and y stand for the spatial locations of the kernel; and W stands for the modulation frequency. Symbols σ_x and σ_y denote the Gaussian function dilation.

A set of 2D Gabor wavelets presented as follows can be used to extract various frequency information in different resolutions and directions:

$$g_{s,n}(x, y) = a^{-2 \cdot s} \cdot g(x', y'), \tag{8}$$

where $a > 1$, $x' = a^{-s} \cdot (x \cdot \cos\theta_n + y \cdot \sin\theta_n) x' = a^{-s}$ $(x\cos\theta_n + y\sin\theta_n)$, and $y' = a^{-s} \cdot (-x \cdot \sin\theta_n + y \cdot \cos\theta_n)$. $\theta_n = n \cdot \pi / K$ for $s = 0, 1, ..., S-1$ and $n = 0, 1, ..., K-1$. Symbols S and K stand for the numbers of scales and directions, respectively.

3 | PROPOSED METHODOLOGY

We propose herein a fusion of local and global features for content-based image retrieval. The local features include those extracted from the UMLBP, Gabor local correlation, and color autocorrelogram, while the global features include those extracted from SVD. Figure 1 shows the block diagram of the proposed content-based image retrieval performing the following procedure: (1) An RGB image is converted to an HSV image; (2) the texture and color features of the HSV image are extracted; that is, the SVD and UMLBP features are extracted from the MRMD filtered domain; (3) the features extracted in Step 2 are combined; (4) the similarity for the feature vectors of query and database images is computed using the Mahalanobis distance
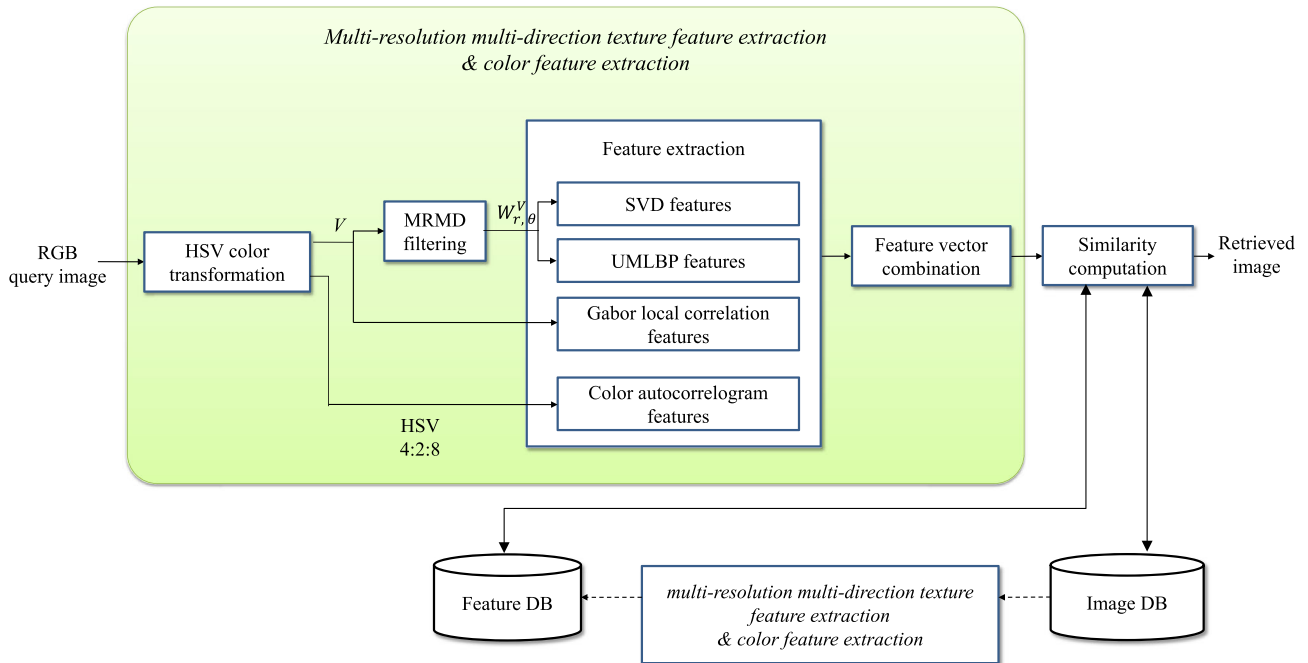

FIGURE 1 Block diagram of the proposed content-based image retrieval method

measure; and (5) the images are retrieved in order of higher rank.

3.1 | Singular value feature extraction based on multi-resolution, multi-direction filtering

The proposed method adopts SVD to measure the slope of singular values.

The process for the feature extraction of SVD is as follows:

Step 1. Convert the query image I to a gray image I_V.

Step 2. Calculate the absolute value $|y_{r,\theta}|$ of $y_{r,\theta}$ filtered from Expression (1) for image I_V.

Step 3. Conduct SVD using Expression (4). Singular values are limited to 0.001 because low singular values are often contaminated by noise. Singular values are approximated as a form of exponential functions [28]:

$$W_{r,\theta}(x) = a_{r,\theta}(x) \cdot e^{-\lambda \cdot x}, \tag{9}$$

where λ is the exponential function slope and x stands for the orders of singular values.

Step 4. Calculate the average over meaningful singular values, except for 1/4 of them in a descending order after the log operation. The expression is represented as follows:

$$W_{r,\theta}^{\mu} = \langle \ln[a_{r,\theta}(x)] \rangle - \lambda \cdot \langle x \rangle, \tag{10}$$

where symbol $\langle \cdot \rangle$ stands for the averaging operation over all xs and $a_{r,\theta}$ is related to energy. The result of (9) has the characteristics of the energy and slope of the singular values.

Step 5. Repeat steps 2 to 4 for all directions.

Step 6. Calculate the average and the standard deviation for the averages obtained in Step 4 in each level.

$$\mu_r = \operatorname*{mean}_{\theta \in \Theta} \left[W_{r,\theta}^{\mu} \right], \tag{11}$$

$$\sigma_r = \operatorname*{std}_{\theta \in \Theta} \left[W_{r,\theta}^{\mu} \right]. \tag{12}$$

The outcomes obtained from Step 6 are chosen as the features for the proposed image retrieval.

The feature vector has eight dimensions in the case of four resolution levels and is represented as follows:

$$f_{\mathrm{svd}} = [[\mu_r], [\sigma_r]]. \tag{13}$$

3.2 | Feature extraction using a uniform-magnitude local binary pattern

UMLBP [10] stands for the uniform magnitude of the local difference in a completed LBP (CLBP_M) [10]. Our previous study represented a local magnitude in a multi-resolution, multi-direction filtered domain. The magnitude local binary pattern (MLBP) should first be utilized for the UMLBP computation.

The MLBP expression is presented as follows [10, 32]:

$$\mathrm{MLBP}_{N,2^{r-1}}(p) = \sum_{n=0}^{N-1} t\big(|y_{r,\theta_n}(p)|, \mu_r(p)\big) \cdot 2^n,$$

$$\mu_r(p) = \operatorname*{mean}_{\theta_n \in \Theta} \big[|y_{r,\theta_n}(p)| \big], \ t(x, c) = \begin{cases} 1, & x \geq c \\ 0, & x < c \end{cases}. \tag{14}$$

where $|y_{r,\theta_n}(p)|$ stands for the absolute value of $y_{r,\theta_n}(p)$ and μ_r is evaluated by computing the $|y_{r,\theta_n}(p)|$ average.

The UMLBP expression is presented as

$$\mathrm{UMLBP}_{N,2^{r-1}}(p) =$$
$$\begin{cases} \sum_{n=0}^{N-1} t\big(|y_{r,\theta_n}(p)|, \ \mu_r(p)\big), & \text{if } U\big(\mathrm{MLBP}_{N,2^{r-1}}(p)\big) \leq 2 \\ N+1, & \text{otherwise} \end{cases}, \tag{15}$$

where U is the operator used to count the number of 1 s in the binary pattern of the output from (14).

Finally, the normalized histogram for the UMLBP from (15) is evaluated as follows in each resolution level:

$$H_r(i) = \frac{1}{|P|} \sum_{p \in P} \delta\big(\mathrm{UMLBP}_{N,2^{r-1}}(p) - i\big), \tag{16}$$

where δ stands for the Kronecker delta; $|P|$ is the image size; and $i \in \{0, 1, 2, \cdots, N, N+1\}$.

The UMLBP features are described as follows:

$$f_{\mathrm{UMLBP}} = \left[H_r^i \right]. \tag{17}$$

It denotes a vector composed of the normalized histogram of $\text{UMLBP}_{N,2^{r-1}}$. The UMLBP feature dimension is $M \cdot (N + 2)$.

3.3 | Feature extraction using the Gabor local correlation

As a Gabor feature, we adopt the Gabor local complex correlation coefficient used in our previous paper. The correlation expression for two images of I_A and I_B is as follows [32]:

$$
\begin{aligned}
&\underset{t \in T}{\text{COR}}[I_A(p), I_B(p)] = \\
&\frac{\underset{t \in T}{\text{mean}}\left[I_A(p) \cdot I_B{}^*(p) - \underset{t \in T}{\text{mean}}[I_A(p)] \cdot \underset{t \in T}{\text{mean}}[I_B{}^*(p)]\right]}{\left[\underset{t \in T}{\text{VAR}}[I_A(p)] \cdot \underset{t \in T}{\text{VAR}}[I_B{}^*(p)]\right]^{\frac{1}{2}}},
\end{aligned}
\tag{18}
$$

where COR, mean, and VAR stand for the correlation coefficient, mean, and variance operators, respectively. Symbols * and T stand for the complex conjugate and a 3×3 window, respectively.

The Gabor local correlation is expressed as follows:

$$
\rho_{s,n}(p) = \text{Re}\left\{\underset{t \in T}{\text{cor}}[J_{s,n}(p), \ J_{s,n}(p - s \cdot \delta\theta_n)]\right\},
\tag{19}
$$

where $J_{s,n}$ is the Gabor-transformed complex image with resolution s and direction n; $\text{Re}\{\cdot\}$ is the real part of the complex number and $\delta\theta_n = (\cos\theta_n, \sin\theta_n)$; and $\rho_{s,n}(p)$ stands for the correlation coefficient between two 3×3 windows with a center pixel p and $(p - s \cdot \delta\theta_n)$, respectively. The image composed of correlation coefficients is obtained using (19).

The features of the Gabor local correlation are finally obtained by computing the global average and the global standard deviation for the image derived in (19):

$$
\mu_{s,n}^\rho = \underset{p \in P}{\text{mean}}[\rho_{s,n}(p)],
\tag{20}
$$

$$
\sigma_{s,n}^\rho = \underset{p \in P}{\text{std}}[\rho_{s,n}(p)],
\tag{21}
$$

where *std* stands for the standard deviation operator.

The feature vector of the Gabor local correlation is composed of the global average vector $[\mu_{s,n}^\rho]$ and the global standard deviation vector $[\sigma_{s,n}^\rho]$:

$$
f_{\text{cor}} = \left[[\mu_{s,n}^\rho], [\sigma_{s,n}^\rho]\right].
\tag{22}
$$

The Gabor feature vector dimension is 32 when four resolution levels and upper four directions are used.

3.4 | Feature extraction using the color autocorrelogram

The color autocorrelogram [2] extracts spatial information for color. This method can effectively be used together with global information. The process for extracting the color autocorrelogram features is presented below.

Step 1. Convert the query image I to an HSV image I_{HSV}.
Step 2. Conduct quantization for the hue, saturation, and value components.
Step 3. Organize the three dimensions as follows:

$$
\begin{aligned}
I_{\text{QHSV}}(p) = &\ I_{\text{QH}}(p) \times \text{Saturation level} \times \text{Value level} \\
&+ I_{\text{QS}}(p) \times \text{Saturation level} + I_{\text{QV}}(p).
\end{aligned}
\tag{23}
$$

Step 4. For each center pixel, count the neighbor pixels whose distances from the center are k and whose values are similar to that of the center.

$$
\begin{aligned}
&\alpha_c^{(k)}(I_{\text{QHSV}}) = \\
&\Pr[I_{\text{QHSV}}(p) = I_{\text{QHSV}}(p') = c \,||\, p - p'| = k \text{ for } p, p' \in P],
\end{aligned}
\tag{24}
$$

where $\Pr[\cdot]$ stands for the probability satisfying the inner condition. Symbols c and k denote the quantized color value and the distance between p and p', respectively.

The feature vector of the color autocorrelogram derived from the (24) is expressed as follows:

$$
f_{\text{autocor}} = [\alpha_c].
\tag{25}
$$

The feature vector dimension is 64 when four, two, and eight histogram bins are used for H, S, and V, respectively.

4 | EXPERIMENT AND RESULTS

The proposed content-based image retrieval uses a fusion of the local and global features extracted from the Gabor local correlation, UMLBP, SVD, and color autocorrelogram.

This experiment involved four tests for the superiority of the proposed method. The first target for comparison was the partial features of the proposed method for the retrieval performance. The second target was the pairs of partial features. The third target was the existing methods that include the color histogram, SCD, CSD, RULBP, CLBP$_{RGB}$, and Autocorrelogram$_{RGB}$. The fourth target was the recent methods [17–21].

The databases used in this experiment were the Corel [33], Corel_MR, Corel_MD, VisTex [34], VisTex_MR, and VisTex_MD databases. The Corel database consists of 990 images with 192×128 resolutions. Each of the 90 images belonged to one of the 11 classes. Most images were of artificial objects.

The VisTex database is composed of 1200 images with 128×128 resolutions. Each of the 16 images belonged to one of the 75 classes. Most images were of homogeneous patterns.

TABLE 1 Description of the text datasets

Dataset	Sample image size	Number of classes	Number of samples per class	Remark
Corel	192×128	11	90	
Corel_MR	192×128 144×96 96×64	11	90	Three resolutions of the Corel image
Corel_MD	80×80	11	90	Eight directions of the Corel image
VisTex	128×128	75	16	
VisTex_MR	128×128 112×112 96×96	75	16	Three resolutions of theVisTex image
VisTex_MD	88×88	75	16	Eight directions of the VisTex image
Corel-1 K	384×256 256×384	10	100	

TABLE 2 Retrieval methods and their dimensions and color spaces

Method	Dimension	Color space	Remark
Color histogram [1]	128	RGB	
SCD [3]	128	HSV	
CSD [3]	128	HMMD	
ULBP [9]	40	V	
CLBP$_{RGB}$ [10]	186	RGB	
Autocorrelogram$_{RGB}$ [2]	216	RGB	6:6:6
UMLBP [10]	40	V	Four resolutions
SVD	8	V	Four resolutions
Autocorrelogram$_{HSV}$ [2]	64	HSV	4:2:8
Gabor correlation [32]	32	V	Four resolutions Eight directions
SVD + UMLBP	48 (8, 40)	V	
SVD + Gabor correlation	40 (8, 32)	V	
SVD + Autocorrelogram$_{HSV}$	72 (8, 64)	V, HSV	
UMLBP + Gabor correlation	72 (40, 32)	V	
UMLBP + Autocorrelogram$_{HSV}$	104 (40, 64)	V, HSV	
Gabor correlation + Autocorrelogram$_{HSV}$	96 (32, 64)	V, HSV	
Proposed	144	HSV	

The Corel_MR and VisTex_MR databases have multi-resolution images, where the Corel_MR database has 192×128, 144×96, and 96×64 resolutions, and the Vis-Tex_MR database has 192×128, 144×96, and 96×64 resolutions.

The Corel_MD and VisTex_MD databases have multi-direction images rotated by $0°$, $45°$, $90°$, $135°$, $180°$, $225°$, $270°$, and $315°$ in the Corel and VisTex databases, respectively.

The Corel-1K [35] database consists of 1000 images with 384×256 and 256×384 resolutions. Each of the 100 images belonged to one of the 10 classes. Most images were of artificial objects and nature. Table 1 provides the details of the seven databases.

Table 2 shows the methods, dimensions, and color spaces used in the experiment. The feature dimension of the proposed method was 144. The SVD among our partial methods used an OpenCV library function.

The distance measure used in this work is the Mahalanobis distance expressed as follows [36]:

$$D\left(f^q, f^d\right) = \left(\sum_{i=1}^{n} |\frac{f_i^q - f_i^d}{\sigma_i}|^M\right)^{\frac{1}{M}}, \quad (26)$$

where f_i^q is the ith component feature of the feature vector f^q extracted from the query image q; f_i^d is the ith component feature of the feature vector f^d extracted from the

database (DB) images; $|\cdot|$ is the absolute value; n is the feature vector dimension; σ_i is the standard deviation of the ith component features of the feature vectors in the entire feature DB; and M is the metric order.

The retrieval performance used in the experiment is precision versus recall [37]:

$$\text{precision} = \frac{|A(q) \cap R(q)|}{A(q)}, \quad (27)$$

$$\text{recall} = \frac{|A(q) \cap R(q)|}{R(q)}, \quad (28)$$

where $|\cdot|$ is the set size; q is a query image; $A(q)$ is the retrieved image set for the query image; and $R(q)$ is the relevant image set for the query image.

Figures 2–6 and Tables 3–5 present the experiment results. The marks on the graphs represent 10, 30, 50, 70, and 89 retrieved images on the Corel database and 5, 10, and 15 retrieved images on the VisTex database.

Figure 2A–C shows the precision versus recall for the proposed method and the methods of our partial features in the Corel, Corel_MR, and Corel_MD databases. The precision gains of our method in the Corel database were 12.2% to 39.44% against the Gabor local correlation, UMLBP, SVD, and color autocorrelogram, respectively.
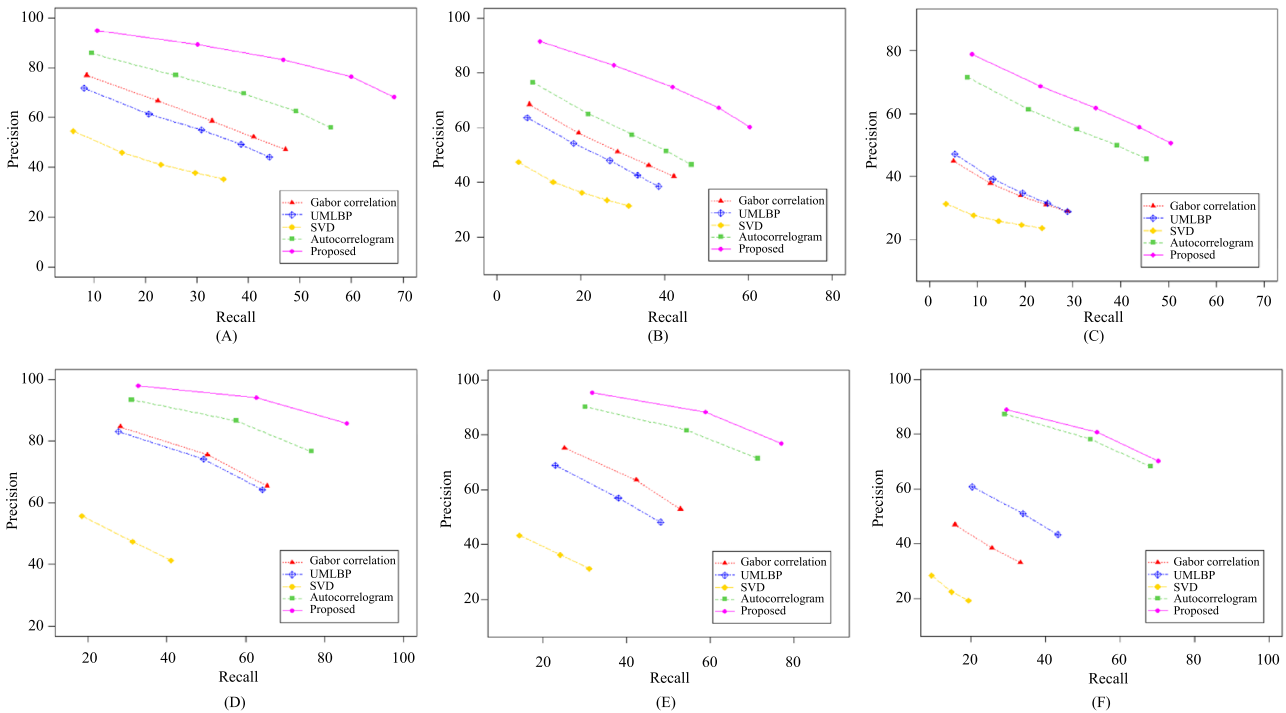

FIGURE 2 Precision versus recall for comparing the proposed method with the methods that used our partial features in the (A) Corel, (B) Corel_MR, (C) Corel_MD, (D) VisTex, (E) VisTex_MR, and (F) VisTex_MD databases
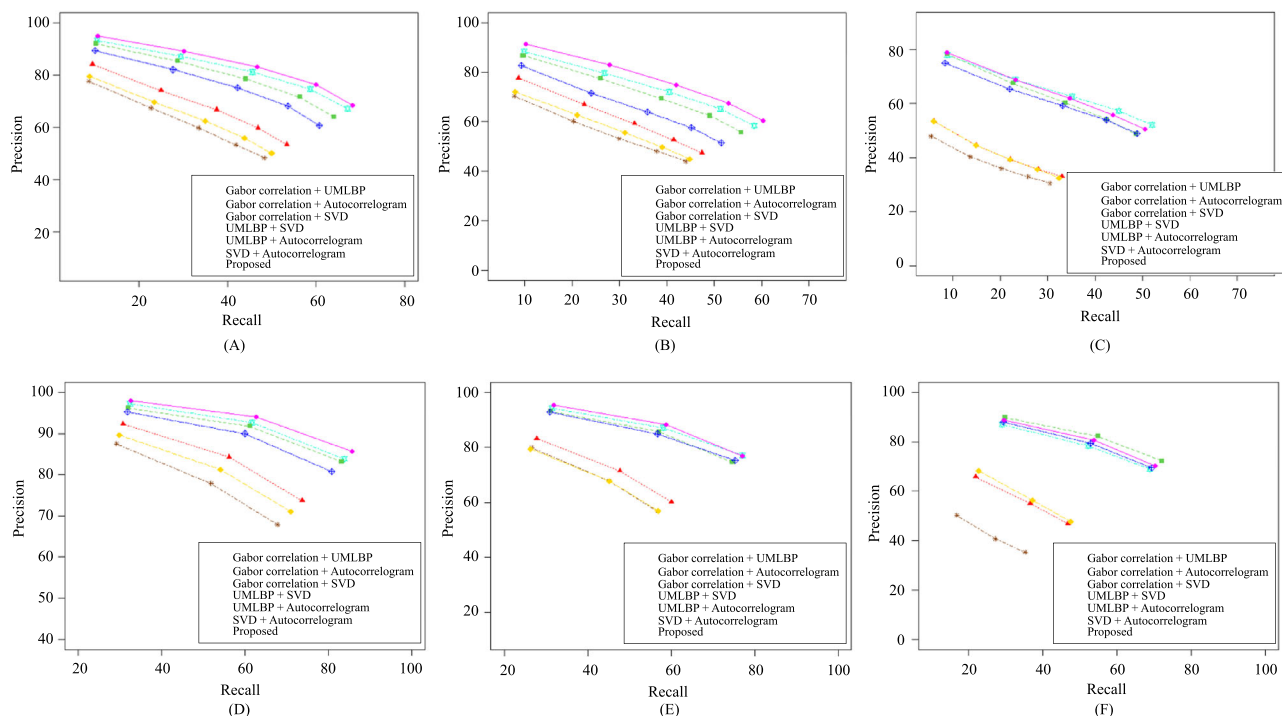
FIGURE 3 Precision versus recall for comparing the combined methods of the two partial methods of the proposed method and the proposed method in the (A) Corel, (B) Corel_MR, (C) Corel_MD, (D) VisTex, (E) VisTex_MR, and (F) VisTex_MD databases
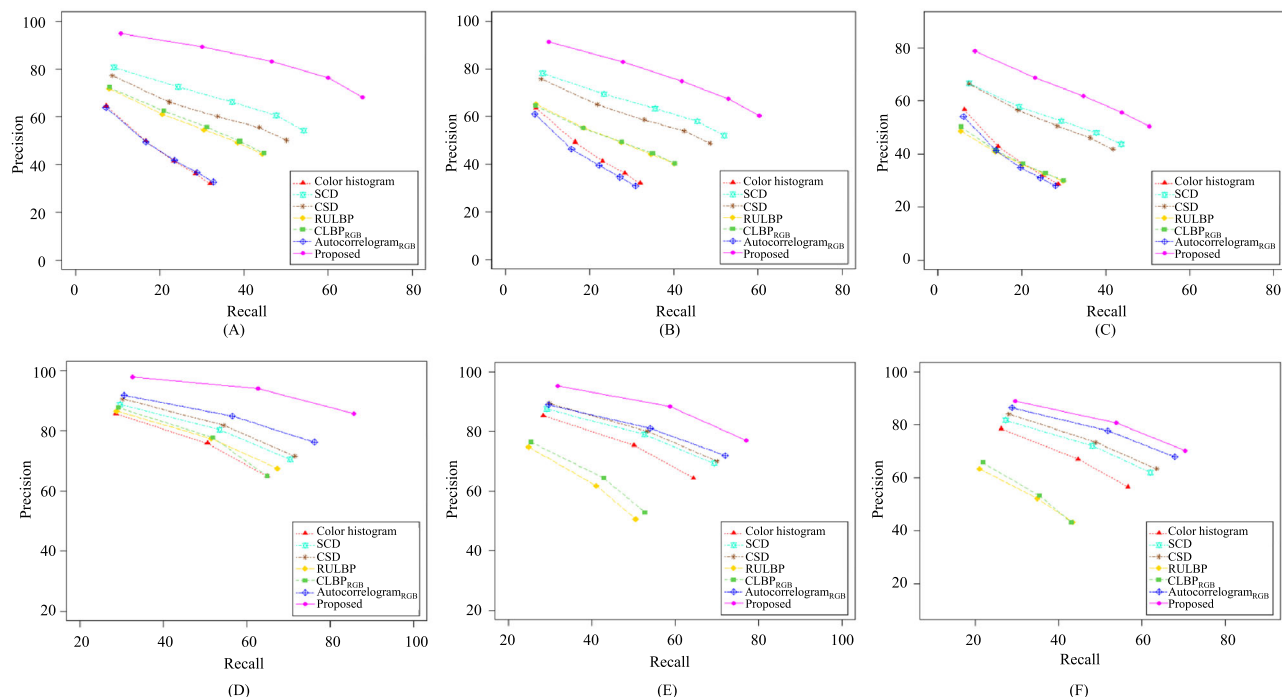

FIGURE 4 Precision versus recall for comparing the existing methods with the proposed method in the (A) Corel, (B) Corel_MR, (C) Corel_MD, (D) VisTex, (E) VisTex_MR, and (F) VisTex_MD databases

The gains on Corel_MR were 15.98% to 37.62%, while those on Corel_MD were 6.48% to 36.52%.

Figure 2D–F demonstrates the precision versus recall for the proposed method and the methods that utilized our partial features in the VisTex, VisTex_MR, and Vis-Tex_MD databases. The gains on the VisTex database were 7% to 44.5% over the Gabor local correlation, UMLBP, SVD, and color autocorrelogram, respectively.
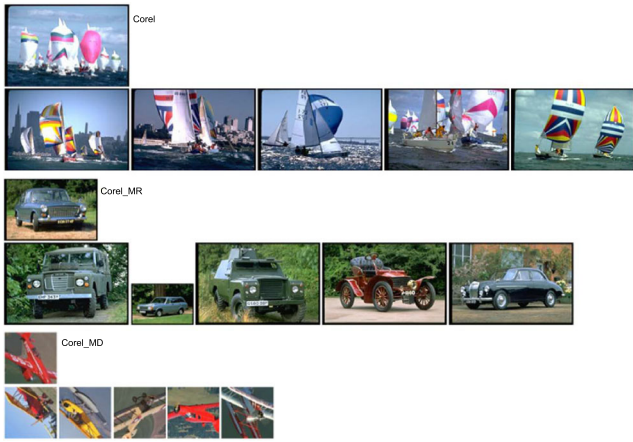
FIGURE 5 Examples of the query images and the top five retrieved images of the proposed method in the Corel, Corel_MR, and Corel_MD databases
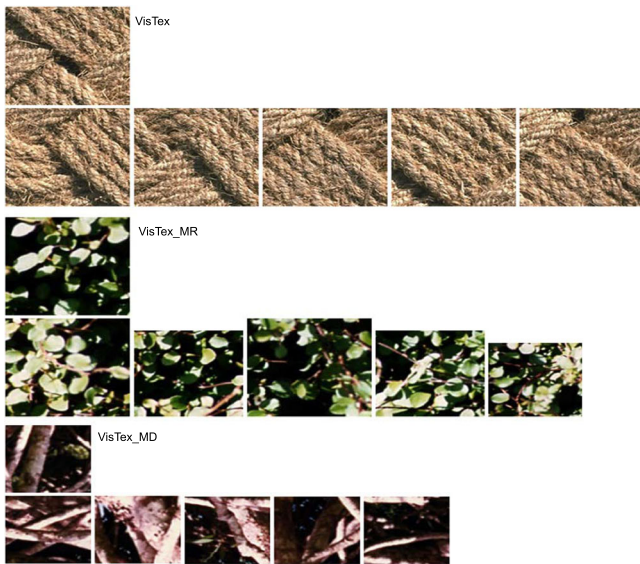

FIGURE 6 Examples of the query images and the top five retrieved images of the proposed method in the VisTex, VisTex_MR, and VisTex_MD databases

The gains on VisTex_MR were 5.77% to 50%, while those on VisTex_MD were 2.03 to 56.53%.

Figure 3A–C presents the precision versus recall for the proposed method and the combined two partial methods using our partial features in the Corel, Corel_MR, and Corel_MD databases. The precision gains of our method in the Corel database were 4.1% to 21.14% against the Gabor local correlation + UMLBP, Gabor local correlation + color autocorrelogram, Gabor local correlation + SVD, UMLBP + SVD, UMLBP + color autocorrelogram, and SVD + color autocorrelogram. The gains on Corel_MR were 5.06% to 20.2%, while those on

Corel_MD were −0.62% to 25.56%. Most showed high values.

Figure 3D–F displays the precision versus recall for the proposed method and the methods that used each pair of our partial features in the VisTex, VisTex_MR, and VisTex_MD databases. The precision gains of our method in the VisTex database were 1.3% to 14.8% against the Gabor local correlation + UMLBP, Gabor local correlation + color autocorrelogram, Gabor local correlation + SVD, UMLBP + SVD, UMLBP + color autocorrelogram, and SVD + color autocorrelogram. The gains on VisTex_MR were 0.7 to 18.87%, while those on VisTex_MD were −1.47 to 24.13%. The results depicted mostly high values.

Figure 4A–C illustrates the precision versus recall for the proposed method and the existing methods in the Corel, Corel_MR, and Corel_MD databases. The precision gains of our method in the Corel database were 15.54% to 37.54% against the color histogram, SCD, CSD, RULBP, CLBP$_{RGB}$, and Autocorrelogram$_{RGB}$. The gains on Corel_MR were 11.18% to 32.88%, while those on Corel_MD were 9.4 to 25.48%.

Figure 4D–F shows the precision versus recall for the proposed method and the existing methods in the VisTex, VisTex_MR, and VisTex_MD databases. The precision gains of our method in the VisTex database were 1.05% to 17.03% against the color histogram, SCD, CSD, RULBP, CLBP$_{RGB}$, and Autocorrelogram$_{RGB}$, respectively. The gains on VisTex_MR were 6.17% to 24.5%, while those on VisTex_MD were 6.3% to 26.93%.

Tables 3 and 4 present the precisions of the methods used in the experiment for the six databases for the top 10 retrieved images. In the Corel database, the precision of our method was 17.79% higher than the average of the precisions of the compared methods. The gains were 20% in Corel_MR, 21.01% in Corel_MD, 10.12% in VisTex, 13.89% in VisTex_MR, and 18.09% in VisTex_MD. Figure 5 depicts the top five retrieved images.

Table 5 presents the precision for the recent methods and the proposed method for the top 10 images retrieved in the Corel-1K database. The average gain of the proposed method was 9.09%, confirming that our method had a higher retrieval performance than the recent methods. The dimensions of the proposed feature vectors were all low, except for one.

The comparison in Table 5 did not include methods with prior information required learning.

5 | DISCUSSION

We proposed herein a method for extracting image features at a low level. The extracted features included color,

TABLE 3 Precision (%) of the tested methods in the Corel, Corel_MR, and Corel_MD databases for the top 10 retrieved images

DBs Methods	Corel	Corel_MR	Corel_MD
Gabor correlation	76.9	68.5	44.8
UMLBP	71.7	63.7	47.1
SVD	54.6	47.5	31.2
Color autocorrelogram	85.8	76.6	71.4
Gabor correlation + UMLBP	84.1	77.6	53.5
Gabor correlation + Color autocorrelogram	93.3	88.4	77.9
Gabor correlation + SVD	77.6	70.3	47.9
UMLBP + SVD	79.4	72.0	53.3
UMLBP + Color autocorrelogram	92.0	86.7	78.2
SVD + Color autocorrelogram	89.4	82.7	74.9
Color histogram	64.6	63.7	56.7
SCD	80.6	78.2	66.6
CSD	77.3	75.8	66.5
RULBP	71.7	65.2	48.6
$CLBP_{RGB}$	72.4	64.5	50.2
Color autocorrelogram$_{RGB}$	63.9	61.0	54.1
Proposed	95.0	91.4	78.7

TABLE 4 Precision (%) of the tested methods in the VisTex, VisTex _MR, and VisTex _MD databases for the top five retrieved images

DBs Methods	VisTex	VisTex _MR	VisTex _MD
Gabor correlation	84.6	75.2	47.0
UMLBP	83.1	68.8	60.8
SVD	55.7	43.2	28.5
Color autocorrelogram	93.4	90.3	87.3
Gabor correlation + UMLBP	92.4	83.2	65.8
Gabor correlation + Color autocorrelogram	97.2	94.2	87.0
Gabor correlation + SVD	87.5	79.9	50.3
UMLBP + SVD	89.6	79.3	68.3
UMLBP + Color autocorrelogram	96.2	93.2	89.8
SVD + Color autocorrelogram	95.3	92.8	88.1
Color histogram	85.8	85.3	78.5
SCD	88.7	87.6	81.8
CSD	90.6	89.5	84.1
RULBP	86.5	74.7	63.4
$CLBP_{RGB}$	87.7	76.4	65.8
Color autocorrelogram$_{RGB}$	91.8	89.0	86.4
Proposed	98.0	95.3	88.9

texture, and shape features. The color autocorrelogram was used to extract the color features. The magnitude of the complete LBP and the Gabor local correlation were used to extract the texture features. Lastly, SVD was used to extract the shape features. The SVD reflects the global structure of an image. The others showed the characteristics of local regions.

Figures 2 and 3 show performance comparisons of the proposed method and the methods that employed our partial features on six DBs. The retrieval performance of

TABLE 5 Comparison of the proposed method with some recent methods in Corel-1K for the top 10 retrieved images

Method	Precision (%)	Feature dimension
Dubey et al. [17]	74.93	2048
Murala et al. [18]	74.80	2048
Bhunia et al. [19]	79.86	312
Ahmed and Mohamed [20]	60.60	30
Dubey et al. [21]	73.84	1280
Proposed	81.90	144

the partial features was high in the order of color autocorrelogram, Gabor correlation, UMLBP, and SVD. Although most of them showed a better performance in the case of larger feature dimensions, the proposed method combining these four features depicted the best retrieval performance in all six DBs.

Figure 4 presents a comparison of the existing methods using each single-type feature and the proposed method for six DBs. The RULBP, CLBP, and color autocorrelogram features used for the comparison were extracted from the RGB region. The experiment result demonstrated that the proposed method has excellent performance in all six DBs. Tables 3 and 4 present numeric details of the retrieval performance shown in Figures 2–4. Figures 5 and 6 illustrate examples of the top five retrieved images from the six DBs using the proposed method. Most of the similar images were retrieved in the upper ranks.

Table 5 compares the latest methods and our proposed method. Our method of a low dimension yielded the highest performance.

6 | CONCLUSIONS

This study proposed a feature extraction method that is a combination of SVD, Gabor local correlation, UMLBP, and color autocorrelogram. The features extracted from the proposed method were not only a combination of color, texture, and shape features but also a fusion of the local and global features. The color feature came from the color autocorrelogram. The texture features were extracted from the UMLBP and Gabor local correlation. Lastly, the shape features came from SVD. The latter exhibited the global structure of an image, while the other features showed local characteristics.

In the experiment, the proposed method was compared with the methods that used partial features,

existing methods, and recent ones. The results showed that our proposed method is superior over other methods.

The main factor for the superior performance of our method is that the fusion of the color, texture, and shape features that possessed the local and global characteristics of an image went well with the other features. The proposed method exhibited the highest retrieval performance among all compared methods, showing a much greater effect on the average than those compared in Corel_MR and Corel_MD DBs. In addition, the dimensions of the proposed method were lower than those of recent techniques with a good performance (Table 5).

A limitation of the proposed method is that its retrieval performance on inhomogeneous images is not that good compared to that on homogeneous images.

Future research must apply the proposed features to semantic-based image retrieval using deep learning.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

ORCID

Hee Hyung Bu https://orcid.org/0000-0003-1637-6523

REFERENCES

1. M. J. Swain and D. H. Ballard, *Color indexing*, Int. J. Comput. Vis. **7** (1991), no. 1, 11–32.
2. J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu, and R. Zabih, *Image indexing using color correlograms*, (Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico), 1997, 762–768.
3. ISO/IEC 15938-3/FDIS Information Technology, *Multimedia content description interface—Part 3: Visual*, ISO/IEC/JTC 1/SC 29/WG 11, Doc. N4358, 2001.
4. X. Duanmu, *Image retrieval using color moment invariant*, (Seventh International Conference on Information Technology: New Generations, Las Vegas, NV, USA), 2010, pp. 200–203.
5. R. M. Haralick, K. Shanmugam, and I. Dinstein, *Texture features for image classification*, IEEE Trans. Syst. Man Cybern. **3** (1973), no. 6, 610–621.
6. J. Han and K. K. Ma, *Rotation-invariant and scale-invariant Gabor features for texture image retrieval*, J. Image Vis. Comput. **25** (2007), no. 9, 1474–1481.
7. C. Anibou, M. N. Saidi, and D. Aboutajdine, *Classification of textured images based on discrete wavelet transform and information fusion*, J. Inf. Process. Syst. **11** (2015), no. 3, 421–437.
8. H. Bu, N. C. Kim, C. Moon, and J. Kim, *Content-based image retrieval using combined color and texture features extracted by multi-resolution multi-direction filtering*, J. Inf. Process. Syst. **13** (2017), no. 3, 464–475.
9. T. Ojala, M. Pietikainen, and T. Maenpaa, *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*, IEEE Trans. Pattern Anal. Mach. Intell. **24** (2002), no. 7, 971–987.

10. Z. Guo, L. Zhang, and D. Zhang, *A completed modeling of local binary pattern operator for texture classification*, IEEE Trans. Image Process. **19** (2010), no. 6, 1657–1663.

11. H. Mahi, H. Isabaten, and C. Serief, *Zernike moments and SVM for shape classification in very high resolution satellite images*, Int. Arab J. Inf. Technol. **11** (2014), no. 1, 43–51.

12. M. R. Teague, *Image analysis via the general theory of moments*, J. Opt. Soc. Am. **70** (1980), no. 8, 920–930.

13. M. K. Hu, *Visual pattern recognition by moment invariants*, IRE Trans. Inf. Theory **8** (1962), no. 2, 179–187.

14. C. S. Lin and C. L. Hwang, *New forms of shape invariants from elliptic Fourier descriptors*, Pattern Recognit. **20** (1987), no. 5, 535–545.

15. H. Lynn Beus and S. H. Tiu, *An improved corner detection algorithm based on chain coded plane curve*, Pattern Recognit. **20** (1987), no. 3, 291–296.

16. S. Selvan and S. Ramakrishnan, *SVD-based modeling for image texture classification using wavelet transformation*, IEEE Trans. Image Process. **16** (2007), no. 11, 2688–2696.

17. S. R. Dubey, S. K. Singh, and R. K. Singh, *Multichannel decoded local binary patterns for content-based image retrieval*, IEEE Trans. Image Process. **25** (2016), no. 9, 4018–4032.

18. S. Murala, R. P. Maheshwari, and R. Balasubramanian, *Directional local extrema patterns: a new descriptor for content based image retrieval*, Int. J. Multimed. Inf. Retr. **1** (2012), 191–203.

19. A. K. Bhunia, A. Bhattacharyya, P. Banerjee, P. P. Roy, and S. Murala, *A novel feature descriptor for image retrieval by combining modified color histogram and diagonally symmetric co-occurrence texture pattern*, Pattern Anal. Appl. **23** (2020), 703–723.

20. A. Ahmed and S. Mohamed, *Implementation of early and late fusion methods for content-based image retrieval*, Int. J. Adv. Appl. Sci. **8** (2022), no. 7, 97–105.

21. S. R. Dubey, S. K. Singh, and R. K. Singh, *Boosting local binary pattern with bag-of-filters for content based image retrieval*, (IEEE UP Section Conference on Electrical Computer and Electronics, Allahabad, India), (2015), pp 1–6. https://doi.org/10.1109/UPCON.2015.7456703

22. I. M. Hameed, S. H. Abdulhussain, and B. M. Mahmmod, *Content-based image retrieval: A review of recent trends*, Cogent Eng. **8** (2021), no. 1, 1927469.

23. N. K. Rout, M. Atulkar, and M. K. Ahirwal, *A review on content-based image retrieval system: present trends and future challenges*, Int. J. Comput. Vis. Robot. **11** (2021), no. 5, 461–485.

24. M. M. Kumar and M. Kumar, *XGBoost: 2D-object recognition using shape descriptors and extreme gradient boosting classifier*, In *Computational methods and data engineering. Advances in Intelligent Systems and Computing*, Vol. **1227**, Springer, 2020, 207–222.

25. P. Chhabra, N. K. Garg, and M. Kumar, *Content-based image retrieval system using ORB and SIFT features*, Neural Comput. Applic. **32** (2020), 2725–2733.

26. I. M. Hameed, S. H. Abdulhussain, B. M. Mahmmod, A. Hussain, *Content based image retrieval based on feature fusion and support vector machine*, (14th International Conference on Developments in eSystems Engineering, Sharjah, United Arab Emirates), 2021, pp. 552–558. https://doi.org/10.1109/DeSE54285.2021.9719539

27. N. C. Kim and H. J. So, *Comments on SVD-based modeling for image texture classification using wavelet transform*, IEEE Trans. Image Process. **22** (2013), no. 12, 5408.

28. en.wikipedia.org/wiki/Singular_value_decomposition

29. info.hiroshima-cu.ac.jp/~miyazaki/knowledge/teche0104.html

30. codeproject.com/Articles/741559/Uniform-LBP-Features-and-spatial-Histogram-Computa

31. mathworks.com/matlabcentral/fileexchange/23253-gabor-filter

32. H. H. Bu, N. C. Kim, and S.-H. Kim, *Content-based image retrieval using a combination of texture and color features*, Hum.-Centric Comput. Inf. Sci. **11** (2021), no. 23, 1–13.

33. Y. D. Chun, N. C. Kim, and I. H. Jang, *Content-based image retrieval using multiresolution color and texture features*, IEEE Trans. Multimed. **10** (2008), no. 6, 1073–1084.

34. R. Picard, C. Graczyk, S. Mann, J. Wachman, L. Picard, and L. Campbell, *Vision texture, Massachusetts Institute of Technology*, Cambridge, MA, (1995). Available: vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html

35. J. Z. Wang, J. Li, and G. Wiederhold, *SIMPLIcity: Semantics-sensitive integrated matching for picture libraries*, IEEE Trans. Pattern Anal. Mach. Intell. **23** (2001), no. 9, 947–963.

36. W. Y. Ma and B. S. Manjunath, *A comparison of wavelet transform features for texture image annotation*, (Proc. International Conference on Image Processing, Washington, DC, USA), 1995, pp. 256–259.

37. D. Comaniciu, P. Meer, K. Xu, and D. Tyler, *Retrieval performance improvement through low rank corrections*, (Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries [CBAIVL], Fort Collins, CO, USA), 1999, pp. 50–54.

AUTHOR BIOGRAPHIES

Hee Hyung Bu She received her BS, MS, and PhD degrees in computer engineering from Mokpo National University (Jeonnam, Republic of Korea), Chonnam National University (Gwangju, Republic of Korea), and Kyungpook National University (Daegu, Republic of Korea) in 2004, 2006, and 2013, respectively. Since September 2019, she has been with the School of Computer Science and Engineering at Kyungpook National University, where she currently is an invited professor. Her research interests include image retrieval, video compression, and image processing.

Nam Chul Kim He received his BS degree in Electronic Engineering from Seoul National University in 1978 and MS and PhD degrees in electrical engineering from Korea Advanced Institute of Science and Technology, Seoul, Republic of Korea, in 1980 and 1984, respectively. He was a full-time professor at the School of Electronics Engineering at Kyungpook National University, Daegu,

Republic of Korea, from March 1984 to August 2020 and became an emeritus professor there. During 1991–1992, he was a visiting scholar in the Department of Electrical and Computer Engineering, Syracuse University, Syracuse, NY, United States. His research interests are image processing and computer vision, biomedical image processing, and image and video coding.

Sung Ho Kim He received his BS degree in electronics from Kyungpook National University, Daegu, Republic of Korea, in 1981 and his MS and PhD degrees in computer science from Korea Advanced Institute of Science and Technology, Seoul, Republic of Korea in 1983 and 1994, respectively. He has been a faculty member of the School of Computer Science and Engineering at Kyungpook National University, Daegu, Republic Korea, since 1986. His research interests include real-time image processing and telecommunication and multi-media systems.