Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-022-06644-5

RESEARCH ARTICLE-COMPUTER ENGINEERING AND COMPUTER SCIENCE

Histogram of Low-Level Visual Features for Salient Feature Extraction

Rubab Mehboob[1] · Ali Javed[2] · Hassan Dawood[1] · Hussain Dawood[3]

Received: 6 August 2021 / Accepted: 23 January 2022

Abstract

Distinctive and robust feature representation plays a crucial role in various multimedia applications. The descriptors invariant to rotations and textural viewpoints are able to extract discriminant features. In this paper, an improved feature descriptor named histogram of low-level visual features (HILL-VF) is proposed to extract distinctive features. In HILL-VF, fused edge maps based on gradient magnitude (FEM-GM) are obtained by using the directional derivative filters of Sobel and Scharr operators in YCbCr color space. Diffusion equation integrates the inherent edge details extracted by Sobel and Scharr edge detection operator by giving higher weights to the chroma components of the image. Moreover, phase maps based on Gabor features (PH-MGF) characterize the phase information by using Gabor filters by varying different orientations. Micro-FEM-GM and PH-MGF are generated by encoding the FEM-GM and PH-MGF into pre-defined intervals based on the selected seed values. These encoded micro-maps are then represented through a 2-D histogram. Experimental evaluations are conducted on four standard benchmarks, i.e., Coil-100k, KTH-TIPS, KTH-TIPS2-a, and -b. Experimental results indicate that we are able to increase the classification accuracy to 96.97%, 87.5%, 97%, and 93.01% on Coil-100, KTH-TIPS, KTH-TIPS2-a, and -b, respectively.

Keywords Feature extraction · Texture · Contrast · Orientation · Viewpoints variation · Rotation invariant · Edges

1 Introduction

The continuous advancement in multimedia tools and applications has promoted ease of accessibility to wider range of images containing diverse content. The proliferation of internet technologies and advancement in multimedia tools have prompted the use of wider range of images containing variety of content [1]. The variation in digital content has led to an enormous expansion of multimedia databases in almost every field such as medical, education, social media, and forensics [2]. Discriminant and robust feature representation is, therefore, a key in an array of multimedia applications. The image representations which are resilient against transformations such as rotations and translations provide necessary cues to the discriminant information contained in the image [3]. Since inspecting an object in an image or a scene in a visual image is complemented by the subjective experience of the user similar to examining an object physically in real world. Therefore, it can be inferred that there exists some association between information processing of visual perception and imagery. The primary visual features, i.e., low-level features are broadly classified into two categories such as global and local features [4]. Global features extract the visual information such as shape, color, textures, and spatial content of an image by taking whole image under consideration. Contrary, the local features characterize the parts of images such as corners and edge details. Similarly, texture is defined as low-level visual features that describe the surface of a material or an object. However, it is of primary importance to give features some meaningful representation that can be easily decoded and its semantics can also be easily understood.

In the last few years, strenuous research has been conducted on numerous descriptors for feature extraction and image representation. A well-known feature descriptor named local binary pattern has been employed to extract the feature representation from the images [5]. A modified variant of local binary pattern (LBP) based on 8-dimensional Krisch mask is proposed for extraction of salient features

✉ Ali Javed
ali.javed@uettaxila.edu.pk

[1] Department of Software Engineering, University of Engineering and Technology-Taxila, Taxila, Punjab 47050, Pakistan

[2] Department of Computer Science, University of Engineering and Technology-Taxila, Taxila, Punjab 47050, Pakistan

[3] Department of Computer and Network Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia

Published online: 08 March 2022

[6]. Two other variants of LBP such as nLBP and dLBP are proposed in [7]. The feature vector of nLBP is constructed by establishing the relations between the order of neighbors within a compact neighborhood. Similarly, the feature vector of dLBP is obtained by establishing the relation between the neighboring pixels and central pixel. Rotation-invariant versions of LBP such as rotation-invariant LBP [8], uniform LBP [9], adaptive LBP (ALBP) [10], SSELBP [11], COV-LBPD [12], COV-LBPD [12], scLBP [13], and multi-scale spatial pyramid LBP (MSSP-LBP) [14] have also been proposed to extract the compact feature representations of images. However, LBP ignores the spatial information of central pixel. Attractive-and-repulsive center-symmetric local binary patterns (ARCS) preserve the characteristics of uniform LPB by extracting gradient and textural characteristics [15]. Genetic programming (GP)-based method (HL-GP) fuses histogram of oriented gradients (HOG) and LBP to achieve rotation invariance and also reduces the high dimensions [16]. Another genetic programming (GP)-based Genetic Texture Descriptor (GTD) is proposed to extract the global textural details [17]. Additionally, the local edge signature (LES) incorporates the statistical information of edges and their orientations in a local neighborhood [17]. However, random selection in GP-based methods does not always guarantee an optimal feature selection. Since local features are invariant to geometric changes in illumination changes, therefore quaternionic extended local binary pattern (QxLBP) encodes discriminative features and aggregates local features by multiresolution pooling [18]. Similarly, local triangular coded pattern (LTCP), binary robust independent elementary features (BRIEF) [19], and local binary difference (LBD), local morphology pattern (LMP), and local directional relation pattern (LDRP) ensure the localization of objects and texture by preserving the significant contour information. A combination of local ternary pattern (LTP) and local directional pattern (LDP) is proposed in [20]. Scale-invariant feature transform (SIFT) [21] and speeded-up robust features (SURF) [22] extract the local image structures while using quantized image gradients. However, these descriptors are insensitive to monotonic illumination changes are also cost-effective. In the local neighborhood, local convex representation (LCR) is proposed to represent a data point in terms of other data points [23]. The major drawback of binarization is the loss of information due to intensity differences and their dimensions also grow exponentially. Normalized difference vector (NDV) [24] and dense micro-block difference (DMD) [25] compute the real-valued intensity difference instead of binary codes. Block intensity and gradient difference (BIGD) captures the variations in local patch by computing the gradient difference [3]. Train features extracted by FV-VGGM [26] and FV-VGGVD [26] for colored images lack the geometric invariance, therefore, unable to recognize textures with high variations. The pre-trained model on weights of ImageNet in FV-Alexnet [26] increases the computational cost.

Sorted random projection (SRP) considers the circular geometry to compress the pixel intensity difference [27]. Local higher-order statistics (LHS) [28] considers a patch to incorporate the high-order statistics of the pixel difference. Circular geometry in the proposed method is sufficient enough to capture the radial variations in a patch of an image while ignoring the pixels present at all other directions. Weber local descriptor captures the textural variations of the image [29]. Basic image features (BIF) extracts 7 features, and each feature represents a specific structure [28] and leads to an increased computational cost. Another method named HMAX [30] uses a top-down approach for finding the desired objects of certain sizes. Rotation-invariant kernels (RIK) have also been proposed for modeling the data by using the complex scalar and rotation-invariant distributions. A local feature-based approach named sparse gradient enhanced features (SGEF) has been proposed by using the sparse data [31]. However, SGEF ignores the essential chromatic information of the images. Sparse Weber-oriented visual features (SWOVF) overcome the imperfections of [32] by extracting the gradient magnitude, edge orientation, and pixel intensities as the significant information for object detection. The technique proposed in [33] uses a cosine kernels to approximate the complex convolution kernels. Randomized nonlinear PCANet (RNPCANet) processes nonlinear data by using kernel methods [34]. In addition, kernel approximation is used to characterize the feature space. A feature descriptor based on complex networks (CN) and uniform local binary pattern (ULBP) is proposed to characterize the image spatial relationships by using three measurements such as clustering coefficient and centrality of eigenvector and degree [35]. The proposed method [36] constructs a framework of capsule neural network by integrating tensor attention blocks, quantization techniques, and wavelet decomposition. Another method [37] based on capsule networks integrates multi-level attention blocks and wavelet decomposition is proposed for extraction of textural characteristics of images. Holistic and hierarchical order-encoding patterns (H2OEP) based on the color information is proposed for texture classification [38]. The method proposed in [39] adopts VGG-19 for extracting salient features of images. The proposed method generates a single feature vector by combining CNN activations from different convolutions layers and then applies the pooling operation [40]. Shi et al. [41] determine that local geometry is more effective in material recognition. Spatially weighted order binary pattern (SWOBP) encodes the internal channel features by using the color gradient [42]. Moreover, the local color differences are decomposed into spatially weighted binary templates. Weight learning schemes are proposed to adaptively learn the weights of features [43]. Also, [43] minimizes the effect of outliers such as noise.

Binary descriptors such as LBP [8], uniform LBP [9], adaptive LBP (ALBP) [10], SSELBP [11], COV-LBPD [12], COV-LBPD [12], scLBP [13], and multi-scale spatial pyramid LBP (MSSP-LBP) [14] result in loss of intensity information due to binarization process. It should be noted that differential excitation and characteristic orientation of WLD [29] extracts contrast and phase-based features. In this paper, an improved feature descriptor homologous to a well-known Weber local descriptor (WLD) is proposed. As WLD extracts contrast- and phase-based features from the images. WLD is executed in two phases such as differential excitation and characteristic orientation. However, WLD is likely to suffer from two problems: (1) positive and negative differences in differential excitation component counteract with each other, (2) characteristic orientation does not effectively reflect the phase information of textures and continuously varying textures, and (3) it is sensitive to rotations. In this paper, an improved feature descriptor named histogram of low-level visual features (HILL-VF) is proposed to overcome the imperfections of WLD. In the first phase, significant edge information is extracted by using two edge detection operators, i.e., Sobel and Scharr in YCbCr color space. Moreover, a diffusion equation is proposed to fuse the extracted edge information by assigning higher weights to the chroma component. In the second phase, Gabor features are employed to extract the phase information of varying textures and patterns in frequency domain. Finally, the extracted contrast- and phase-based features are encoded by using seed values to reduce the cost of high dimensions. The main contributions of the proposed method are as follows:

- Fused edge maps based on gradient magnitude (FEM-GM) integrate the horizontal and vertical edge contours obtained by Scharr and Sobel edge detectors by assigning high weights to the chroma components.
- The explicit edge orientation information in Phase Maps based on Gabor Features (PH-MGF) is obtained by using Gabor features in frequency domain.
- Micro-FEM-GM and PH-MGF are generated by using seed values. Both features are then integrated to generate a final feature vector.

The rest of the paper is organized as follows: Sect. 2 presents the proposed methodology of HILL-VF, Sect. 3 presents the experimental results and discussion. Finally, Sect. 4 concludes the proposed work and also provides future dimensions.

2 The Proposed Approach

In this paper, an improved feature descriptor named histogram of low-level visual features (HILL-VF) is proposed for extraction of salient features. The pipeline (see Fig. 1) of our HILL-VF consists of two stages, i.e., feature extraction and their integration. HILL-VF is based on the extraction

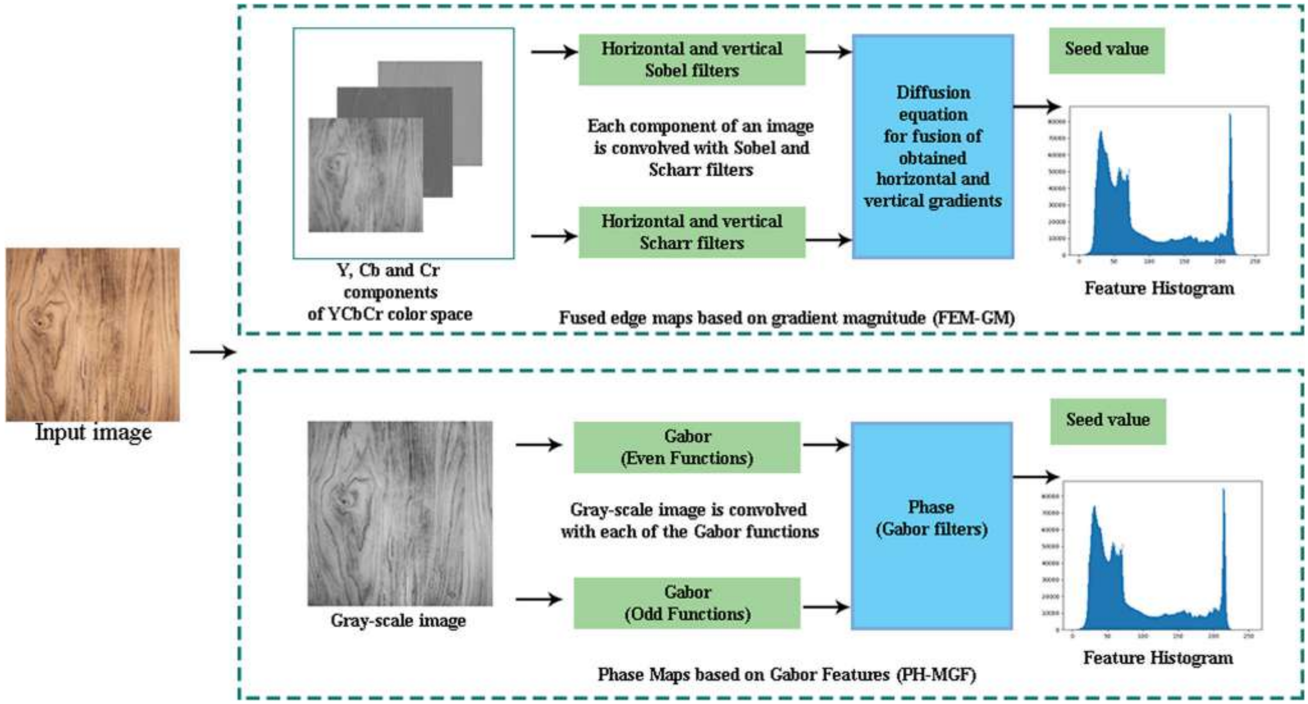
Input image
Y, Cb and Cr components of YCbCr color space
Horizontal and vertical Sobel filters
Each component of an image is convolved with Sobel and Scharr filters
Horizontal and vertical Scharr filters
Diffusion equation for fusion of obtained horizontal and vertical gradients
Seed value
Feature Histogram
Fused edge maps based on gradient magnitude (FEM-GM)
Gray-scale image
Gabor (Even Functions)
Gray-scale image is convolved with each of the Gabor functions
Gabor (Odd Functions)
Phase (Gabor filters)
Seed value
Feature Histogram
Phase Maps based on Gabor Features (PH-MGF)

Fig. 1 Pipeline of feature extraction scheme of HILL-VF

of low-level features such as color, edges (continuities and discontinuities), and orientation of edges. Low-level visual features being invariant to translations and less sensitive to noise and occlusions are considered in this approach for extraction of the salient features. In the proposed HILL-VF, FEM-GM extracts the inherent edge details of the input image by using the directional derivative filters of the Sobel and Scharr edge detection operators in YCbCr color space. However, the strength of horizontal and vertical edge information is obtained by assigning higher weights to the chroma components of the image. Moreover, the PH-MGF is proposed to characterize the phase information in the transform domain. Prior to the feature integration, both micro-FEM-GM and PH-MGF are explicitly encoded into the pre-defined intervals. These pre-defined intervals are set based on the selected seed values. The integrated feature set is then represented through a 2-D histogram.

2.1 Feature Extraction

2.1.1 Fused Edge Maps Based on Gradient Magnitude (FEM-GM)

Edge contours represent the image structures with larger spatial extents efficiently. This is most likely due to the fact that these cues are invariant to changes in color, texture, and lightning conditions. FEM-GM in the proposed HILL-VF is designed with the motivation to (1) Preserve the inherent discontinuities and continuities, (2) Minimize the trade-off between the false positives and negatives, (3) Preserve the precise location of the potential (obvious) edge contours. In gray-scale images, the abrupt variations in pixel intensities may increase the likelihood of detection of false positives. RGB color space is not close to human perception. In RGB color space, the homogeneous and sharper variations are likely to observe within the subject area and along the boundaries of an object, respectively. Therefore, in the proposed FEM-GM, the input image $f(x, y)$ is transformed from RGB color space to YCbCR color space. In YCbCr color space, Y represents the luma component whereas Cb (Blue-Y) and Cr (Red-Y) represent the chroma component of an input image. Equation (1) shows the transformation of input image $f(x, y)$ to YCbCr color space. Inspired by the efficiency and effectiveness of Sobel and Scharr operators, the directional derivative filters, i.e., Sobel and Scharr operators are used to extract the inherent edge details along horizontal and vertical directions. Since Sobel is sensitive to diagonal edges and may lead to inaccurate computations in gradient magnitude. Therefore, Scharr edge detection operator being anisotropic in nature is used for extraction of edge contours. The Scharr operator is an optimized kernel that minimizes the weighted mean-squared angular error (WMS-AE). Unlike, Prewitt, Average, Robert, and Canny, Scharr is invariant to rotations. Moreover, the decreasing weights along diagonals give the compact feature representation. Equations (2)–(3) give the strength of edges extracted by Sobel edge detection operator along horizontal and vertical directions of transformed input image $(f'(x, y))$. Equations (6)–(7) give the strength of edges extracted by Scharr edge detection operator along horizontal and vertical directions of transformed input image $(f'(x, y))$.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 65.738 & 129.059 & 25.064 \\ -37.945 & -74.494 & 112.439 \\ 112.439 & -94.154 & -18.285 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{1}$$

Here, R, G, and B represent the red, green, and blue channels of the input image $(f(x, y))$, respectively. Y, Cb, Cr represent the luma, chroma (Blue-Y), and (Red-Y) of the transformed image $(f'(x, y))$, respectively.

$$gx(x, y, j) = f'(x, y, j) \circledast \text{Sobel}_{\text{horizontal}}. \tag{2}$$

Here, j refers to the components of the transformed input image $(f'(x, y))$. The value of J ranges from 1 to 3. Here, $\circledast$ represents the convolution of the transformed input image $(f'(x, y))$ with Sobel kernel (horizontal). When the value of J is 1, it refers to the luma (Y) component of the transformed image and $gx(x, y, 1)$ represents the corresponding horizontal gradients. Likewise, when the value of j is set to 2, it refers to the Cb (Blue-Y) component of the image, and $gx(x, y, 2)$ represents the corresponding horizontal gradients. Similarly, when the value of j is set to 3, it refers to the Cr (Red-Y) component of the image, and $gx(x, y, 3)$ represents the corresponding horizontal gradients. x and y refer to the spatial coordinates of the image.

$$gy(x, y, j) = f'(x, y, j) \circledast \text{Sobel}_{\text{vertical}}. \tag{3}$$

Here, j refers to the components of the transformed input image $(f'(x, y))$. The value of J ranges from 1 to 3. Here, $\circledast$ represents the convolution of the transformed input image $(f'(x, y))$ with Sobel kernel (vertical). When the value of J is 1, it refers to the luma (Y) component of the transformed image and $gy(x, y, 1)$ represents the corresponding vertical gradients. Likewise, when the value of j is set to 2, it refers to the Cb (Blue-Y) component of the image, and $gy(x, y, 2)$ represents the corresponding vertical gradients. Likewise, when the value of j is set to 3, it refers to the Cr (Red-Y) component of the image, and gy(x,y,3) represents the corresponding vertical gradients. x and y refer to the spatial coordinates of the image. The stride of 1 is set to preserve the continuities. Two Kernels of Sobel edge detection operator ($\text{Sobel}_{\text{horizontal}}$) and $\text{Sobel}_{\text{vertical}}$ are given as follows:

$$\text{Sobel}_{\text{horizontal}} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \tag{4}$$

$$\text{Sobel}_{\text{horizontal}} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}. \tag{5}$$

$$gxx(x, y, j) = f'(x, y, j) \circledast \text{Scharr}_{\text{horizontal}}. \tag{6}$$

Here, j refers to the components of the transformed input image ($f'(x, y)$). The value of J ranges from 1 to 3. Here, $\circledast$ represents the convolution of the transformed input image ($f'(x, y)$) with Scharr kernel (horizontal). When the value of J is 1, it refers to the luma (Y) component of the transformed image and $gxx(x, y, 1)$ represents the corresponding horizontal gradients. Likewise, when the value of j is set to 2, it refers to the Cb (Blue-Y) component of the image, and $gxx(x, y, 2)$ represents the corresponding horizontal gradients. Likewise, when the value of j is set to 3, it refers to the Cr (Red-Y) component of the image, and $gxx(x, y, 3)$ represents the corresponding horizontal gradients. x and y refer to the spatial coordinates of the image.

$$gyy(x, y, j) = f'(x, y, j) \circledast \text{Scharr}_{\text{vertical}}. \tag{7}$$

Here, j refers to the components of the transformed input image ($f'(x, y)$). The value of J ranges from 1 to 3. Here, $\circledast$ represents the convolution of the transformed input image ($f'(x, y)$) with Scharr kernel (vertical). When the value of J is 1, it refers to the luma (Y) component of the transformed image and $gyy(x, y, 1)$ represents the corresponding horizontal gradients. Likewise, when the value of j is set to 2, it refers to the Cb (Blue-Y) component of the image, and $gyy(x, y, 2)$ represents the corresponding horizontal gradients. Likewise, when the value of j is set to 3, it refers to the Cr (Red-Y) component of the image, and $gyy(x, y, 3)$ represents the corresponding horizontal gradients. x and y refer to the spatial coordinates of the image. The stride of 1 is set to preserve the continuities. Two Kernels of Scharr edge detection operator ($\text{Scharr}_{\text{horizontal}}$) and $\text{Scharr}_{\text{vertical}}$ are given as follows:

$$\text{Scharr}_{\text{horizontal}} = \begin{bmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{bmatrix}. \tag{8}$$

$$\text{Scharr}_{\text{vertical}} = \begin{bmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{bmatrix}. \tag{9}$$

Fundamentally, the strength of horizontal and vertical gradients obtained by Sobel and Scharr edge detection operators is obtained by the proposed diffusion equation.

$$\begin{aligned} \text{FEG} - \text{GM}_{\text{hor}} = {} & (gx(x, y, 1) \times gxx(x, y, 1)) \\ & + h \times (gx(x, y, 2) \times gxx(x, y, 2)) \\ & + \frac{h^2}{2}(gx(x, y, 3) \times gxx(x, y, 3)) \end{aligned} \tag{10}$$

$$\begin{aligned} \text{FEG} - \text{GM}_{\text{ver}} = {} & (gy(x, y, 1) \times gyy(x, y, 1)) \\ & + h \times (gy(x, y, 2) \times gyy(x, y, 2)) \\ & + \frac{h^2}{2}(gy(x, y, 3) \times gyy(x, y, 3)). \end{aligned} \tag{11}$$

Here, h is the penalty parameter. Since chromatic components, i.e., Cb and Cr have less detailed information of the image contents. Therefore, the penalty parameter, i.e., h is added while fusing gradients of chroma components. It has two senses: (1) to add the visual fidelity to the chroma components in the resultant edge map, and (2) assign weightage to the chroma components in the resultant edge map. It is important to mention that the h is a pre-defined constant and its value is set to 0.017455 in our proposed method. The strength of FEG-GM_{hor}, and FEG-GM_{ver} is computed as follows:

$$\text{FEG} - \text{GM} = \sqrt{\text{FEG} - \text{GM}_{\text{hor}}^2 + \text{FEG} - \text{GM}_{\text{ver}}^2}. \tag{12}$$

2.1.2 Phase Maps Based on Gabor Features (PH-MGF)

The magnitude of edge contours incorporate the strong semantic information of object boundaries and textural structures. The edge maps incorporate the implicit information regarding the location of pixels, their strength (consisting of horizontal, vertical, and diagonal edge information). Contrary, the orientation maps of edges incorporate explicit information of the contents of the image such as the geometry of shapes and textural patterns. The phase information in the proposed HILL-VF is extracted by using Gabor filters [44]. Since natural images are blend of textures, patterns, and have variable viewpoints. Therefore, transform domain is best suited to analyze such textures, patterns, and viewpoints. In HILL-VF, Gabor filter is utilized to extract the significant phase information with finer distinction of different textures. Gabor filters are used with the motivation to obtain the optimal localization of significant edges or contours. Generally, a 2-D Gabor filter is a band-pass filter. Gaussian function based on modulated orientation and frequencies is used as a transfer function in Gabor filter. The Gabor filter analyzes the input image ($f(x, y)$) in Fourier domain. It can be computed as follows:

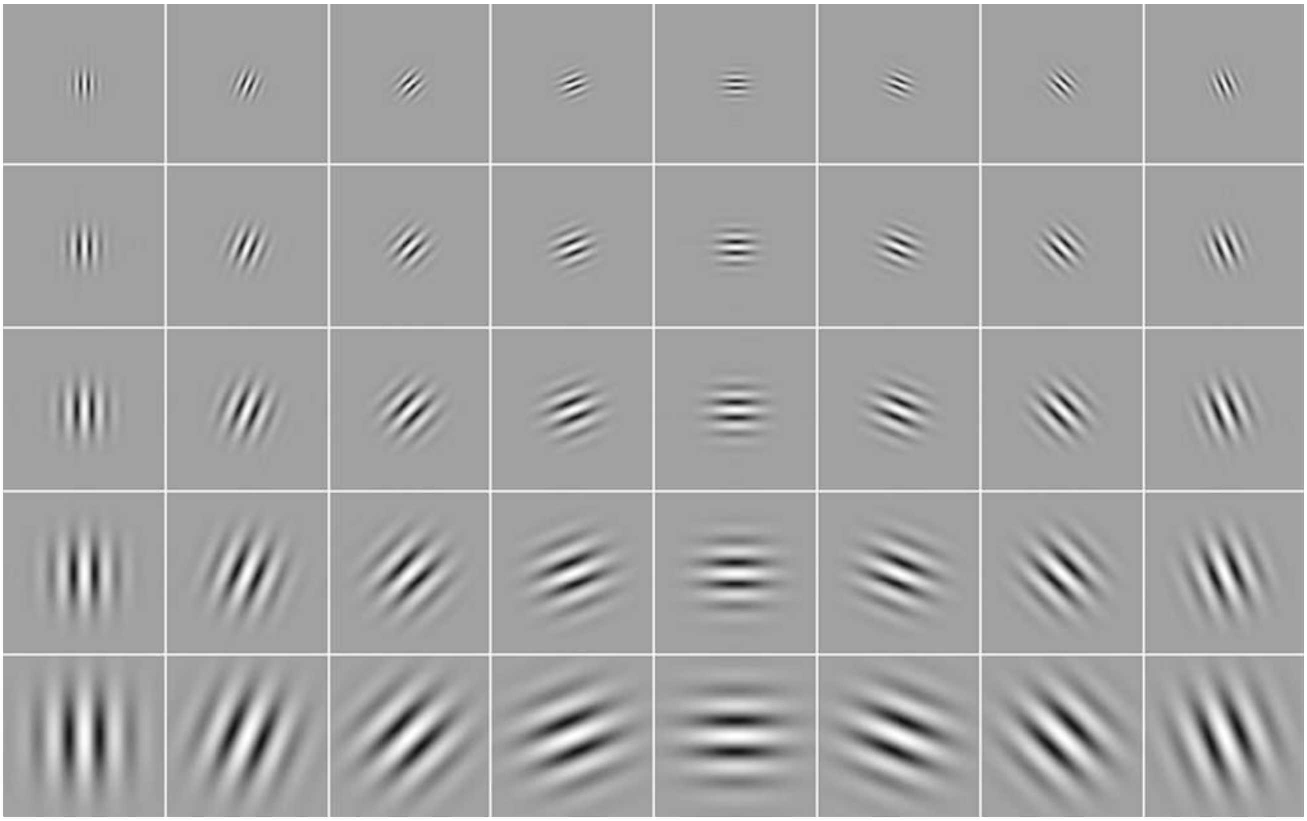

Fig. 2 Even- and odd-symmetric functions of Gabor filter [45]

$$G(x, y, f, \theta) = \frac{1}{2\pi\sigma_x\sigma_y} \times \exp\left(-\left[\frac{x'^2}{2\sigma_x^2} + \frac{y'^2}{2\sigma_y^2}\right]\right) \times \exp(i2\pi(u_0 x + v_0 y) = \frac{1}{2\pi\sigma_x\sigma_y} \times \exp\left(-\left[\frac{x'^2}{2\sigma_x^2} + \frac{y'^2}{2\sigma_y^2}\right]\right) \times exp(i(2\pi * f * x')). \quad (13)$$

Here, x and y refer to the spatial coordinates of the input image ($f(x, y)$). σ_x andσ_y represent the standard deviation of the Gaussian kernel in x and y directions, respectively. Besides, $x' = x\cos\theta + ysin\theta$ whereas $y' = -x\sin\theta + y\cos\theta$ refers to the rotated coordinates. The spatial frequencies, i.e., $u_0 = f\cos\theta$ and $v_0 = fsin\theta$. Gabor filter consists of a real part (even-symmetric component) and a complex part (an odd-symmetric component). Figure 1 shows the real and complex parts of a Gabor filter. Gaussian envelop in frequency domain centered at (u_0, v_0) is computed as follows:

$$\text{Gabor}(u, v) = \exp\left(-\left[\frac{u'^2}{2\sigma_u^2} + \frac{v'^2}{2\sigma_v^2}\right]\right) \quad (14)$$

Here, in frequency domain $x_0 = u - u_0$ and $y_0 = v - v_0$. Therefore, $u' = x_0cos\theta + y_0sin\theta$. Also, $v' = -x_0sin\theta + y_0cos\theta$. σ_u and σ_v refer to the spatial frequencies. The real and imaginary parts of a Gabor filter are computed by eliminating the DC component located at $G(0, 0)$. Even- and odd-symmetric functions are given by modifying Eq. (13) as follows:

$$\text{even}(x, y, g, \theta) = \frac{1}{2\pi\sigma_x\sigma_y} \times \exp\left(-\left[\frac{x'^2}{2\sigma_x^2} + \frac{y'^2}{2\sigma_y^2}\right]\right) \times -\exp(-2\pi^2\sigma_0^2). \quad (15)$$

Here, $sigma_x = \sigma_0$ by eliminating the DC component.

$$\text{odd}(x, y, g, \theta) = \frac{1}{2\pi\sigma_x\sigma_y} \times \exp\left(-\left[\frac{x'^2}{2\sigma_x^2} + \frac{y'^2}{2\sigma_y^2}\right]\right) \times \exp(-i2\pi * f * x'). \quad (16)$$

The input image ($f(x, y)$) is convolved with even and odd functions to obtain the phase information. Worth noting, another parameter, i.e., the spatial frequency (λ) is set to 2 in our proposed PH-MGF. However, the valid value of λ ranges from 2 to ∞, Then, the orientation is given by $\frac{360}{4} = 90°$. The resultant phase maps (PH-MGF) are obtained by convolving the input image $f(x, y)$ with the Gabor filter.

2.2 Generation of Feature Vector

To account for task-relevance and context of classification, it is presumed that the extracted feature sets, i.e., FEG-GM, and PH-MGF incorporate more enduring representations to the extent that passes a certain kind of relevance. It is, therefore, necessary to register the extracted features into a single feature set. Direct integration of both the FEG-GM, and PH-MGF may lead to an increased computation cost. Prior to the feature integration process, seed values ranging between 15 and 300 are generated to encode the FEG-GM, and PH-MGF into selected intervals. The seed values are used to generate the micro-feature maps because of the following three reasons: (1) to reduce the high-dimensional features to low dimensions, (2) to reduce the computational cost, and (3) to reduce the redundant information. These seed values are then considered to encode FEG-GM, and PH-MGF into intervals and generate the micro-FEG-GM, and PH-MGF maps. Let p and q be the selected seed values for the encoding of FEG-GM, and PH-MGF maps, respectively. The micro-FEG-GM, and PH-MGF maps are computed as follows:

$$\begin{aligned}\mathrm{Micro}_{\mathrm{FEG-GM}}(i) &= \{(x, y)|(x, y) \in \mathrm{FEG-GM} \\ &= i, 1 \leq i \leq p\} \qquad (17)\\ \mathrm{Micro}_{\mathrm{PH-MGF}}(i) &= \{(x, y)|(x, y) \in \mathrm{PH-MGF} \\ &= i, 1 \leq i \leq q\}. \qquad (18)\end{aligned}$$

Here, x and y represents the spatial coordinates and i be the resultant value obtained in micro-FEG-GM, and PH-MGF maps. p and q are the desired seed values for encoding FEG-GM, and PH-MGF maps. The integrated feature set is then fed into the linear kernel of support vector machine (SVM) for classification. However, SVM is optimized by setting the value of C. Nonetheless, C is a regularization parameter used to control the trade-off between the train error and norm of weights. Therefore, the value of C is set to 1 in our proposed HILL-VF.

3 Experimental Results

Three standard image datasets, i.e., Columbia Object Image Library (Coil-100) [46], Kth-tips [47], KTH-TIPS2-a and KTH-TIPS2-b [48] are used to evaluate the performance of proposed HILL-VF. Characteristic features and description of train and test sets of three benchmarks are described in Sect. 3.1. Effect of variation of parameters on proposed HILL-VF is analyzed in Sect. 3.2. The impact of different subsets of features and computation time are analyzed in Sect. 3.3. Section 3.4 presents the effect of different color spaces on proposed FEG-GM. Section 3.5 presents different accuracies obtained by varying SVM kernels and different classification frameworks. The performance of conventional edge detection operators and proposed FEG-GM is analyzed in Sect. 3.6. Finally, Sect. 3.7 presents the performance comparison of proposed HILL-VF against the state-of-the-art methods on Coil-100, KTH-TIPS, and KTH-TIPS2-a and -b.

3.1 Datasets

Three standard image datasets, i.e., Columbia Object Image Library (Coil-100) [46], KTH-TIPS [47], KTH-TIPS2-a and KTH-TIPS2-b [48] are used to evaluate the performance of proposed HILL-VF. Coil-100 consists of 7200 colored images(RGB) of 100 objects. Each object has 72 different images. Moreover, each object varies with respect to geometry, viewpoints, and reflectance characteristics. The motorized turntable against a black background was rotated through 360 degrees to capture variable viewpoints of an object. Further, the pose interval was set to 5 degrees while capturing the images of the objects. Therefore, each object has 72 different viewpoints. Coil-100 is partitioned into two subsets, i.e., TS1 and TS2. TS1 consisting of 3200 images is used to train the proposed method. TS2 consisting of 7200 images is used to evaluate the generalization performance of the proposed method. Another dataset named KTH-TIPS consists of variable textures, illuminations, viewpoints, and scales is used for performance evaluation of proposed method. KTH-TIPS image database was created as an extension of the CUReT database in two different ways, i.e., (1) by varying scales, viewpoints, and illumination, and (2) images of the other samples of a subset of 11 different materials were captured in variable settings. The KTH-TIPS2 databases, i.e., a and b consist of images of 4 different samples of 11 different materials with varying viewpoints, illumination, and scale. Cross-validation scheme is used to avoid the problem of overfitting. Independent training and testing sets for each of the datasets are created and used for the performance evaluation of proposed method. KTH-TIPS dataset is partitioned into two subsets, i.e., TS3 and TS4. TS3 consists of 50%, i.e., 400 images are used for training and rest of the 50% images, i.e., 410 are used to test the generalization performance of the proposed HILL-VF. The KTH-TIPS2-a and -b are partitioned into five folds. Four folds are used for training and one fold is used for testing (without replacement). This process is repeated five times. Only those folds are retained for training and testing which give the highest accuracies on these datasets. Moreover, KTH-TIPS2-a and -b datasets are partitioned into two subsets, i.e., TS5 and TS6. TS5 consisting of 80% of the images, i.e., 3564 is used for training and TS6 consisting of 20% of the images, i.e., 1188 is used to evaluate the generalization performance of the proposed method. Table 1 depicts the characteristic features of each of the dataset used for performance evaluation of the

Table 1 Characteristic features of Coil-100, KTH-TIPS, and KTH-TIPS2-a and b

Dataset	# Of classes	Size	# Of images	Train images	Test images	Characteristics
Coil-100	100	128 × 128	7200	3200	7200	1 sample has 72 different viewpoints
KTH-TIPS	10	200 × 200	810	400	410	1 sample has 3 different viewpoints, 3 illuminants, and 9 different scales
KTH-TIPS2-a	11	200 × 200	4752	3564	1188	4 samples have 3 different viewpoints, 4 illuminants, and 9 different scales
KTH-TIPS2-b	11	200 × 200	4752	3564	1188	4 samples have 3 different viewing angles, 4 illuminants, and 9 different scales

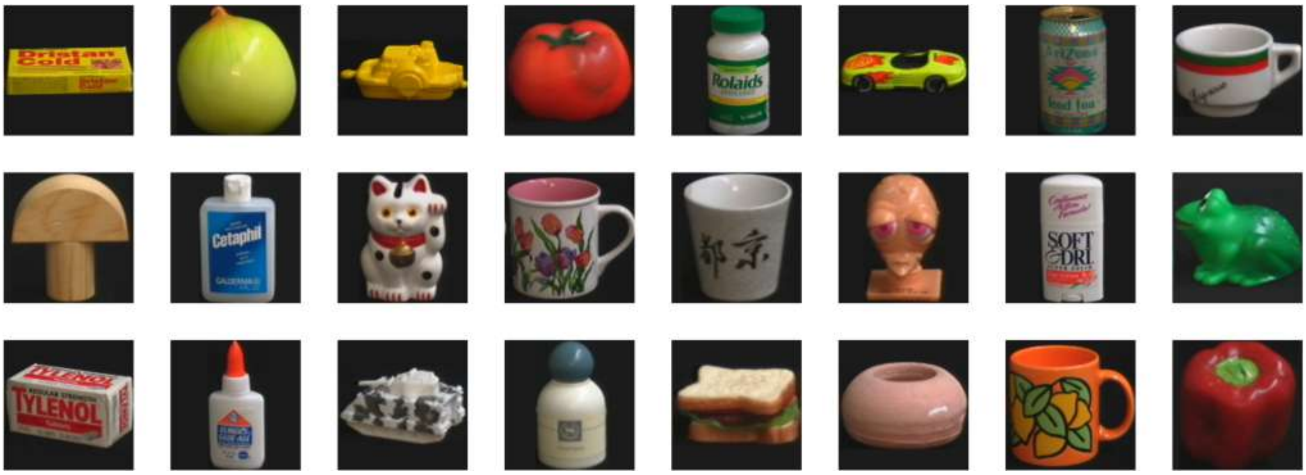

Fig. 3 Visual samples from Coil-100

proposed method. Figs. 3, 4, 5, and 6 present the few of the visual samples of Coil-100, KTH-TIPS, KTH-TIPS2-a, and -b, respectively.

3.2 Parameter Estimation for Proposed HILL-VF

In this section, experiments are conducted on KTH-TIPS to evaluate the generalization performance of the proposed HILL-VF by varying parameters. A linear support vector machine (SVM) is used for classification of proposed method. Since wavelength and orientation have substantial effect on the performance of proposed HILL-VF. In the proposed method, λ is used to control the width of the strips of a Gabor filter. Therefore, several variants of wavelength $\lambda \in \{1, 2, 4, 6, 8\}$, and orientation $O \in \{0, 30, 60, 90, 120, 150, 180\}$ have been used for experimental evaluations. Table 2 summarizes the performance (accuracy) of proposed method by varying wavelength (λ), and orientation (O). From the results reported in Table 2, it is noticed that best accuracy of 87.5% is obtained with the values of λ and O is 2 and 90, respectively. Furthermore, results reported in Tables 10, 11, and 12 also signify that the comparable results are obtained on Coil-100, KTH-TIPS2-a, and -b with same values of λ and O. This confirms that the choice of values of λ and O to extract the significant phase information is independent of the type of textural characteristics and variable viewpoints present in these datasets. The decrease in performance (from 55.25% to 52.75%) is observed with different values of Orientation (O) such as 0° , 30°, and 60° and with $lambda = 2$. The increase in intra-class variably has resulted in this graceful degradation. Unfortunately, the worst generalization performance of 36% is obtained with the values of $\lambda = 8$ and $O = 180°$. Since Gabor filters have strong orientation selectivity, therefore, increasing the orientations greater than 90° results in the loss of significant textural details. Moreover, the orientation levels less than 90° are not sufficient to represent the inherent textural details of the different materials. Further, higher values of λ lead to the thicker stripes. The values of $\lambda > 2$ most likely occlude the potential edge information and lead to the poor discrimination of different textures. The Gabor function is sampled in its zero crossings with values of $\lambda > 2$ and $O = 90°$

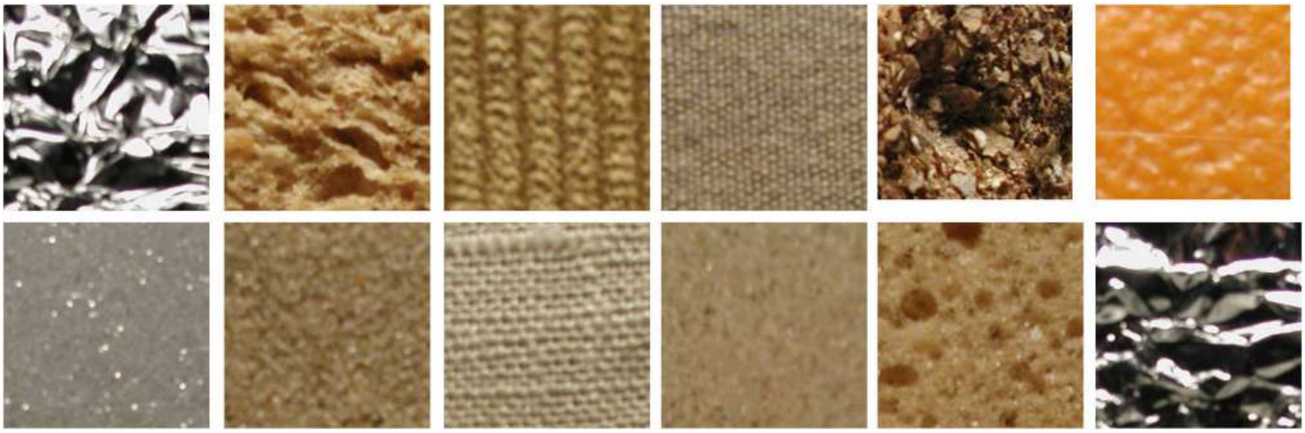

Fig. 4 Visual samples from KTH-TIPS

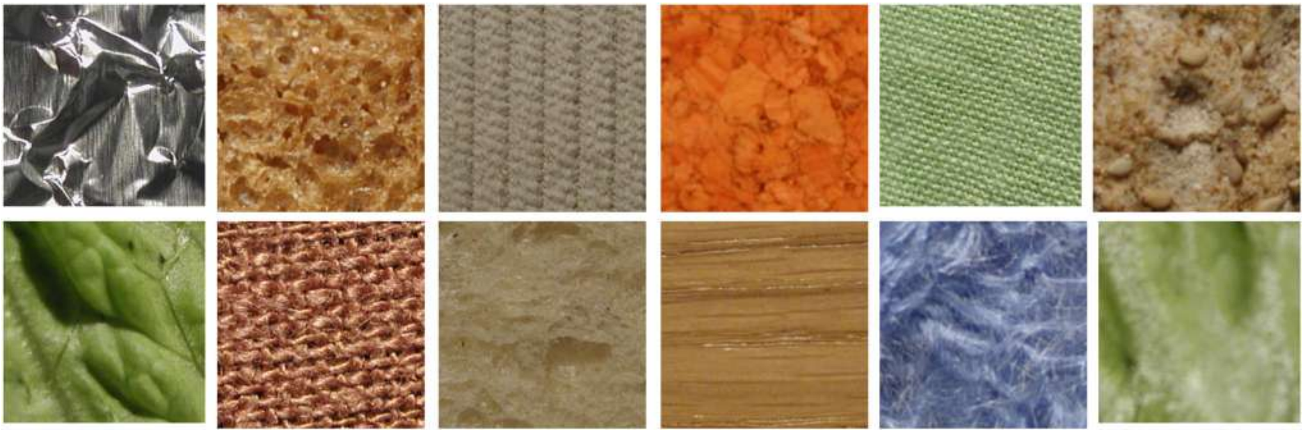

Fig. 5 Visual samples from KTH-TIP2-a

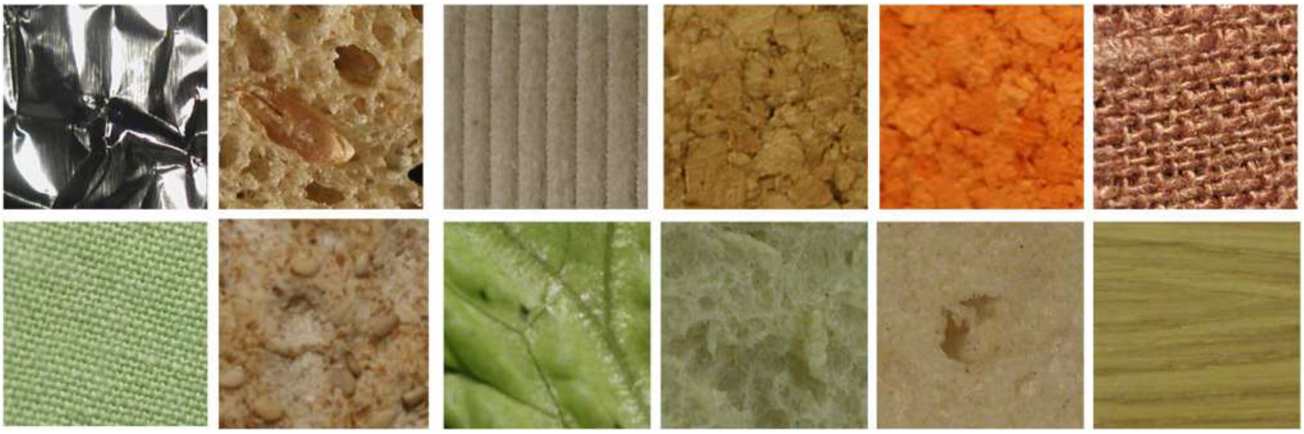

Fig. 6 Visual samples from KTH-TIP2-b

Table 2 Performance (accuracy) of proposed HILL-VF on KTH-TIPS by varying wavelength and orientation

Wavelength (λ)	Orientation (O)						
	0	30	60	90	120	150	180
1	52.45	51.25	52.75	76.0	53.25	54.2	53.0
2	55.25	54.75	52.75	**87.5**	54.0	54.5	53.0
4	53.75	52.75	54.5	82.25	53.25	54.75	51.5
6	51.5	53.0	51.5	78.0	54.75	53.75	50.25
8	52.75	57.5	54.75	78.0	55.25	42.5	36.5

Bold value indicates the highest accuracy obtained on each dataset

Table 3 Accuracy obtained on Coil-100 by varying p and q

Levels for p	Levels for q					
	50	100	150	200	250	300
7	90.05	93.45	94.10	94.5	93.5	95.75
15	91.75	92.56	94.50	94.10	95.0	**96.97**
19	92.50	93.50	93.05	92.75	91.25	96.25

Bold values indicate the highest accuracy obtained on each dataset

leading to the inclusion of false negatives in resultant edge maps. Therefore, for determining the inherent details 90° and $\lambda = 2$ are considered as the optimal choice in HILL-VF. Encoding levels (p and q) for construction of micro-feature maps are also the most critical parameters. In the proceeding section, the performance (accuracy) of the HILL-VF is evaluated by varying these two parameters. For p and q, the seed values are randomly generated between 15 and 300. The experimental results obtained on few of the random values of $p \in \{7, 15, 19\}$ and $q \in \{50, 100, 150, 200, 250, 300\}$ are summarized in Table 3 on Coil-100. It is evident from the results reported in Table 3 that highest accuracy of 96.72% is obtained on Coil-100 when the seed value of p is 15 and of q is 300. From Table 3, a slight increase in performance is observed when the encoding levels for q are increased from 50 to 300 when the value of p is 15. Similar trend is also observed when value of p is 7 and that of q varies from 50 to 300. Contrarily, a slightly different behavior is observed when the value of p is 19. However, this combination ($p = 19$ and $q = 300$) yielded an accuracy of 96.25% (closer to 96.75%) when the value of p is 19 and that of q is 300. This may be due to the fact the features selected in this combination are highly discriminating and are also complementary to other features. However, the size of feature vector obtained for this combination ($p = 15$ and $q = 300$) may not be cost-effective. Lower values of q (less than 300) may be detrimental for extracting the texture characteristics from small spatial regions. Results presented in Table 3 clearly indicate that none of the combination except $p = 15$ and $q = 300$ is able to produce the significant results. Therefore, to obtain a significant feature representation, the seed values of 15 and 300 are recommended for p and q in the proposed method.

3.3 Effect of Feature Selection on Performance of Proposed HILL-VF

Feature selection intends to analyze the high-dimensional data in order to reduce high dimensions, eliminate the irrelevant details and increase the accuracy. High-dimensional feature vector increases the computational cost and also leads to model overfitting. In this section, experiments are conducted on the train sets of Coil-100, KTH-TIPS, KTH-TIPS2-a, and KTH-TIPS2-b to evaluate the generalization performance of proposed HILL-VF by selecting different subsets of features. Principal component analysis (PCA) is exploited for selection of subset of features. Different subsets of features $sub \in \{100, 200, 300, 400, 500\}$ are selected from the extracted set of features. It is important to mention that the train sets are considered with 80% of the data for training and 20% for testing. Table 4 contrasts the performance of proposed method by selecting the aforementioned subsets of features and without feature selection. It is evident from Table 4 that accuracy goes on decreasing from 92.9% to 90.3% when feature subsets are increased from 100 to 500. Likewise, the degradation in performance, i.e., 84.4 to 49.3% is observed on KTH-TIPS when number of feature subsets are increased from 100 to 500. Similar behavior is observed on KTH-TIPS2-a and -b. However, in Table 4, last column presents the accuracy obtained by without selecting any subsets of features. Promising results of 93.25, 91.2, 94.0, and 87.4% are obtained on Coil-100, KTH-TIPS, KTH-TIPS2-a, and -b, respectively, without selecting features. Unfortunately, the proposed method fails to outperform on selected subsets of features by PCA. The selected feature subsets by PCA may reduce the computation time, however, may not always be the optimal features. This may be hypothesized that PCA ignores the variations among the features extracted by HILL-VF and results in significant information loss. Further, the proposed method extracts the compact and salient features which are significant enough to represent the discriminant information of images.

Fast computation especially with the images of higher dimensions is essentially required in real-time applications. The proposed HILL-VF is implemented with Windows 10 operating system. Implementation and experimental evalua-

Table 4 Performance comparison of proposed HILL-VF Accuracy obtained by with and without feature selection on Coil-100, KTH-TIPS, KTH-TIPS2-a, and KTH-TIPS2-b

Datasets	No. of selected components					Without selection
	100	200	300	400	500	
Coil-100	92.9	93.0	92.7	91.0	90.3	**93.25**
KTH-TIPS	84.4	81.7	52.2	49.3	49.3	**91.2**
KTH-TIPS2-a	89.9	87.2	85.5	83.4	81.4	**94.0**
KTH-TIPS2-b	87.4	84.3	81.4	78.1	76.6	**87.4**

Bold values indicate the highest accuracy obtained on each dataset

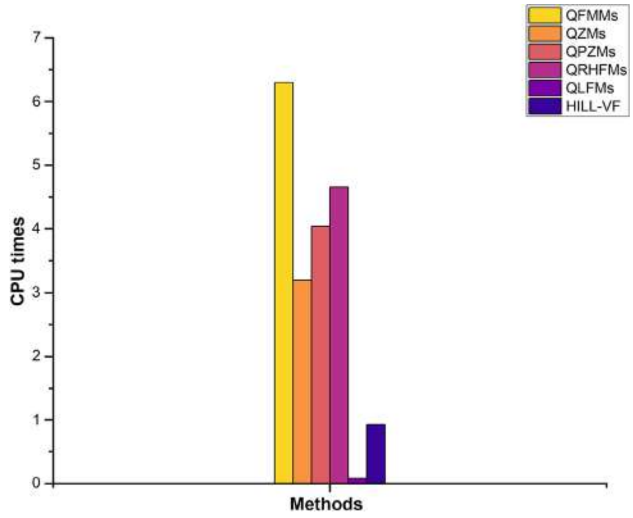

Fig. 7 CPU times for Coil-100 dataset

Table 5 Accuracy obtained on Coil-100, KTH-TIP, KTH-TIPS2-a, and KTH-TIPS2-b by varying the color spaces in FEG-GM

Datasets	Color spaces			
	RGB	YIQ	HSV	**YCbCr**
Coil-100	95.3	95.8	93.72	**96.97**
KTH-TIPS	64.0	66.5	61.25	**87.50**
KTH-TIPS2-a	78.28	96.296	83.50	**97.89**
KTH-TIPS2-b	80.80	73.81	84.68	**93.01**

Bold values indicate the highest accuracy obtained on each dataset

tions are performed with MATLAB (R2017a) on a desktop computer with an 8 GB RAM, and Intel(R) Core(TM) m3-7Y30 CPU @ 1.00GHz 1.61 GHz. Figure 7 contrasts the computation time of proposed method with state-of-the-art methods on Coil-100 dataset. Coil-100 dataset consists of 7200 images of size 128×128. It is worth noting here that computation time is measured as elapsed CPU time not the wall clock time. Most of the results presented in Fig. 7 are obtained from [49]. Figure 7 illustrates that QLFMs consumed the minimum time of 0.0838 s to process an image. It has been observed that the proposed HILL-VF takes 0.925 s (second highest) to process a single image on average. However, the computation time is proportional to the size of feature vector obtained against an image.

3.4 Effect of Different Color Spaces on Performance of Proposed HILL-VF

Experiments are conducted on Coil-100, KTH-TIPS, KTH-TIPS2-a, and -b to evaluate the generalization performance of proposed FEG-GM by varying different color spaces. Table 5 summarizes the accuracy (%) obtained on each of the datasets by varying color spaces. From Table 5, it is noticed that accuracies of 95.3, 64, 78.28, and 80.8% are obtained on Coil-100, KTH-TIPS, KTH-TIPS2-a, and -b, respectively, with RGB color space. In RGB color space, the proposed method is not able to outperform because of the two reasons: (1) it is non-uniform and also illumination sensitive, (2) the channels in RGB color space are highly correlated, therefore, change in intensities leads to the changes in all the three channels. Further, the proposed method is able to achieve the accuracy of 95.8, 66.5, 96.296, and 73.81 on Coil-100, KTH-TIPS, KTH-TIPS2-a, and -b, respectively, with YIQ color space. Since YIQ is more powerful color space compared to the RGB, therefore, a slight improvement in accuracy is observed on Coil-100, KTH-TIPS, KTH-TIPS2-a. The slight improvement in performance is observed because YIQ is perceptually non-uniform. However, the proposed method is unable to yield effective results with YIQ color space because there exists a strong correlation in YIQ color space due to linear transformation. The accuracies of 93.72, 61.25, 83.50, 84% on Coil-100, KTH-TIPS, KTH-TIPS2-a, and -b, respectively, with HSV. The HSV color space is unable to achieve significant results because the rotations in Hue component of HSV with high chroma values could be transformed into a nonexistent color. It is evident from the results reported in Table 5 that the proposed method outperformed with the accuracies of 96.97, 87.50, 97.89, and 93.01% is obtained on Coil-100, KTH-TIPS, KTH-TIPS2-a, and -b, respectively, with YCbCr color space. YCbCr color space outperformed because of the following few reasons: (1) it is able to extract the compact feature representation, (2) also reduces the false alarms, and (3) YCbCr reduces the pixel values in RGB color space to half, i.e., 255 to 127. Therefore, YCbCr color space is able to enhance the recognition performance of proposed FEG-

Table 6 Accuracy (%) obtained by varying kernels of SVM on KTH-TIPS, KTH-TIPS2-a, and KTH-TIPS2-b

Datasets	SVM kernels					
	Linear	Quadratic	Cubic	Fine Gaussian	Medium Gaussian	Coarse Gaussian
KTH-TIPS	**91.2**	85.6	83.4	89.8	69.8	38.8
KTH-TIPS2-a	**94.5**	89.8	94.0	94.0	81.3	51.3
KTH-TIPS2-b	87.3	92.5	92.9	**93.0**	78.3	48.2

Bold values indicate the highest accuracy obtained on each dataset

GM compared to the other color spaces. Further, YCbCr in the proposed HILL-VF is not only able to detect the contour information effectively but also minimizes the effect of outliers.

3.5 Effect of Varying Kernels (SVM) and Different Classifiers on Performance of Proposed HILL-VF

In this section, experiments are performed on the train sets of KTH-TIPS, KTH-TIPS2-a, and -b to evaluate the generalization performance of proposed HILL-VF by varying the kernels of SVM. Major kernels of SVM include linear, quadratic, cubic, fine, medium, and coarse Gaussian. The 80% of the data of train sets of KTH-TIPS, KTH-TIPS2-a, and -b is used for training and rest of the 20% of the data of train sets of KTH-TIPS, KTH-TIPS2-a, and -b is used for testing of proposed method. Results are reported in Table 6. Surprisingly, significant increase in accuracy, i.e., 92.9% is obtained on KTH-TIPS2-b when cubic svm is used. Unfortunately, HILL-VF shows the worst generalization performance of 38.8, 51.3, and 48.2% on KTH-TIPS, KTH-TIPS2-a, and KTH-TIPS2-b, respectively, when course Gaussian kernel is used for classification. Moreover, in Gaussian kernels, the parameters such as sigma and cost factor have substantial effects on the performance. However, parameter estimation in these kernels leads to the increased computational cost. SVM with nonlinear kernels such as fine, medium, and coarse Gaussian may lead to overfitting due to irregularity and noise. The greatest improvement in the accuracy, i.e., of 91.2 and 94.5% is observed on KTH-TIPS and KTH-TIPS2-a when the linear kernel of SVM is used. Linear kernel of SVM is considered as the optimal kernel for classification of proposed method. Moreover, linear kernel of SVM trains faster as compared to the other kernels and parameter optimization is also not required. Table 7 presents the accuracy obtained on KTH-TIPS databases by varying different classification frameworks such as SVM, K-Nearest Neighbor (KNN), decision trees, and linear discriminant analysis (LDA). Table 7 shows that the greatest improvement in performance, i.e., 91.2, 94.5, and 87.3% is observed when SVM with linear kernel is used on KTH-TIPS, KTH-TIPS2-a, and KTH-TIPS2-b, respectively. The accuracy of proposed KNN is slightly decreased to 78, 93.9, and 83.1% when KNN is used for classification. This decline in accuracy is observed because KNN considers the significant features of HILL-VFF as irrelevant features or noise. Moreover, the selection of value of k for KNN is also computationally intensive. Unfortunately, the worst generalization performance of HILL-VF is observed when decision trees and LDA are used for classification. The extracted feature set of HILL-VF may suffer the problem of overfitting when decision tree is used for classification. Since mean values in LDA are shared among classes, therefore, is unable to find a linearly separable boundary between the classes. Moreover, LDA yields the worst accuracies because it overemphasizes the large distances at the expense of increased in inter-class variability.

Table 7 Accuracy (%) obtained by varying classifiers on KTH-TIPS, KTH-TIPS2-a, and KTH-TIPS2-b

Datasets	Classifiers			
	SVM	KNN	Decision trees	LDA
KTH-TIPS	**91.2**	78.0	66.9	68.3
KTH-TIPS2-a	**94.5**	93.9	65.0	61.6
KTH-TIPS2-b	**87.3**	83.1	63.0	60.6

Bold values indicate the highest accuracy obtained on each dataset

3.6 Comparison of Proposed FEG-GM with State-of-the-Art Edge Detection Operators

In the proceeding section, experiments are performed on KTH-TIPS2-b in order to evaluate the generalization performance of proposed FEG-GM and state-of-the-art edge detection operators. Table 8 contrasts the performance (accuracies) obtained by the proposed FEG-GM and state-of-the-art edge detection operators. The performance of the FEG-GM is compared with the edge detection operators such as Prewitt, Average, Gaussian, and Difference of Gaussian (DoG). Since the proposed FEG-GM fuses Sobel and Scharr edge detection operator in YCbCr color space. It is important to mention that size of Gaussian kernel is set to 3, whereas in DoG, the size of kernels is set to 5 and 3, respectively. In Table 8, 1st row contains the results obtained by fusing Scharr with Prewitt, Average, Gaussian, and DoG in YCbCr color space. In

Table 8 Accuracy obtained on KTH-TIPS2-b by proposed FEG-GM and state-of-the-art edge detection operators

Operators	Prewitt	Average	Gaussian	Difference of Gaussian (DoG)	Proposed FEG-GM
Scharr	67.0	79.09	79.29	86.53	**93.01**
Sobel	69.7	78.03	78.80	88.2	

Bold value indicates the highest accuracy obtained on each dataset

addition, 2nd row in Table 8 contains the results obtained by fusing Sobel edge detection operator with Prewitt, Average, Gaussian, and DoG. It has been observed from the results reported in Table 8 that none of combination except Scharr and Sobel is able to increase the performance of proposed method. It is evident from the results reported in Table 8, that proposed FEG-GM has achieved a significant high accuracy of 93.01% on KTH-TIPS2-b. DoG reduces the overall contrast of the image, therefore, is not able to extract the salient feature information of the image. The Gaussian and Average edge detection operators smoothen the inherent edge details, consequently, significant feature information may be lost. Prewitt operator is sensitive to noise and also produces inaccurate edge details when the magnitude of extracted edges decreases. It is evident from the results reported in Table 8 that none of the combinations except Scharr and Sobel is able to increase the recognition accuracy. The diffusion equation in proposed FEG-GM outperforms the state-of-the-art edge detection operators because of the following few reasons: (1) preserves the inherent textural details, (2) preserves the variable viewpoints, (3) preserves the chromatic information by assigning higher weights to the chroma components of the images, and 4) it is also invariant to rotations.

3.7 Comparison of Proposed HILL-VF with the State-of-the-Art Methods

In this section, we have compared the results obtained by proposed method (HILL-VF) with state-of-the-art methods of salient feature extraction. The performance of proposed HILL-VF on KTH-TIPS is compared with state-of-the-art methods such as SSELBP [11], BIF [50], COV-LBPD [12], WLD [29], SIFT + IFV, DMD +IFV [25], BIGD + IFV [3], LES+GTD [17], LDRP [51], LTCP [52], HL-GP [16], and ARCS [15]. Similarly, the performance of proposed HILL-VF on KTH-TIPS2-a is compared with state-of-the-art methods such as LBP [25], scLBP [13], MLEP [53], MS-BIF [50], VZ-MRS [25], LHS [28], SIFT+IVF [25], NDV [24], BIGD+VLAD [3], BIGD + IFV [3], and HL-GP [16], LES + GTD [17]. Likewise, the performance of proposed HILL-VF on KTH-TIPS2-b is compared with the state-of-the-art methods such as MC-SBP [54], CDL [55], S_H-SVM [56], Timofte [57], FV-Alexnet [26], FV-VGGM [26], FV-VGGVD [26], BIGD+VLAD [3], BIGD + IFV [3], HL-GP [16], LES+GTD [17], and ARCS [15]. Similarly, the performance on Coil-100 is compared with the state-of-the-art methods such as Gaussian-CKN [58], Cos-CKN-RF [33], Cos-CKN-OP [33], PCANet [34], RNPCANet [34], SGEF [31], SIFT [30], HMAX [30] , RIK [59], SWOVF [32], LCR [23], LCR + d [23], and LLMC + d [23]. Tables 9, 10, 11, and 12 present the performance comparison of the proposed HILL-VF against state-of-the-art methods on KTH-TIPS, KTH-TIPS2-a, and b, and Coil-100, respectively. It has been observed from the results reported in Tables 10, 11, and 12 that the proposed HILL-VF outperformed the state-of-the-art methods with the accuracy of 97.89, 93.01, and 96.97% on KTH-TIPS2-a and -b, and Coil-100, respectively. Unfortunately, the proposed HILL-VF is not able to increase the classification accuracy on KTH-TIPS because the proposed HILL-VF feature representation is not enough to represent the discriminant textural characteristics of KTH-TIPS. Another reason for lower accuracy on KTH-TIPS is due to evaluation protocol followed in our proposed method. In the proposed method, only 3 samples are used for training and only one sample is used for testing. However, the proposed HILL-VF outperformed the two challenging benchmarks of KTH-TIPS2 compared the state-of-the-art methods. SIFT [25] constructs the histogram for each key point by binning the gradient orientations and magnitudes. However, the quantized orientation information results in the loss of discriminant information. Moreover, the BIGD descriptor [3] incorporates both intensity and gradient information at multiple orientations without quantization. However, the computations of BIGD descriptor lead to an increased computational cost. Train features extracted by FV-VGGM [26] and FV-VGGVD [26] for colored images lack the geometric invariance, therefore, unable to recognize textures with high variations. The pre-trained model on weights of ImageNet in FV-Alexnet [26] increases the computational cost. LBP and its variants [11–13,25] consider on the central pixel as most significant. Moreover, these methods consider only short-ranged local regions to achieve the rotation invariance. BIF [50] preserves the scale information while increasing the dimensions of representations. WLD [29] considers only 4 pixels at horizontal and vertical positions to compute the orientation information. LHS lacks naive and coarse encoding, therefore, is unable to represent the significant feature information. In SIFT and HMAX [30] losses the necessary chromatic information required for texture classification. Kernels in Gaussian-CKN [58], Cos-CKN-RF [33], and Cos-CKN-OP [33] need to be optimized.

Table 9 Performance comparison of proposed HILL-VF with state-of-the-art methods on KTH-TIPS

Methods	Accuracy (%)
SSELBP [11]	98.1
BIF [50]	98.5
COV-LBPD [12]	98.0
WLD [29]	91.1
SIFT+IFV [25]	97.3
DMD+IFV [25]	97.6 ± 1.6
BIGD+IFV [3]	**98.8** ± **1.1**
LES+GTD [17]	83.2
LDRP [51]	77.72
LTCP [52]	80.9
HL-GP [16]	84.5
ARCS [15]	71.6
Proposed	87.5

Bold values indicate the highest accuracy obtained on each dataset

Table 10 Performance comparison of proposed HILL-VF with state-of-the-art methods on KTH-TIPS2-a

Methods	Accuracies (%)
BIGD+VLAD [3]	81.2 ± 2.5
BIGD+IFV [3]	81.3 ± 3.6
LBP [25]	69.80
scLBP [13]	78.39
MLEP [53]	75.57
MS-BIF [50]	71.56
VZ-MRS [25]	62.35
LHS [28]	73.0
SIFT+IVF [25]	76.6
NDV [24]	77.1
HL-GP [16]	82.6
LES + GTD [17]	78.9
Proposed	**97.89**

Bold values indicate the highest accuracy obtained on each dataset

Since contrast-based feature information is more resilient to illuminations changes compared to the intensities, the FEG-GM in the proposed HILL-VF incorporates the variations in contrast locally and therefore improves distinctiveness. The proposed HILL-VF outperformed the state-of-the-art methods because of the following reasons: (1) The proposed HILL-VF fuses the inherent edge details and textural characteristics of images while minimizing the noisy artifacts and preserving rotation invariance. (2) Gabor feature extracts the discriminant orientation information of dominant textures, and (3) Micro-feature maps in the proposed HILL-VF balance the computational complexity and dimensionality constraints.

Table 11 Performance comparison of proposed HILL-VF with state-of-the-art methods on KTH-TIPS2-b

Methods	Accuracy (%)
MC-SBP [54]	71.6
CDL [55]	76.3
S_H-SVM [56]	80.1
Timofte [57]	66.3
FV-Alexnet [26]	77.9
FV-VGGM [26]	79.9
FV-VGGVD [26]	88.2
BIGD+VLAD [3]	81.4 ± 3.1
BIGD + IFV [3]	82.7 ± 4.5
HL-GP [16]	77.3
LES+GTD [17]	83.7
ARCS [15]	78.8
Proposed	**93.01**

Bold values indicate the highest accuracy obtained on each dataset

Table 12 Performance comparison of proposed HILL-VF on Coil-100 with state-of-the-art methods

Methods	Accuracy
Gaussian-CKN [58]	90.25 ± 0.95
Cos-CKN-RF [33]	90.02 ± 1.35
Cos-CKN-OP [33]	90.15 ± 1.12
PCANet [34]	88.23 ± 0.76
RNPCANet [34]	90.35 ± 1.15
AWFKM [43]	61.71 ± 1.90
SGEF [31]	92.4
SIFT [30]	87.2
HMAX [30]	77.0
RIK [59]	95.8
SWOVF [32]	94.43
LCR [23]	76.63
LCR + d [23]	91.5
LLMC + d [23]	80.06
Proposed	**96.97**

Bold values indicate the highest accuracy obtained on each dataset

4 Conclusion

In this paper, a feature descriptor named histogram of low-level visual features (HILL-VF) is proposed for extraction of salient features. In the proposed HILL-VF, directional derivative filters of Sobel and Scharr operators in YCbCr color space are fused to extract the contrast-based features.

The diffusion equation in the proposed method fuses the extracted edge information by assigning higher weights to the chroma component of YCbCr. Moreover, Gabor features in the proposed HILL-VF effectively characterize the phase information. Micro-feature maps of contrast- and phase-based information are then generated by encoding the contrast and phase information into pre-defined levels based on the selected seed values. These encoded micro-maps are then represented through a 2-D histogram. Yet, promising results are obtained on the standard benchmarks of Coil-100, KTH-TIPS, KTH-TIPS2-a and b. The proposed HILL-VF outperformed the state-of-the-art methods of object recognition and texture feature extraction. The discriminant set of features in the proposed HILL-VF is effective against rotations and viewpoints variations. However, is sensitive to image scaling. The future work will incorporate the shape-based features along with contrast- and phase-based for extraction of significant feature information.

References

1. Hou, J.; Yang, S.; Lin, W.; Zhao, B.; Fang, Y.: Learning image aesthetic assessment from object-level visual components. arXiv:2104.01548 (2021)
2. Wang, T.; Borji, A.; Zhang, L.; Zhang, P.; Lu, H.: A stagewise refinement model for detecting salient objects in images. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4019–4028 (2017)
3. Hu, Y.; Wang, Z.; AlRegib, G.: Texture classification using block intensity and gradient difference (bigd) descriptor. Signal Process. Image Commun. **83**, 115770 (2020)
4. Nazir, A.; Nazir, K.: An efficient image retrieval based on fusion of low-level visual features. arXiv:1811.12695 (2018)
5. Ojala, T.; Pietikainen, M.; Harwood, D.: Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In: Proceedings of 12th international conference on pattern recognition, vol. 1. IEEE, pp. 582–585 (1994)
6. Shabat, A.M.; Tapamo, J.R.: A comparative study of local directional pattern for texture classification. In: 2014 World Symposium on Computer Applications an d Research (WSCAR). IEEE, pp. 1–7 (2014)
7. Kaya, Y.; Ertuğrul, Ö.F.; Tekin, R.: Two novel local binary pattern descriptors for texture analysis. Appl. Soft Comput. **34**, 728–735 (2015)
8. Ojala, T.; Pietikainen, M.; Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
9. Ojala, T.; Maenpaa, T.; Pietikainen, M.; Viertola, J.; Kyllonen, J.; Huovinen, S.: Outex-new framework for empirical evaluation of texture analysis algorithms. In: Object Recognition Supported by User Interaction for Service Robots, vol. 1. IEEE, pp. 701–706 (2002)
10. Guo, Z.; Zhang, L.; Zhang, D.; Zhang, S.: Rotation invariant texture classification using adaptive lbp with directional statistical features. In: 2010 IEEE International Conference on Image Processing. IEEE, pp. 285–288 (2010)
11. Hu, Y.; Long, Z.; AlRegib, G.: Scale selective extended local binary pattern for texture classification. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 1413–1417 (2017)
12. Hong, X.; Zhao, G.; Pietikäinen, M.; Chen, X.: Combining lbp difference and feature correlation for texture description. IEEE Trans. Image Process. **23**(6), 2557–2568 (2014)
13. Ryu, J.; Hong, S.; Yang, H.S.: Sorted consecutive local binary pattern for texture classification. IEEE Trans. Image Process. **24**(7), 2254–2265 (2015)
14. Liu, Y.-Y.; Chen, M.; Ishikawa, H.; Wollstein, G.; Schuman, J.S.; Rehg, J.M.: Automated macular pathology diagnosis in retinal oct images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. Med. Image Anal. **15**(5), 748–759 (2011)
15. Ruichek, Y.; et al.: Attractive-and-repulsive center-symmetric local binary patterns for texture classification. Eng. Appl. Artif. Intell. **78**, 158–172 (2019)
16. Hazgui, M.; Ghazouani, H.; Barhoumi, W.: Genetic programming-based fusion of hog and lbp features for fully automated texture classification. In: The Visual Computer, pp. 1–20 (2021)
17. Ghazouani, H.; Barhoumi, W.: Genetic programming-based learning of texture classification descriptors from local edge signature. Expert Syst. Appl. **161**, 113667 (2020)
18. Song, T.; Xin, L.; Gao, C.; Zhang, T.; Huang, Y.: Quaternionic extended local binary pattern with adaptive structural pyramid pooling for color image representation. Pattern Recogn. **115**, 107891 (2021)
19. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P.: Brief: Binary robust independent elementary features. In: European Conference on Computer Vision. Springer, pp. 778–792 (2010)
20. Chahi, A.; Ruichek, Y.; Touahni, R.; et al.: Local directional ternary pattern: a new texture descriptor for texture classification. Comput. Vis. Image Underst. **169**, 14–27 (2018)
21. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
22. Bay, H.; Tuytelaars, T.; Van Gool, L.: Surf: Speeded up robust features. In: European Conference on Computer Vision. Springer, pp. 404–417 (2006)
23. Yang, T.; Li, C.-G.: Local convex representation with pruning for manifold clustering. In: 2019 IEEE Visual Communications and Image Processing (VCIP). IEEE, pp. 1–4 (2019)
24. Zhang, W.; Zhang, W.; Liu, K.; Gu, J.: A feature descriptor based on local normalized difference for real-world texture classification. IEEE Trans. Multimedia **20**(4), 880–888 (2017)
25. Mehta, R.; Egiazarian, K.: Texture classification using dense microblock difference. IEEE Trans. Image Process. **25**(4), 1604–1616 (2016)
26. Cimpoi, M.; Maji, S.; Vedaldi, A.: Deep filter banks for texture recognition and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3828–3836 (2015)
27. Liu, L.; Fieguth, P.; Kuang, G.; Zha, H.: Sorted random projections for robust texture classification. In: 2011 International Conference on Computer Vision. IEEE, pp. 391–398 (2011)
28. Sharma, G.; ul Hussain, S.; Jurie, F.: Local higher-order statistics (lhs) for texture categorization and facial analysis. In: European Conference on Computer Vision. Springer, pp. 1–12 (2012)
29. Chen, J.; Shan, S.; He, C.; Zhao, G.; Pietikäinen, M.; Chen, X.; Gao, W.: Wld: a robust local image descriptor. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1705–1720 (2009)
30. Elazary, L.; Itti, L.: A Bayesian model for efficient visual search and recognition. Vision. Res. **50**(14), 1338–1352 (2010)
31. Sudhakaran, S.; James, A.P.: Sparse distributed localized gradient fused features of objects. Pattern Recogn. **48**(4), 1538–1546 (2015)

32. Shabbir, S.; Majeed, N.; Dawood, H.; Dawood, H.; Xiu, B.: Integrating the local patches of weber orientation with sparse distribution method for object recognition. Arab. J. Sci. Eng. **44**(4), 3603–3618 (2019)
33. Mohammadnia-Qaraei, M.R.; Monsefi, R.; Ghiasi-Shirazi, K.: Convolutional kernel networks based on a convex combination of cosine kernels. Pattern Recogn. Lett. **116**, 127–134 (2018)
34. Qaraei, M.; Abbaasi, S.; Ghiasi-Shirazi, K.: Randomized nonlinear pca networks. Inf. Sci. **545**, 241–253 (2021)
35. Huang, Z.: Cn-lbp: complex networks-based local binary patterns for texture classification. arXiv:2105.06652 (2021)
36. Liu, X.; Shan, C.; Zhang, Q.; Cheng, J.; Xu, P.: Compressed wavelet tensor attention capsule network. In: Security and Communication Networks, vol. 2021 (2021)
37. Tao, Z.; Wei, T.; Li, J.: Wavelet multi-level attention capsule network for texture classification. IEEE Signal Process. Lett. **28**, 1215–1219 (2021)
38. Song, T.; Feng, J.; Wang, Y.; Gao, C.: Color texture description based on holistic and hierarchical order-encoding patterns. In: 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, pp. 1306–1312 (2021)
39. Nsimba, C.B.; Levada, A.L.: Combining deep and manifold learning for nonlinear feature extraction in texture images. In: 2020 28th European Signal Processing Conference (EUSIPCO). IEEE, pp. 1552–1555 (2021)
40. Fradi, H.; Fradi, A.; Dugelay, J.-L.: Multi-layer feature fusion and selection from convolutional neural networks for texture classification. In: VISIGRAPP (4: VISAPP), pp. 574–581 (2021)
41. Shi, F.; Guo, J.; Zhang, H.; Yang, S.; Wang, X.; Guo, Y.: Glavnet: global-local audio-visual cues for fine-grained material recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14 433–14 442 (2021)
42. Song, T.; Feng, J.; Wang, S.; Xie, Y.: Spatially weighted order binary pattern for color texture classification. Expert Syst. Appl. **147**, 113167 (2020)
43. Nie, F.; Chang, W.; Li, X.; Xu, J.; Li, G.: Adaptive feature weight learning for robust clustering problem with sparse constraint. In: ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 3125–3129 (2021)
44. Kovesi, P.D.: Matlab and octave functions for computer vision and image processing. In: Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia, vol. 147, p. 230. http://www.csse.uwa.edu.au/~pk/research/matlabfns (2000)
45. Kim, J.; Um, S.; Min, D.: Fast 2d complex Gabor filter with kernel decomposition. IEEE Trans. Image Process. **27**(4), 1713–1722 (2017)
46. Marée, R.; Geurts, P.; Piater, J.; Wehenkel, L.: Random subwindows for robust image classification. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1. IEEE, pp. 34–40 (2005)
47. Fritz, M.; Hayman, E.; Caputo, B.; Eklundh, J.-O.: The kth-tips database (2004)
48. Mallikarjuna, P.; Targhi, A.T.; Fritz, M.; Hayman, E.; Caputo, B.; Eklundh, J.-O.: The kth-tips2 database. In: Computational Vision and Active Perception Laboratory, pp. 1–10. Stockholm (2006)
49. Hosny, K.M.; Darwish, M.M.: Invariant color images representation using accurate quaternion Legendre–Fourier moments. Pattern Anal. Appl. **22**(3), 1105–1122 (2019)
50. Crosier, M.; Griffin, L.D.: Using basic image features for texture classification. Int. J. Comput. Vision **88**(3), 447–460 (2010)
51. Dubey, S.R.: Local directional relation pattern for unconstrained and robust face retrieval. In: Multimedia Tools and Applications, vol. 78, no. 19, pp. 28063–28088 (2019)
52. Arya, R.; Vimina, E.R.: Local triangular coded pattern: a texture descriptor for image classification. IETE J. Res. (2021). https://doi.org/10.1080/03772063.2021.1919222
53. Zhang, J.; Liang, J.; Zhao, H.: Local energy pattern for texture classification using self-adaptive quantization thresholds. IEEE Trans. Image Process. **22**(1), 31–42 (2012)
54. Nguyen, T.P.; Vu, N.-S.; Manzanera, A.: Statistical binary patterns for rotational invariant texture classification. Neurocomputing **173**, 1565–1577 (2016)
55. Wang, R.; Guo, H.; Davis, L.S.; Dai, Q.: Covariance discriminative learning: a natural and efficient approach to image set classification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 2496–2503 (2012)
56. Harandi, M.; Salzmann, M.; Porikli, F.: Bregman divergences for infinite dimensional covariance matrices. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1003–1010 (2014).
57. Timofte, R.; Van Gool, L.: A training-free classification framework for textures, writers, and materials. In: BMVC, vol. 13, p. 14 (2012)
58. Mairal, J.; Koniusz, P.; Harchaoui, Z.; Schmid, C.: Convolutional kernel networks. arXiv:1406.3332 (2014)
59. Hamsici, O.C.; Martinez, A.M.: Rotation invariant kernels and their application to shape analysis. IEEE Trans. Pattern Anal. Mach. Intell. **31**(11), 1985–1999 (2008)