IEEE Access

Received September 16, 2021, accepted September 25, 2021, date of publication September 29, 2021,
date of current version October 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3116225

Directional Magnitude Local Hexadecimal Patterns: A Novel Texture Feature Descriptor for Content-Based Image Retrieval

AYESHA KHAN[1], ALI JAVED[2], (Member, IEEE),
MUHAMMAD TARIQ MAHMOOD[3], (Senior Member, IEEE),
MUHAMMAD HAMZA ARIF KHAN[1], AND IK HYUN LEE[4], (Member, IEEE)

[1]Department of Software Engineering, University of Engineering and Technology Taxila, Taxila 47050, Pakistan
[2]Department of Computer Science, University of Engineering and Technology Taxila, Taxila 47050, Pakistan
[3]Future Convergence Engineering, School of Computer Science and Engineering, Korea University of Technology and Education, Cheonan-si 31253, Republic of Korea
[4]Department of Mechatronics Engineering, Korea Polytechnic University, Siheung-si, Gyeonggi-do 15073, Republic of Korea

Corresponding authors: Ali Javed (ali.javed@uettaxila.edu.pk) and Ik Hyun Lee (ihlee@kpu.ac.kr)

This work was supported in part by BK21 FOUR, in part by the National Research Foundation of Korea (NRF) funded by the Korean Government, Ministry of Science and ICT (MSIT) under Grant 2021R1F1A1050022, and in part by the Priority Research Centers Program funded by the Ministry of Education under Grant 2017R1A6A1A03015562.

ABSTRACT Social media platforms such as Twitter, Facebook, and Flicker, and the evolution of digital image capturing devices have resulted in the generation of a massive number of images. Thus, we experienced an exponential growth in digital image repositories in the last decade. Content-based image retrieval (CBIR) has been extensively employed to reduce the dependency on textual annotations for image searching. Effective feature descriptor is mandatory to retrieve the most relevant images from the repository. Additionally, CBIR methods often experience the semantic gap problem, which must also be addressed. In this paper, we propose a novel texture descriptor, Directional Magnitude Local Hexadecimal Patterns (DMLHP), based on the texture orientation and magnitude to retrieve the most relevant images. The objective of the proposed feature descriptor is to examine the relationship between the neighboring pixels and their adjacent neighbors based on texture orientation and magnitude. Our DMLHP texture descriptor is capable of capturing the texture and semantic information of the images effectively with the same visual content. Furthermore, the proposed method employs a learning-based approach to lessen the semantic gap problem and to improve the understanding of the contents of query images to retrieve the most relevant images. The presented descriptor provides remarkable results by achieving the average retrieval precision (ARP) of 66%, 92%, 83%, average retrieval recall (ARR) of 66%, 92%, 83%, average retrieval specificity (ARS) of 99%, 99%, 76%, and average retrieval accuracy (ARA) of 98%, 99%, 85% on the AT&T, MIT Vistex, and Brodatz Texture image repositories, respectively. Our experiments reveal that the proposed DMLHP descriptor achieves far better performance, i.e., 95% on AT&T, 92% on BT, and 99% on MIT Vistex, when used with a learning-based approach over a non-learning-based approach (similarity measure). Experimental results show that the proposed texture descriptor outperforms state-of-the-art descriptors such as LNIP, LTriDP, LNDP, LDGP, LEPSEG, and CSLBP for CBIR.

INDEX TERMS Content-based image retrieval, orientation-based pattern, magnitude-based pattern, similarity-based approach, learning-based approach.

I. INTRODUCTION

The tremendous evolution of digital cameras and the Internet has resulted in the generation of a massive amount of multimedia content over the last couple of decades. Additionally, social media platforms also amplified this huge collection of images and videos in cyberspace. Effective storage and retrieval of the multimedia content is challenging [1] due to huge multimedia repositories, semantic understanding of media in the presence of complex backgrounds, and the

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

FIGURE 1. Related visual textural appearance of two sample images belonging to MIT Vistex image repository.

FIGURE 2. Close visual and semantic appearance between two different semantic categories of MIT Vistex image repository.

semantic gap between computers and humans. Therefore, there is a need to develop more effective image retrieval systems that are robust to the above-mentioned limitations. Text-based image retrieval (TBIR) is commonly employed for retrieval where the search is based on automatic and manual annotation of images. TBIR-based techniques use similar image text to search and retrieve relevant images from the repositories. It has been observed that the description of the texture images using the text becomes difficult at times because different users employ distinct keywords for annotation. This reveals the limitation of text descriptors being subjective, which results in low retrieval accuracy. Additionally, it is difficult to express the entire visual contents of an image in words, so, TBIR may produce irrelevant results. To overcome the limitations of TBIR systems, researchers introduced the concept of a content-based image retrieval system. CBIR addresses the limitation of TBIR, as CBIR does not need manual annotation to retrieve visually similar images [2]. A CBIR system is based on the visual contents of the images described in the low-level features, that is, texture, shape, color, and spatial locations to build the feature repository. CBIR gives more prominent attention to the local and global information such as the color, texture, and shape of an image. In the CBIR system, the image is provided as an input query instead of feeding the textual query.

We have witnessed a tremendous evolution in CBIR systems over the years to solve many retrieval issues in large repositories. However, some open issues in the image retrieval domain must still be addressed. Working on the semantic gap for image repositories of large sizes is still a challenging problem. The semantic gap refers to the limitation of low-level feature representation of images in describing the actual visual perception of the image, that is, human semantics. Visual similarity belonging to two different semantic categories reduces the performance of the CBIR system because images that have no semantic relation are retrieved. Fig. 1 represents two sample images from the MIT Vistex image repository, which are similar in terms of visual perception and image semantics. The degree of similarity in terms of visual content among these images are the same due to low-level features, that is, color, shape, and texture as well as high-level image semantics. Actually, these two images belong to different semantic classes where the first image belongs to the ''Food'' class, whereas, the second image belongs to the ''Fabric'' class. The CBIR system may retrieve irrelevant images due to common visual contents like color, shape, and texture. Similarly, we can observe from Fig. 2 that the similar-looking image ''GroundWaterCity'' is returned against the inquiry image ''Clouds''. These images seem to be semantically related due to the presence of common visual contents such as the sky and water, but they actually belong to two different semantic groups. These images describe the semantic gap issue between low-level feature representations and high-level user semantics.

Existing local texture descriptors such as local binary patterns (LBP), local ternary patterns (LTP), and others compute limited directional information and ignore the magnitude information. The retrieval performance of these texture descriptors can be further enhanced by capturing more directional and higher-magnitude information from the neighboring pixels. This observation motivated us to develop a novel texture descriptor by considering the aforementioned limitations of existing texture descriptors. The proposed DMLHP feature descriptor is capable of effectively capturing the characteristics of an image based on texture orientation and magnitude. The texture orientation-based pattern (TOBP) encodes more detailed discriminative information in sixteen directions of the neighborhood region, using first-order derivatives. The amount of intensity variation is computed at 0°, 45°, 90°, and 135° directions to illustrate maximum changes by, using a texture magnitude-based pattern, (TMBP). To represent the image features, the two patterns are formulated (i.e., texture orientation is extracted from TOBP, and magnitude information is extracted from TMBP). Each image is represented as a fusion of texture orientation and magnitude patterns. The histograms are obtained by concatenating both patterns, and the classifier is trained using the proposed features. The learning-based approach is required to bridge the semantic gap problem in CBIR systems. The proposed descriptor is also used with a learning-based approach to enhance the CBIR accuracy. The results show that the proposed texture descriptor achieves high retrieval and classification performance in both the traditional (similarity) as well as the learning approaches.

The major contributions of this work are as follows:

1) We propose a novel texture descriptor that is robust to noise, color, pose variations, and a variety of natural and artificial regular textures.
2) We present an effective features fusion of texture orientation and texture magnitude to extract more discriminative information from the input image.
3) Our method addresses the issue of the semantic gap between high-level semantics and local features.

The upcoming sections of the article are outlined as follows: In Section II, we provided a comprehensive overview of existing state-of-the-art CBIR systems. The detailed procedure of the proposed texture descriptor is described in Section III. In Section IV, we present the performance evaluation of the proposed CBIR method in detail. Finally, Section V concludes the proposed work.

II. RELATED WORK

This section provides a critical investigation of existing state-of-the-art CBIR techniques. Effective image representation is required for accurate and highly relevant image retrieval. For this purpose, researchers have proposed different local and global descriptors over the last few decades. Although global descriptors are more computationally efficient than the local descriptors, they are unable to perform well under certain conditions, e.g., scaling, rotation, and viewpoint changing. On the other hand, local features are more robust under the above-mentioned limitations, and are better able to capture the complex texture information in the images [3]. Local descriptors have recently received considerable attention in the community of image retrieval because of their robustness to illumination changes, noise, and variations in pose. Different local feature descriptors like BRISK [4], SURF [5], SIFT [6], and HOG [7] were used for effective feature extraction. SIFT has been widely employed as a local feature descriptor because of its invariance to scale, rotation, transition in lighting, and 3D camera perspective properties. Various drawbacks are associated with the original SIFT descriptor such as high computational cost and slow processing. SURF improves the computational efficiency of features computation as compared to SIFT. SURF performs well on blur and rotated images, but does not perform well when illumination or viewpoint change. The enhanced SURF keypoint descriptor was introduced in [8], which combined the color information to improve the accuracy of keypoint-based descriptors. HOG [7] is an effective feature descriptor used for object detection. The HOG descriptor finds the edges and shape of an object in a localized portion. Therefore, it is invariant against the photometric and local geometric variations.

Many extensions were made for conventional local binary pattern descriptors such as multichannel adder and decoder local binary pattern (MadLBP), modified local binary pattern, multiscale local binary pattern, pyramid local binary pattern, and local derivative radial pattern (LDRP). The dimensionality of the conventional patterns is increased by concatenating the LBP histograms derived from each channel. MaLBP and MdLBP were introduced in [9] to obtain the LBP's cross-channel co-occurrence details. However, the traditional LBP dimensionality issue is reduced by using the multi-channel decoded-based LBP fusion scheme. Mamta *et al.* [10] presented a variation of the local binary pattern that reduced the execution time by considering only four local neighbors information. In [11], LBP is used as a texture descriptor to extract surface textures of multi-object images in face recognition applications; since the majority of the approaches employed a fixed scale LBP that was not effective to extract the structural features. Manish *et al.* [12] used the LBP to extract texture features and decomposed the LBP image into multiple scales through DWT. Some of the multi-scale LBP methods consider only boundary pixels while ignoring the inner block information. Prashant *et al.* [13] introduced multi-scale LBP, which efficiently extracts dominant features at multiple blocks of 3×3, 5×5, and 7×7 windows. In [14], the conventional LBP is extended to the spatial pyramid domain to improve computational efficiency. In [15], LDRP used multi-level encoding in various directions to overcome the problem of information loss of existing local patterns. Cevik *et al.* [49] introduced a directional local gradient-based descriptor (DLGBD) for face recognition. DLGBD calculated the relationship of the reference pixel with its neighboring pixels based on ternary encoding (9 states) by considering both predecessor and successor pixels. Moreover, this approach examined the relationship of neighboring pixels with their adjacent neighbors by utilizing the mean information of the successor and predecessor adjacent neighbors. However, this descriptor is unable to extract spatial structure information in various directions. To overcome the problem of missing information, this descriptor can be further improved by considering texture orientation, texture magnitude, and edge orientation-based information in different directions.

As we experience an extensive range of colors and textures in natural images, feature descriptors must effectively capture such rich information of colors and textures. In [16], color histogram features, texture discrete wavelet transform features, and edge histogram (EH) features were combined to create a new CBIR system. Li *et al.* [17] introduced neighbor intensity co-occurrences local ternary pattern (NI-CLTP) for image representation. The Gabor filters were integrated with NI-CLTP to extract the texture details at different scales and directions. By integrating the different cross channel combinations of HSV, Megha *et al.* [18] implemented a multi-channel local ternary pattern to capture texture-chromatic characteristics. In [19], color features were used in combination with a local directional pattern for CBIR. Feature normalization was performed before integrating the histograms of color and LDP features. In [50], Rehan *et al.* fused the color and texture features where CM was employed to extract the color features, whereas Gabor wavelet and discrete wavelet transforms were used to extract the texture data. Moreover, color and edge directivity descriptor (CEDD) was employed

to improve the feature representation. To overcome the challenge of extracting features from the most visible objects, Rehan *et al.* [51] presented a CBIR method to extract the texture and color features of the most salient objects. Texture features were extracted through bandelet transform and fused with the color histogram features in the HSV domain to improve the feature space capability. SVM was used to determine the semantic association.

In CBIR, several descriptors have been introduced to combine the benefits of the local visual contents of the images. To integrate color spatial structure information of an image with local visual contents, low-level local features were used with the color features in [20]. Low-level salient visual features were extracted using the accelerated segment test feature descriptor. Then color features were extracted and segmented using the L*a*b* color space model. The fusion of these features performed well compared to the existing CBIR methods. Vibhav *et al.* [21] extracted color and shape features, using the color moments (CM) and invariant moments (IM) respectively. The feature fusion of color and shape gives significant precision in comparison to the previous CBIR frameworks. Prashant *et al.* [22] used a local ternary wavelet gradient pattern to capture shape and texture characteristics in images at various resolutions. Since the descriptor computes the features at different image scales, this feature extraction process makes it computationally more expensive. The existing descriptors such as structure element histogram (SEH) and color difference histogram (CDH) integrated the texture with color features and have certain limitations. SEH was not rotation invariant, and thus did not employ the idea of symmetry with the center pixel, whereas CDH's scale and rotation invariance properties are affected by the usage of strongly correlated adjacent data. The rotation and scale-invariant hybrid descriptor [52] (RSHD) was introduced to address the issues associated with both descriptors by incorporating the rotation invariant structure elements. RSHD fused the texture and color features. To accurately describe an image, the fusion of all local visual features has both advantages and drawbacks. However, the aforementioned integrated low-level local visual features are greatly improved in terms of high-level semantic meanings, but are computationally complex in terms of retrieval time due to the large dimension of the feature vector.

Color-based descriptors are unable to achieve reasonable accuracy for images, where multiple objects have similar colors [20]. Shape features often fail to extract more information in the presence of noise, occultation, and non-rigid deformations. As a result, images captured from a single point of view are sensitive to it. The content of many real-world objects, such as clouds, trees, valley, food, and fabric, can be better described using texture. As a result, texture plays an important role in describing high-level semantics for feature extraction. Many texture-based descriptors have been introduced in the CBIR domain. Bhavana *et al.* [23] presented a computationally efficient descriptor called the dual cross pattern. A local DCP sampling was done in eight directions to extract each pixel information twice. DCP has improved the grouping approach by joint Shannon entropy to reduce information loss. Ranjit *et al.* [24] presented a fused feature descriptor by integrating the threshold local binary AND pattern and local adjacent neighborhood average difference pattern for CBIR. This method provides higher accuracy over other local features-based methods (e.g. SURF, SIFT, LBP); however, this also comes with increased features computation cost. Satya *et al.* [25] implemented a local mean differential excitation pattern (LMDeP) to extract features from the images of noisy texture. The LMDeP descriptor encoded the correlation between center pixels and their neighbors through differential excitation, rather than a gray level difference. The contourlet tetra pattern was introduced in [26] for efficient image retrieval. This descriptor employed the contourlet transform to determine the direction instead of using spatial first-order derivatives to evaluate the direction. The contourlet transform used the laplacian pyramid and directional filter bank to decompose an image in multi-directional scales. Faiq *et al.* [27] introduced an amended LTP version called the extended local ternary pattern for CBIR. This descriptor presented an automatic threshold calculation mechanism for ternary code generation as compared to a static threshold used in LTP. Existing patterns such as CSLBP and CSLTP encoded a restricted number of pixels, which makes them unable to work effectively in an unconstrained environment. The feature lengths of these descriptors, on the other hand, are a main source of concern. In [53], the center symmetric quadruple pattern (CSQP) has been introduced to address these drawbacks by encoding the pixels in a large neighborhood in diagonally opposite quadruple space. CSQP generated an 8-bit pattern from 16 pixels in the immediate vicinity. CSQP has a clear advantage in terms of computational complexity. The primary issue associated with the existing intensity order-based descriptors such as local intensity order pattern (LIOP) and soft ordinal spatial intensity distribution (soft OSID) is to increase the size of the descriptor with a little increase in the neighboring pixels. Shiv *et al.* [54] introduced intensity order-based interleaved local descriptor (IOLD), which is based on the division of N neighbors into k interleaved sets has proven to substantially achieve low time complexity while maintaining reasonable performance under noisy conditions. The internal spatial structural knowledge of an image is represented by texture features, which are more descriptive of high-level image semantic observations than color features. However, existing texture descriptors have certain limitations such as they are sensitive to noise, are less efficient, fail to capture finer variations and have smooth image regions that must be addressed to achieve better retrieval performance [24].

All the above-mentioned methods employed the traditional similarity matching-based approach (Euclidian distance, Manhattan distance, L2, and Canberra distance) to perform the retrieval task. Existing methods also employed

the learning-based approaches for CBIR. Gajanan *et al.* [28] presented a directional magnitude local triplet pattern to effectively extract both the orientation and structural microscopic details. This feature descriptor improves the retrieval performance using similarity distance measure as well as the Artificial neural networks learning approach. In [29], a feature descriptor consisting of SIFT, LBP, LDP, LTP, and HOG was proposed to better extract the patch-level information from the images. The performance of the patch-based descriptor was also evaluated with classifiers, which include Support vector machine (SVM) and Random forest. Chandani *et al.* [30] integrated the SIFT and Gabor descriptors. The method applied both the learning as well as non-learning-based approaches. Experimental results indicate better performance when these features were used with the SVM. Dakshata *et al.* [31] employed the region-based descriptor based on Zernike-'s and Hu's seven-moments with the SVM classifier for CBIR. According to the experimental results, Hu's seven moments seem to do best in almost all classes of the CT database in terms of accuracy. In [32], EH features were used in combination with color auto-correlogram, color moment, and Gabor wavelet transforms to train the SVM for classification.

Deep learning-based convolutional neural networks (CNNs) have recently succeeded impressively in computer vision and image processing applications. Different types of deep neural networks have been suggested for image retrieval and classification tasks, such as compact root bilinear- convolutional neural network (CRB-CNN) [36], deep belief network (DBN) [37] method, ensembles CNN [38], Alex Net CNN [39], 1D grey-level co-occurrence matrix (GLCM)-CNN [41], deep belief network (DBN) and stacked autoencoder (SAID). Ahmad et. al [36] presented a bilinear CNN-based architecture that utilizes a pre-trained CNN network (VGG-m or VGG-16) to maintain high discriminative image representations at a compact scale. This significantly improves the retrieval efficiency in terms of extraction and search time. Furthermore, various distance measures (Euclidean, Manhattan, and City block) indicate the highest retrieval accuracy. In [37], the DBN method was introduced for learning effective image representations. Hamreras *et al.* [38] introduced ensembles of CNNs for the CBIR task. In comparison to individual CNNs, it was demonstrated in [38] that using CNN ensembles is very effective in producing a strong image representation. Shah *et al.* [39] suggested a CNN-based deep learning algorithm for feature extraction to improve the CBIR retrieval performance. Benco *et al.* [41] presented 1D GLCM with 25 layers of CNN. In [42], the image retrieval problem was examined using two deep learning approaches: DBN & SAID.

The common issues associated with the above approaches further inspired the researchers to progressively investigate CBIR and design more robust feature descriptors to improve the retrieval accuracy. In CBIR, texture-based descriptors have been actively explored over the years to effectively capture the discriminative information from the images. The proposed DMLHP texture descriptor is an effort in this direction to capture maximum discriminative information from the images via orientation and magnitude patterns.

III. PROPOSED METHODOLOGY

This section provides a detailed discussion of the proposed CBIR method that presents a novel directional magnitude local hexadecimal pattern descriptor. The flow of the proposed method is illustrated in Fig. 3. The proposed method initially applies image resizing as a prepossessing step to prepare the images for further processing. Afterward, we employ the novel DMLHP descriptor to extract the features of each image in the training and testing sets. In the next step, histograms of the texture orientation and magnitude patterns are computed and fused together to generate the feature vector. Next, we train the classifier using this feature vector. After training the classifier, we apply the same procedure to the selected query image. Finally, we compute the similarity between the feature values of the query image and the images store in the repository.

A. DIRECTIONAL MAGNITUDE LOCAL HEXADECIMAL PATTERN FEATURE DESCRIPTOR

Over the last few years, we have observed various local patterns such as LBP, MLBP, LDP, and ELTP, designed for CBIR. Existing local texture patterns have certain limitations such as sensitivity to noise and lighting conditions, failure to capture finer variations, multiple objects, and complex background problems. All these limitations motivated us to propose a novel directional magnitude local hexadecimal pattern (DMLHP) descriptor. The framework of the proposed feature descriptor, based on the texture orientation and texture magnitude, is shown in Fig. 4. The TOBP effectively extracts more discriminative information from the images, as our pattern computes 16 aspects of directional information based on the horizontal, diagonal, and vertical derivatives. Furthermore, we compute the TMBP using horizontal, diagonal, and vertical derivatives that capture more detailed edge information from the images. Moreover, we employed the proposed descriptor with learning methods to enhance the classification performance of CBIR.

For a given image $I(x, y)$, we compute the 1^{st} order derivatives at the grayscale value of the surrounding pixels along 0°, 45°, 90°, and 135° directions defined as $I^1_\alpha(g_s)|_{\alpha=0°,45°,90°,135°}$. The 1^{st} order derivative at the grayscale value of the center pixel along 0°, 45°, 90°, and 135° directions are computed as:

$$I^1_{0°}(g_c) = I(g_h) - I(g_c) \tag{1}$$
$$I^1_{45°}(g_c) = I(g_d) - I(g_c) \tag{2}$$
$$I^1_{90°}(g_c) = I(g_v) - I(g_c) \tag{3}$$
$$I^1_{135°}(g_c) = I(g_{db}) - I(g_c) \tag{4}$$

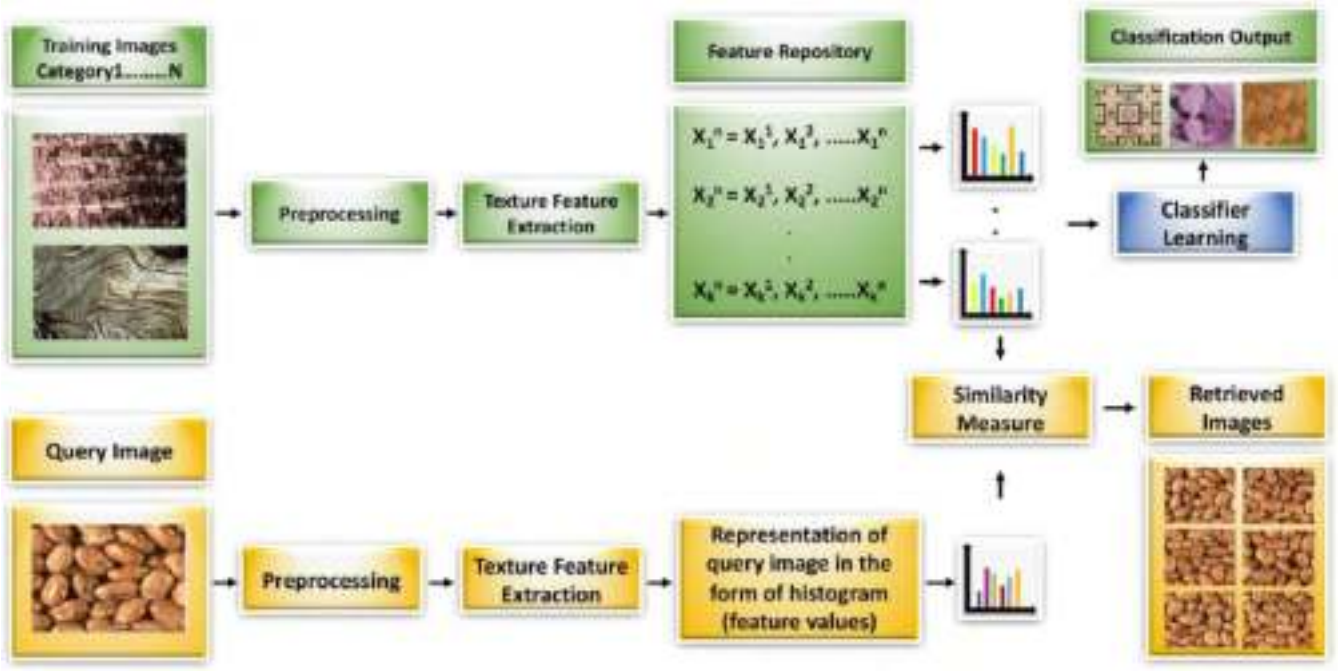


FIGURE 3. Block diagram of the proposed work with similarity-based and learning-based approaches.

where g_h, g_d, g_v, and g_{db} denote the horizontal, diagonal, vertical, and diagonal-back neighborhoods direction of the center pixel, respectively. Based on the 1^{st} order derivative values of center pixel, the direction of the center pixel is computed using (5), as shown at the bottom of the page.

These are 16 possible directional values of the center pixel, and the image is transformed into 16 distinct values i.e., directions. Let $I^1_{dir.}(g_s)|_{s=1...,8}$ denote the direction of 3×3 neighborhoods such as eight surrounding neighbors, for which directions can be calculated in the same manner.

$$I^1_{dir.}(g_c) = \begin{cases} 1, & I^1_{0^\circ}(g_c) \geq 0 \text{ and } I^1_{45^\circ}(g_c) \geq 0 \text{ and } I^1_{90^\circ}(g_c) \geq 0 \text{ and } I^1_{135^\circ}(g_c) \geq 0 \\ 2, & I^1_{0^\circ}(g_c) \geq 0 \text{ and } I^1_{45^\circ}(g_c) \geq 0 \text{ and } I^1_{90^\circ}(g_c) \geq 0 \text{ and } I^1_{135^\circ}(g_c) < 0 \\ 3, & I^1_{0^\circ}(g_c) \geq 0 \text{ and } I^1_{45^\circ}(g_c) \geq 0 \text{ and } I^1_{90^\circ}(g_c) < 0 \text{ and } I^1_{135^\circ}(g_c) \geq 0 \\ 4, & I^1_{0^\circ}(g_c) \geq 0 \text{ and } I^1_{45^\circ}(g_c) \geq 0 \text{ and } I^1_{90^\circ}(g_c) < 0 \text{ and } I^1_{135^\circ}(g_c) < 0 \\ 5, & I^1_{0^\circ}(g_c) \geq 0 \text{ and } I^1_{45^\circ}(g_c) < 0 \text{ and } I^1_{90^\circ}(g_c) \geq 0 \text{ and } I^1_{135^\circ}(g_c) \geq 0 \\ 6, & I^1_{0^\circ}(g_c) \geq 0 \text{ and } I^1_{45^\circ}(g_c) < 0 \text{ and } I^1_{90^\circ}(g_c) \geq 0 \text{ and } I^1_{135^\circ}(g_c) < 0 \\ 7, & I^1_{0^\circ}(g_c) \geq 0 \text{ and } I^1_{45^\circ}(g_c) < 0 \text{ and } I^1_{90^\circ}(g_c) < 0 \text{ and } I^1_{135^\circ}(g_c) \geq 0 \\ 8, & I^1_{0^\circ}(g_c) \geq 0 \text{ and } I^1_{45^\circ}(g_c) < 0 \text{ and } I^1_{90^\circ}(g_c) < 0 \text{ and } I^1_{135^\circ}(g_c) < 0 \\ 9, & I^1_{0^\circ}(g_c) < 0 \text{ and } I^1_{45^\circ}(g_c) \geq 0 \text{ and } I^1_{90^\circ}(g_c) \geq 0 \text{ and } I^1_{135^\circ}(g_c) \geq 0 \\ 10, & I^1_{0^\circ}(g_c) < 0 \text{ and } I^1_{45^\circ}(g_c) \geq 0 \text{ and } I^1_{90^\circ}(g_c) \geq 0 \text{ and } I^1_{135^\circ}(g_c) < 0 \\ 11, & I^1_{0^\circ}(g_c) < 0 \text{ and } I^1_{45^\circ}(g_c) \geq 0 \text{ and } I^1_{90^\circ}(g_c) < 0 \text{ and } I^1_{135^\circ}(g_c) \geq 0 \\ 12, & I^1_{0^\circ}(g_c) < 0 \text{ and } I^1_{45^\circ}(g_c) \geq 0 \text{ and } I^1_{90^\circ}(g_c) < 0 \text{ and } I^1_{135^\circ}(g_c) < 0 \\ 13, & I^1_{0^\circ}(g_c) < 0 \text{ and } I^1_{45^\circ}(g_c) < 0 \text{ and } I^1_{90^\circ}(g_c) \geq 0 \text{ and } I^1_{135^\circ}(g_c) \geq 0 \\ 14, & I^1_{0^\circ}(g_c) < 0 \text{ and } I^1_{45^\circ}(g_c) < 0 \text{ and } I^1_{90^\circ}(g_c) \geq 0 \text{ and } I^1_{135^\circ}(g_c) < 0 \\ 15, & I^1_{0^\circ}(g_c) < 0 \text{ and } I^1_{45^\circ}(g_c) < 0 \text{ and } I^1_{90^\circ}(g_c) < 0 \text{ and } I^1_{135^\circ}(g_c) \geq 0 \\ 16, & I^1_{0^\circ}(g_c) < 0 \text{ and } I^1_{45^\circ}(g_c) < 0 \text{ and } I^1_{90^\circ}(g_c) < 0 \text{ and } I^1_{135^\circ}(g_c) < 0 \end{cases} \tag{5}$$

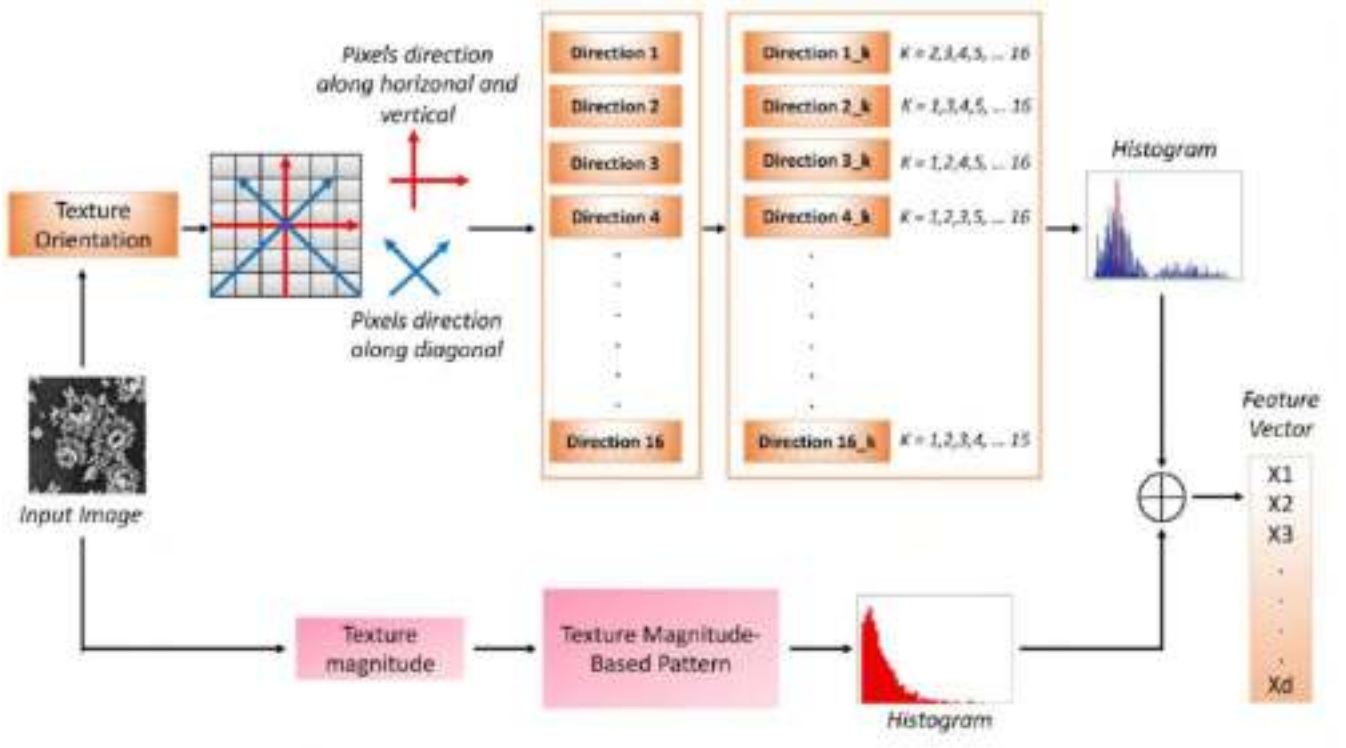


FIGURE 4. Framework of the proposed feature descriptor based on the texture orientation and texture magnitude.

The 2^{nd} order derivative of the center pixel is defined as:

$$\begin{aligned} TOBP^2 = \{&T_1(I^1_{dir.}(g_c), I^1_{dir.}(g_1)),\\ &T_1(I^1_{dir.}(g_c), I^1_{dir.}(g_2)),\\ &\ldots, T_1(I^1_{dir.}(g_c), I^1_{dir.}(g_s))\}|_{s=8} \end{aligned} \tag{6}$$

$$\begin{aligned} &(I^1_{dir.}(g_c), I^1_{dir.}(g_s))\\ &= \begin{cases} 0, & (I^1_{dir.}(g_c) = I^1_{dir.}(g_s))\\ I^1_{dir.}(g_s), & \text{otherwise} \end{cases} \end{aligned} \tag{7}$$

Finally, we obtained an 8-bit texture orientation-based pattern for each pixel by comparing the $(n-1)^{th}$ order derivatives of center pixel direction with all the eight surrounding neighbors' direction, using (7) and (7). Then the orientation pattern is separated into 15 binary patterns. Similarly, for the n^{th} order texture orientation-based pattern, the $(n-1)^{th}$ order derivatives in horizontal, diagonal, vertical, and diagonal-back directions, denoted as $I^{n-1}_{\alpha}(g_s)|_{\alpha=0°,45°,90°,135°}$, is computed as:

$$\begin{aligned} TOBP^n = \{&T_1(I^{n-1}_{dir.}(g_c), I^{n-1}_{dir.}(g_1)),\\ &T_1(I^{n-1}_{dir.}(g_c), I^{n-1}_{dir.}(g_2)),\\ &\ldots, T_1(I^{n-1}_{dir.}(g_c), I^{n-1}_{dir.}(g_s))\}|_{s=8} \end{aligned} \tag{8}$$

An example of the TOBP calculation procedure for a center pixel highlighted with yellow color and surrounding neighbors highlighted with red color are presented in Fig. 5. The direction of the center pixel and each surrounding neighbor are calculated using (5). We have discussed in Fig. 5 that if the direction of the center pixel is the same as that of surrounding neighbors, assign 0 to the corresponding bit of TOBP according to (7), and if the direction of the center pixel is different from that of the surrounding neighbors, retain the corresponding bit of TOBP with the direction of the surrounding neighbors using (7). In Fig. 6, for center pixel 5, let the direction of the center pixel $I^1_{dir.}(g_c)$ obtained using (5) be 14. The direction of the first neighboring pixel is 15, which is different than the direction of the center pixel, so the first bit of TOBP is retained with the same neighboring pixel value, which is 15 based on (7). Similarly, for the second neighborhood pixel 3, the calculated direction is 11, which is again different than the center pixel direction. Hence, the second bit of TOBP is coded with 11. For the third neighborhood pixel 6, the direction is 14, which is the same as the direction of the center pixel; thus, the pattern is coded with 0 according to (7). Furthermore, the remaining neighborhood directions are different from that of the center pixel, and thus, the other bits of TOBP are coded with 2 3 8 16 and 5 respectively. The resultant 8-bit TOBP is 15 11 0 2 3 8 16 5. Afterward, an 8-valued orientation code for each direction is separated into 15 binary patterns based on the direction of the central pixel.

In the illustrated example shown in Fig. 6, the direction of the central pixel $I^1_{dir.}(g_c)$ obtained using (5) is 14, then 2^{nd} order TOBP is divided into 15 binary patterns, and calculated as follows:

$$\begin{aligned} &TOBP^2|_{direction=1,2,3,4,5,6,7,8,9,10,11,12,13,15,16}\\ &\quad = \sum_{s=1}^{s} 2^{(s-1)} \times T_2(TOBP^2(g_c))|_{direction=\alpha} \end{aligned} \tag{9}$$

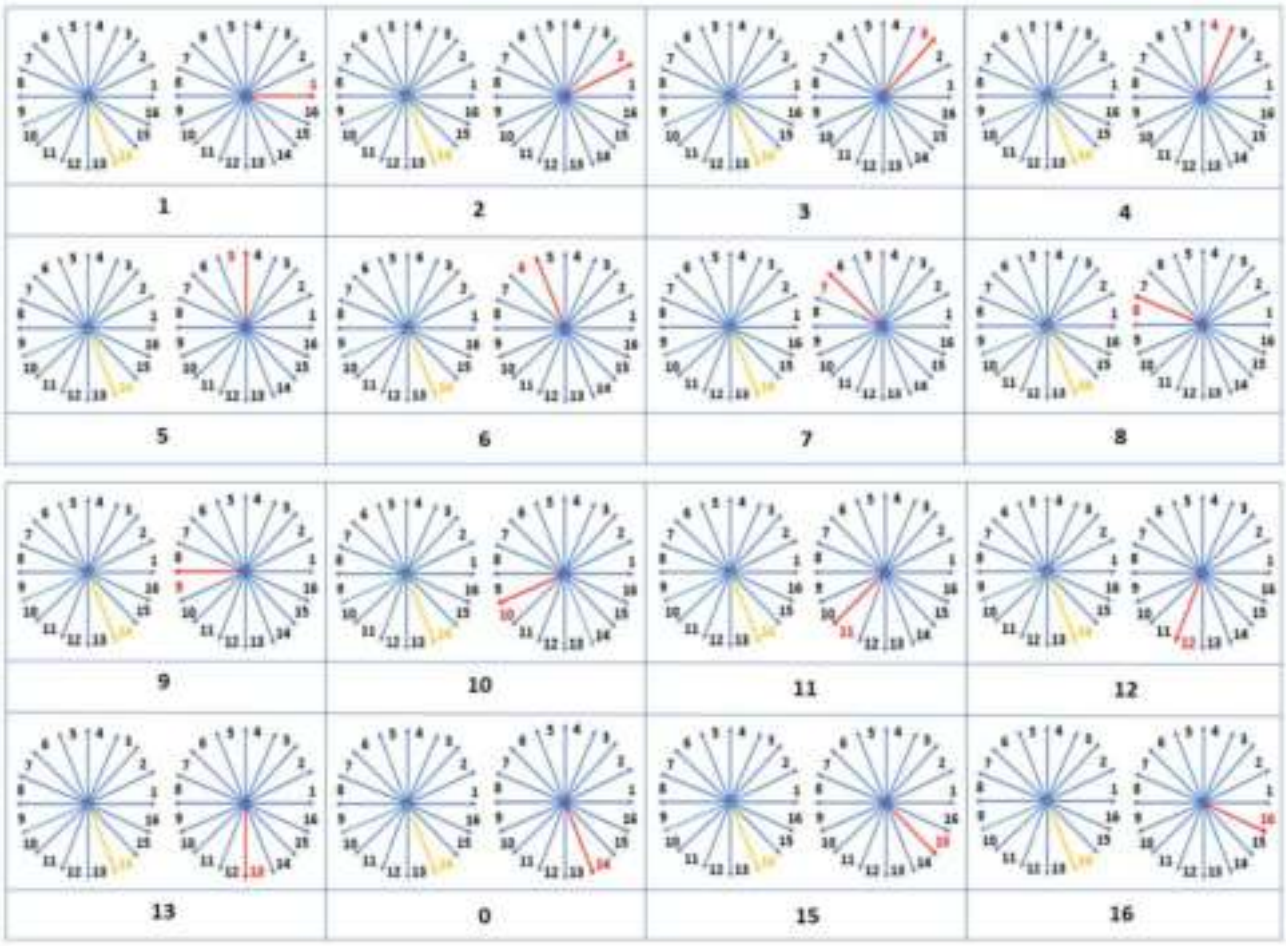


FIGURE 5. Calculation of texture orientation-based pattern values for the center pixel direction '14' based on the direction of surrounding neighbors. The yellow line represents the center pixel direction, while the red line represents the direction of surrounding neighbors.

where function T_2 is defined as:

$$T_2(TOBP^2(g_c))|_{direction=\alpha} = \begin{cases} 1, & TOBP^2(g_c) = \alpha \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

and $\alpha \; \varepsilon \; \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16\}$.

For the direction of center pixel, which is 14 in our case, the formation of binary patterns for the other 15 directions, i.e., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, and 16, can be seen in Fig. 7. As our 8-bit TOBP value is 15 11 0 2 3 8 16 5, the binary patterns are generated based on this TOBP value. The first binary pattern is obtained by retaining 1, where the bit of TOBP value is 1 and 0 is placed for the remaining bits. As TOBP do not contain any value of 1, the binary pattern is 0 0 0 0 0 0 0 0. The second binary pattern 0 0 0 1 0 0 0 0 is obtained by keeping 1 where the TOBP value is 2 and the remaining bits are coded with 0. Similarly, the remaining thirteen binary patterns 0 0 0 0 1 0 0 0, 0 0 0 0 0 0 0 0, 0 0 0 0 0 0 0 1, 0 0 0 0 0 0 0 0, 0 0 0 0 0 0 0 0, 0 0 0 0 0 1 0 0, 0 0 0 0 0 0 0 0, 0 0 0 0 0 0 0 0, 0 1 0 0 0 0 0 0, 0 0 0 0 0 0 0 0, 0 0 0 0 0 0 0 0, 1 0 0 0 0 0 0 0, and 0 0 0 0 0 0 1 0 are formed separately for texture orientation-based pattern values 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, and 16, respectively. In the same way, the other 15 orientation patterns are generated for the remaining directions of the center pixels, that is, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15 and 16. Based on the 8-bit orientation code, each direction is transformed into 15 binary patterns as per (9). In this way, we calculated a total of 240 (16 × 15) binary patterns.

Although for every binary pattern, the sign information is more important, the magnitude information also plays a significant role, which is ignored in the binary pattern. However, the magnitude information effectively captures the edge and gradient structure over other texture descriptors such as LBP. The idea of LBP guided us to introduce a novel magnitude pattern for image retrieval. Since compact texture information lies along the horizontal, diagonal, and vertical directions, the main aim of the proposed pattern is to examine the relationship of neighboring pixels with their adjacent neighbors by utilizing the magnitude information in horizontal, diagonal, vertical, and diagonal-back directions. The 241^{st} TMBP is computed from the magnitude of horizontal,

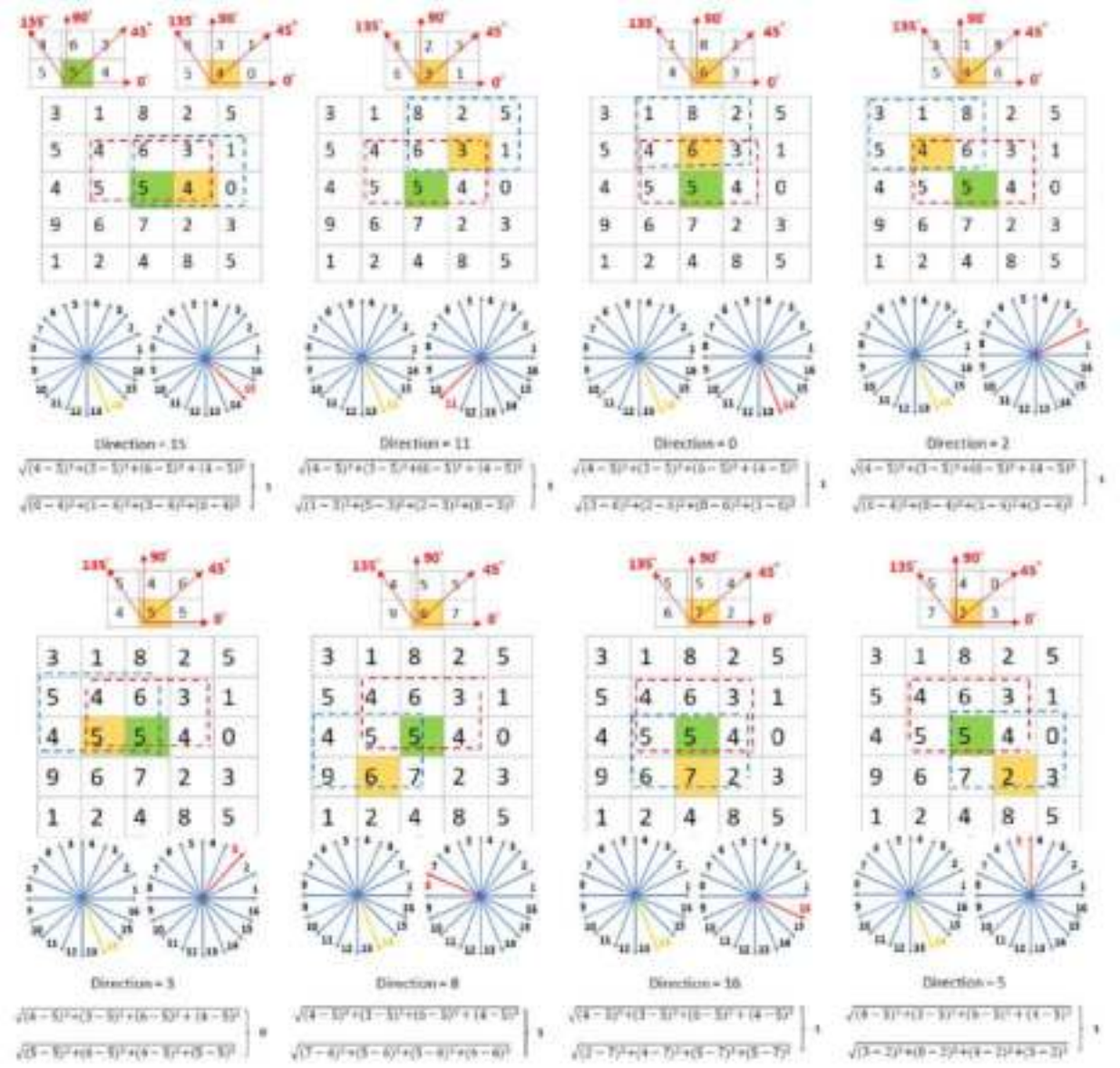


FIGURE 6. Example showing the calculation of texture orientation-based patterns and the texture Magnitude-based Pattern. For the TOBP, the bit is coded with the direction of surrounding neighbors when the center pixel and surrounding neighbors having different directions, or else '0'. The red box represents the center pixel '5' with its adjacent neighbors 4, 3, 6, 4 in 0°, 45°, 90°, and 135° directions, respectively, while the blue box represents the eight surrounding neighbors, i.e., 4, 3, 6, 4, 5, 6, 7, 2, of the center pixel with their adjacent neighbors in 0°, 45°, 90°, and 135° directions. To compute the TMBP, a value of 1 is assigned when the magnitude of the surrounding neighborhood is greater than the center pixel; otherwise, 0 is assigned.

diagonal, vertical, and diagonal-back 1^{st} order derivatives using the (12).

$$M_I^1(g_S) = \sqrt{(I^1_{0^\circ}(g_s))^2 + (I^1_{45^\circ}(g_s))^2 + (I^1_{90^\circ}(g_s))^2 + (I^1_{135^\circ}(g_s))^2} \tag{11}$$

$$TMBP = \sum_{s=1}^{s} 2^{(s-1)} \times T_3(M_I^1(g_S) - M_I^1(g_c))|_{s=8} \tag{12}$$

$$T_3(x) = \begin{cases} 1, & x >= 0 \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

Here T_3 is a function, where x is calculated based on the difference between the magnitude of surrounding neighbors and the magnitude of a center pixel as can be seen in (12), that is, $x = M_I^1(g_S) - M_I^1(g_c)$. The magnitude of the center pixel $M_I^1(g_c)$ is 2.6, which is calculated by using (12). Similarly, the magnitudes of the surrounding neighbors $M_I^1(g_S)$ are calculated. For the 1^{st} neighborhood pixel 4, the magnitude is 5.4, which is greater than the magnitude of the center pixel. Hence, the first bit of TMBP is coded with 1. Moreover, the magnitude of the second neighbor is 5.8, which is again greater than the magnitude of the center pixel. Hence, we assign the value of 1 to the corresponding bit of TMBP.

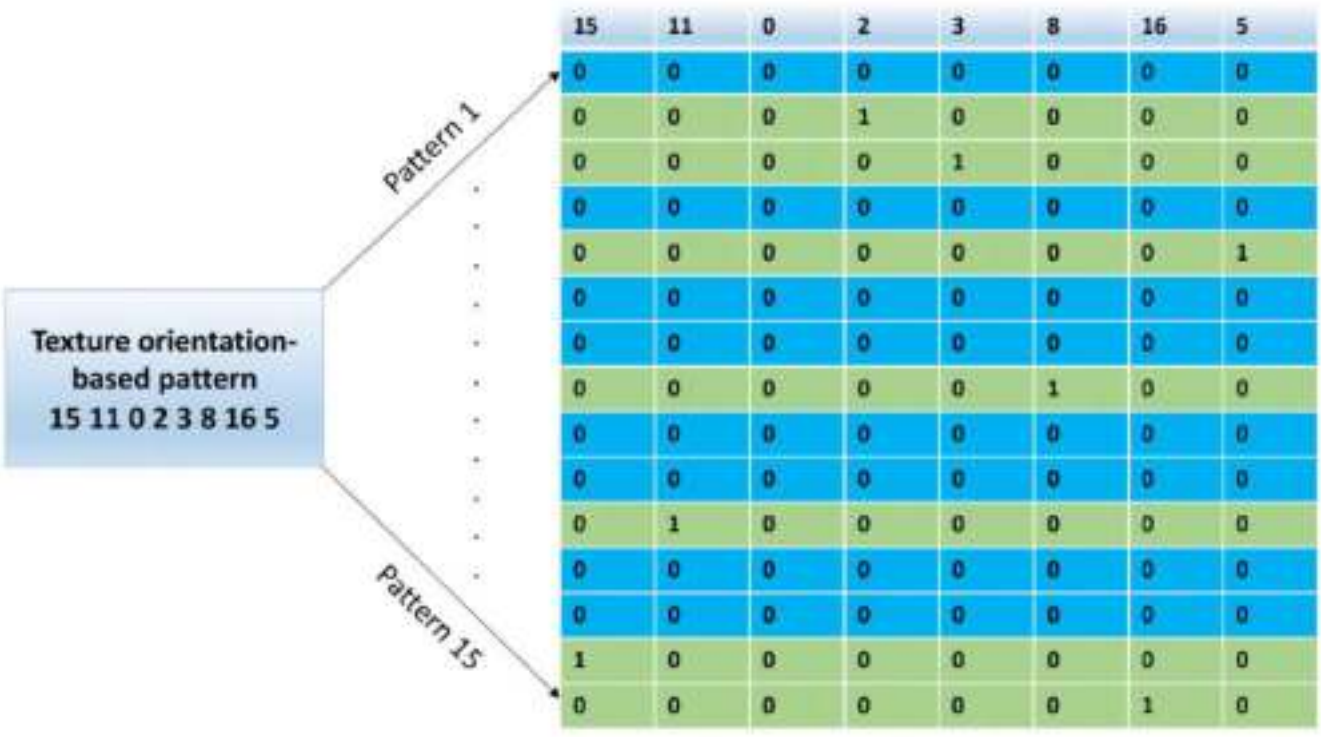


FIGURE 7. Binary pattern generation with blue rows represent the non-occurring pattern values (1, 4, 6, 7, 9, 10, 12, 13) while green rows represent the occurring pattern values (2, 3, 5, 8, 11, 15, 16).

For the third and fourth neighbors, TMBP is coded with 1. For the fifth neighbor 5, the magnitude is 1.4, which is less than the magnitude of the center pixel. So, in this case, the magnitude pattern is coded with 0. Based on the neighboring magnitude, other bits are coded with 1 1 1. The resultant TMBP is 1 1 1 1 0 1 1 1.

After extracting the local image pattern LIP (TOBP and 241^{st} TMBP) of each pixel (j, k), we obtain the histogram of TOBP and TMBP, using (14). The final feature descriptor is formed by concatenating the two histograms of our TOBP and TMBP patterns, which are Hist TOBP and Hist TMBP as shown in (16).

$$H_s(l) = \frac{1}{M \times N} \sum_{j=1}^{M} \sum_{k=1}^{N} T_4(LIP(j,k), l) \tag{14}$$

where $l\varepsilon\,[0, s(s-1)+2]$ is the maximum LIP pattern value, $M \times N$ represents the size of input image, $x = LIP(j,k)$, $y = l$, and function T_4 is defined as:

$$T_4(x,y) = \begin{cases} 1, & x = y \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

$$Hist^{DMLHP} = \left[Hist^{TOBP} \parallel Hist^{TMBP} \right] \tag{16}$$

The algorithm of the proposed work is described below:

B. IMAGE CLASSIFICATION

Ensemble classifiers have demonstrated their effectiveness for various classification tasks in computer vision. We employed the proposed descriptor to train the Ensemble subspace discriminant (ESD) for classification. Later, we performed the testing on the unseen images of the repository. Next, we used the proposed descriptor to extract the features of the query image selected by the user, and the trained classifier determined the category of the query image. We employed the ESD classifier for the proposed method, as we obtained the best results on ESD when used with our novel features. Another benefit of subspace discriminant-based ensemble methods is their ability to handle large image repositories because they are fast, accurate and easy to interpret. In our case, ESD took a reasonable computational time as it has a prediction speed of 40 observations per second.

The range of evaluation parameters for the Ensemble subspace discriminant classifier in our work is as follows: the number of learners – range [10,500], learning rate – range [0.001,1], and the subspace dimension – range [1, 319]. In our experiment, we explored different hyper-parameter values to enhance the classification performance of different classifiers, namely ESD, SVM, KNN, and XGBoost. We trained these classifiers on different hyper-parameters (mentioned below) and reported the results on those parameters, where we achieved the best results. We tuned the following parameters for ESD and set the learning rate = 0.1, the number of learners = 30, and the subspace dimension = 160, as we achieved optimal results in these settings. For SVM, we set the kernel function = Gaussian and kernel scale = 18 with one vs all multiclass method, whereas for KNN, we set the

Algorithm 1 Proposed Algorithm

Input: Image Repository $C = I_1, I_2, \ldots, I_n$, Query Image X
Output: Classification Output, Retrieved Images $S = S_1, S_2, S_3, \ldots, S_n$

prepossessing
for *TOBP* **do**
  $I^1_{\alpha}(g_s)|_{\alpha=0°,45°,90°,135°}$; // First-order derivative
  $I^1_{dir.}(g_s)|_{s=1,\ldots,8}$; // Pixels orientation
  if *Direction 1:* **then**
    $[TOBP\_1]$; // 8-bit texture orientation-based pattern
    $[TOBP\_1\_k], k = \{2, 3, 4, ..16\}$; // Separate into 15 binary patterns
  end
  else if *Direction 2:* **then**
    $[TOBP\_2]$; // 8-bit texture orientation-based pattern
    $[TOBP\_2\_k], k = \{1, 3, 4, ..16\}$; // Separate into 15 binary patterns
  end
  else if *Direction 16:* **then**
    $[TOBP\_16]$; // 8-bit texture orientation-based pattern
    $[TOBP\_16\_k], k = \{1, 2, 3, ..15\}$; // Separate into 15 binary patterns
  end
  $Hist^{TOBP}$; // Histograms of orientation patterns
end
for *TMBP* **do**
  $M^1_I(g_s)|_{s=1,\ldots,8}$; // Pixels magnitude
  $[TMBP]$; // 8-bit texture magnitude-based pattern
  $Hist^{TMBP}$; // Histograms of magnitude patterns
end
$Hist^{DMLHP} = \left[Hist^{TOBP} \parallel Hist^{TMBP}\right]$; // Histogram concatenation
$X_i \, \varepsilon \, R^n$; // Construct feature space
$D(Q, IR)$; // Query image comparison
$\{S_1, S_2, \ldots., S_n\}$; // Retrieval output
Image classification methods
Classification output

method = weighted KNN, number of neighbors = 1, and distance metric = Euclidean. With regard to XGBoost, we set the maximum depth = 30, maximum round = 500, and evaluation metric = mlogloss.

C. IMAGE MATCHING

In case of the CBIR approach, our goal is to retrieve the top k most similar images against a query image based on the distance. We employed the weighted Manhattan distance (Weighted L1 norm) for this purpose due to its ability to yield robust results. Moreover, the weighted Manhattan distance outperforms other similarity measures, such as Euclidean, Minkowski, and chi-square, on higher-dimensional data [40]. Thus, the image matching function can be expressed as

$$DM(p, q) = min \sum_{i=1}^{n} \left| \frac{p_i - q_i}{1 + p_i + q_i} \right| \tag{17}$$

Here p_i and q_i represent the feature vectors of the images present in the repository and target image, respectively, whereas $DM(p, q)$ is the image retrieval function of CBIR that retrieves the most visually similar images based on most to least similarity.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section provides a discussion on different experiments conducted to evaluate the performance of our method. We evaluated our method on three standard image repositories, that is, AT&T [33], Brodatz Texture (BT) [34], and MIT Vistex [35]. All of these comparative methods such as LNIP [43], LTriDP [44], LNDP [45], LDGP [46], LEPSEG [47], and CSLBP [48] have also used AT&T, Brodatz Texture, and MIT Vistex open datasets with the same experimental setup used for our proposed approach. The details of these image repositories and evaluation metrics are also provided in this section.

A. PERFORMANCE EVALUATION PARAMETERS

The performance of the proposed descriptor is evaluated using the Average retrieval precision (ARP), Average retrieval recall (ARR), Average retrieval specificity (ARS), and Average retrieval accuracy (ARA). Precision is defined as the ratio of the total number of relevant images retrieved to the total number of images retrieved. We computed the precision as follows:

$$P_i = \frac{I_R}{I_T} \tag{18}$$

where I_R and I_T represent the number of relevant images retrieved and the total number of retrieved images, respectively in response to the inquiry image represented by i.

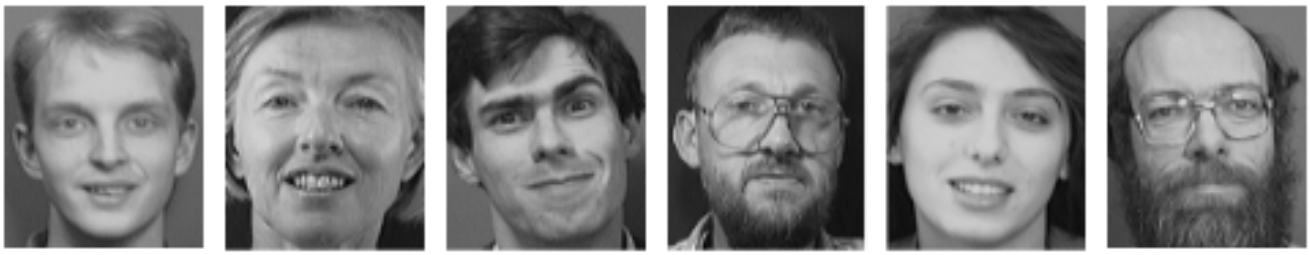
The recall is another evaluation parameter that denotes the ratio of the total number of relevant images retrieved to the total number of relevant images in the repository.

$$R_i = \frac{I_R}{I_C} \tag{19}$$

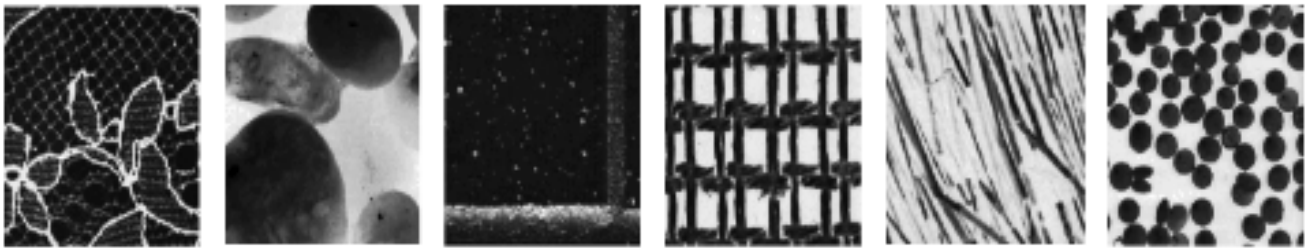
Here, I_R is the number of relevant images retrieved, and I_C is the total number of images in each category of the repository in response to the inquiry image

TABLE 1. Details of image repositories.

Image Repository	Total Images	Total Categories	% of the training images	% of the testing images
AT&T	400	40	80	20
Brodatz Texture	2800	112	80	20
MIT Vistex	480	30	70	30



(a) AT&T



(b) Brodatz Texture

(c) MIT Vistex

FIGURE 8. Image samples from different image repositories.

represented by i.

$$C_k = \frac{1}{I_C} \sum_{i=1}^{I_C} P_i \tag{20}$$

$$ARP = \frac{1}{T_C} \sum_{k=1}^{T_C} C_k \tag{21}$$

Here, C_k represents the precision of the k^{th} category of the image repository, and T_C represents the total number of categories present in the image repository. Similarly, ARR is calculated by using (23).

$$G_k = \frac{1}{I_C} \sum_{i=1}^{I_C} R_i \tag{22}$$

$$ARR = \frac{1}{T_C} \sum_{k=1}^{T_C} G_k \tag{23}$$

In (22) and (23), G_k represents the recall of k^{th} category of the image repository.

TABLE 2. Average retrieval rate with recent CBIR techniques using NT values of 10, 25, 16 on AT&T, Brodatz Texture, and MIT Vistex Image repositories, respectively.

Image Repository	Average Retrieval Rate (ARR %)						
	LNIP	LTriDP	LNDP	LDGP	LEPSEG	CSLBP	Proposed
AT&T	57	54	53	45	36	43	66
Brodatz Texture	79	76	75	64	63	53	83
MIT Vistex	90	86	85	80	78	73	92

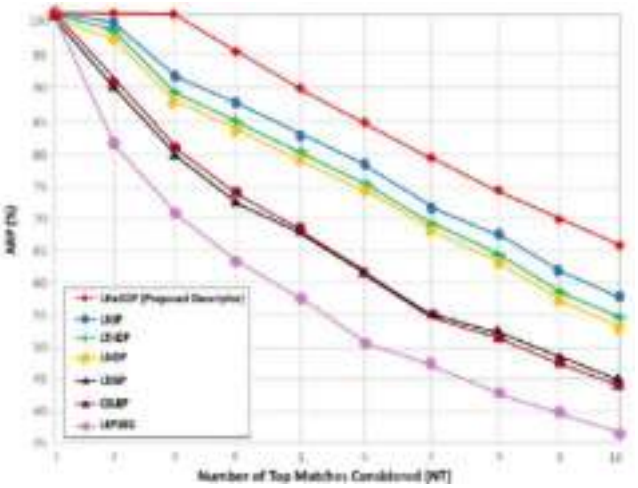


FIGURE 9. Average retrieval precision of DMLHP on different values of NT with state-of-the-art methods on the AT&T image repository.

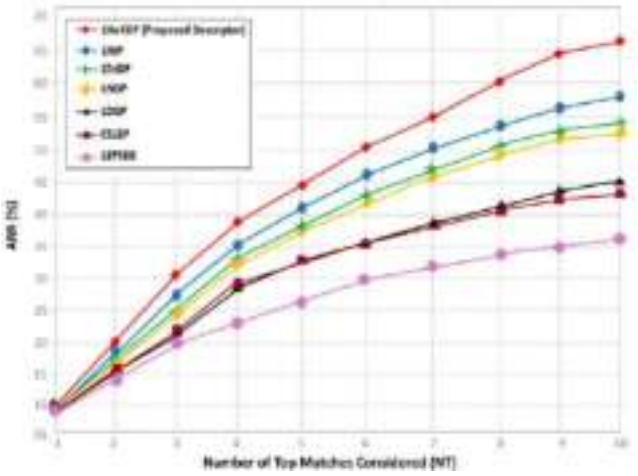


FIGURE 10. Average retrieval recall of DMLHP on different values of NT with state-of-the-art methods on the AT&T image repository.

Specificity is the ratio of a total number of correctly labeled negative images to the total number of negative images, and the specificity is computed by the following equations:

$$S_K = \frac{I_{TN}}{I_{TN} + I_{FP}} \tag{24}$$

$$ARS = \frac{1}{T_C} \sum_{k=1}^{T_C} S_K \tag{25}$$

Here, S_k represents the specificity of the k^{th} category of the image repository, T_c is the total number of categories present in the image repository, I_{TN} is the number of non-matched images that are correctly identified, and I_{FP} is the number of non-matched images that are not correctly identified.

Accuracy is another evaluation parameter that is calculated as

$$A_K = \frac{I_{TP} + I_{TN}}{I_{TP} + I_{TN} + I_{FP} + I_{FN}} \tag{26}$$

$$ARA = \frac{1}{T_C} \sum_{k=1}^{T_C} A_K \tag{27}$$

Here, A_K represents the accuracy of the k^{th} category of the image repository, I_{TP} is the number of matched images that are correctly identified, and I_{FN} is the number of matched images that are not correctly identified.

B. DESCRIPTION OF THE IMAGE REPOSITORIES

To evaluate the performance of any CBIR method, it must be tested on a diverse and challenging image repository. For this purpose, we selected three standard image repositories having a wide range of image themes. Moreover, these repositories are diverse in terms of pose variations, noise, occlusions, and a variety of natural and artificial regular textures. Most commonly-used natural scene and rich textural image repositories such as MIT Vistex, BT, and AT&T face image repositories are used for image retrieval tasks. Every experiment is repeated multiple times and average retrieval precision, recall, specificity, and accuracy values are reported. We split each image repository into an 80-20 training-testing ratio for experimentation. The division of image repositories is shown in Table 1. Some samples of each category of image repository are shown in Fig. 8.

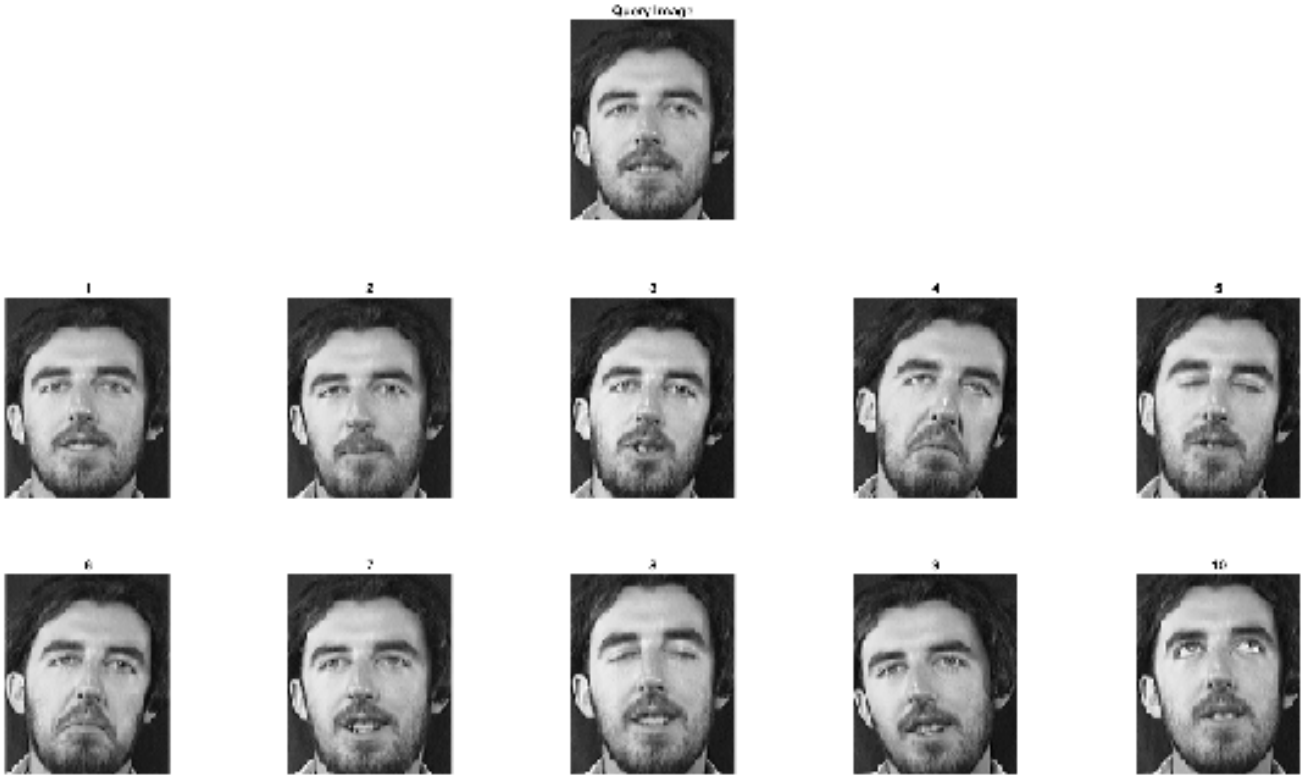


FIGURE 11. Sample image is selected, and the top 10 relevant images are retrieved from the AT&T image repository.

C. PERFORMANCE ANALYSIS ON THE AT&T IMAGE REPOSITORY

The AT&T image repository consists of 400 images partitioned into 40 different categories, and each category consists of 10 images having a resolution of 92×112. Some sample images from 40 categories of the AT&T image repository are depicted in Fig. 8(a). We performed two different experimental analysis, that is, similarity matching-based approach and the learning-based approach.

For the similarity matching-based approach, we randomly select an inquiry image from each category of the image repository to measure the retrieval accuracy of the proposed descriptor. The proposed texture features (TOBP + TMBP) are extracted for a given inquiry image and compared with the feature values of the images stored in the image repository based on the similarity index, i.e., weighted Manhattan distance. For this image repository, the number of top matches (NT) is retrieved in a group of $1, 2, 3, \ldots, 10$ images. The proposed descriptor gives a retrieval accuracy of 66% ARR, using weighted Manhattan distance. For the learning-based approach, the set of 320 random images from the AT&T image repository is selected for training, and the remaining 80 images are used for testing. To measure the robustness of the proposed descriptor, we evaluated the CBIR results, using our features with the ESD classifier. We obtained a high average precision rate of 95%. These results signify the effectiveness of this learning-based approach for CBIR.

1) PERFORMANCE COMPARISON AGAINST DIFFERENT STATE-OF-THE-ART DESCRIPTORS

To show the robustness of the proposed descriptor, a comparative analysis of the proposed descriptor against existing state-of-the-art descriptors is provided in terms of ARP and ARR in Figs. 9 and 10, respectively. From these results on the AT&T image repository, we can conclude that the ARP and ARR are inversely related by varying the value of NT. As the value of NT is increased, the ARR is also increased because of the high true-positive rate while the ARP is decreased due to the high false-positive rate. However, the proposed texture descriptor gives an ARR of 66% on a maximum value of NT, that is, 10, which is outstanding as compared to recent CBIR texture-based descriptors. The ARR indicates that our method yields better image retrieval performance over LNIP by 9%, LTriDP by 12%, LNDP by 13%, LDGP by 21%, LEPSEG by 30%, and CSLBP by 23% as shown in Table 2.

We also tested the robustness of the proposed method for the face recognition task. For this purpose, we designed an experiment to compare the performance of the proposed method against the existing state-of-the-art DLGBD descriptor [49] for facial recognition on the face image repository, AT&T, which is diverse in terms of variations in pose, face angles, gender, and race. Moreover, this image repository also includes the face images of people with and without glasses. For this experiment, we used 80% of the images from the AT&T image repository for training and the remaining

TABLE 3. Comparison of the proposed descriptor with the similarity approach and classification approach.

Image Repository	Proposed descriptor with similarity-based approach				Proposed descriptor with classification-based approach															
					ESD				SVM				KNN				XGBoost			
AT&T	Prec %	Rec %	Sp %	Acc %	Prec %	Rec %	Sp %	Acc %	Prec %	Rec %	Sp %	Acc %	Prec %	Rec %	Sp %	Acc %	Prec %	Rec %	Sp %	Acc %
	66	66	99	98	95	96	99	95	86	87	99	86	85	86	99	85	81	79	99	81
Brodatz Texture	83	83	76	85	92	94	99	92	87	90	99	87	85	87	99	85	89	88	99	89
MIT Vistex	92	92	99	99	99	98	99	98	98	98	99	98	92	94	99	92	85	86	99	85

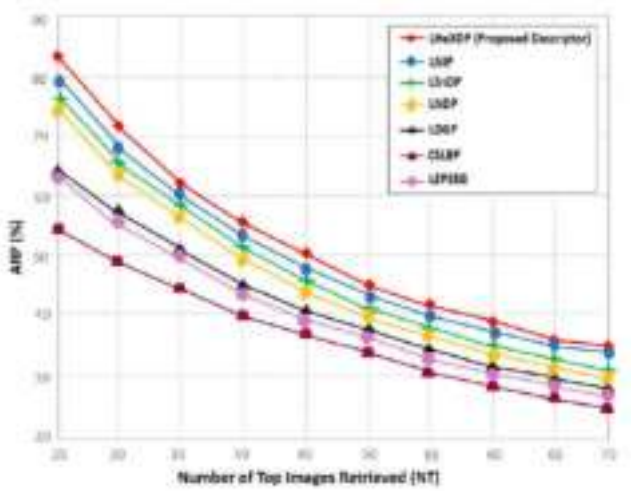


FIGURE 12. Average precision performance assessment based on DMLHP on different values of NT with other state-of-the-art methods on the Brodatz Texture image repository.

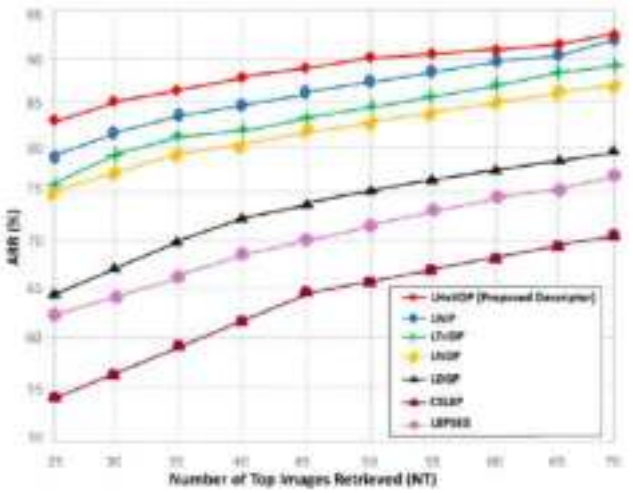


FIGURE 13. Average recall performance assessment based on DMLHP on different values of NT with other state-of-the-art methods on the Brodatz Texture image repository.

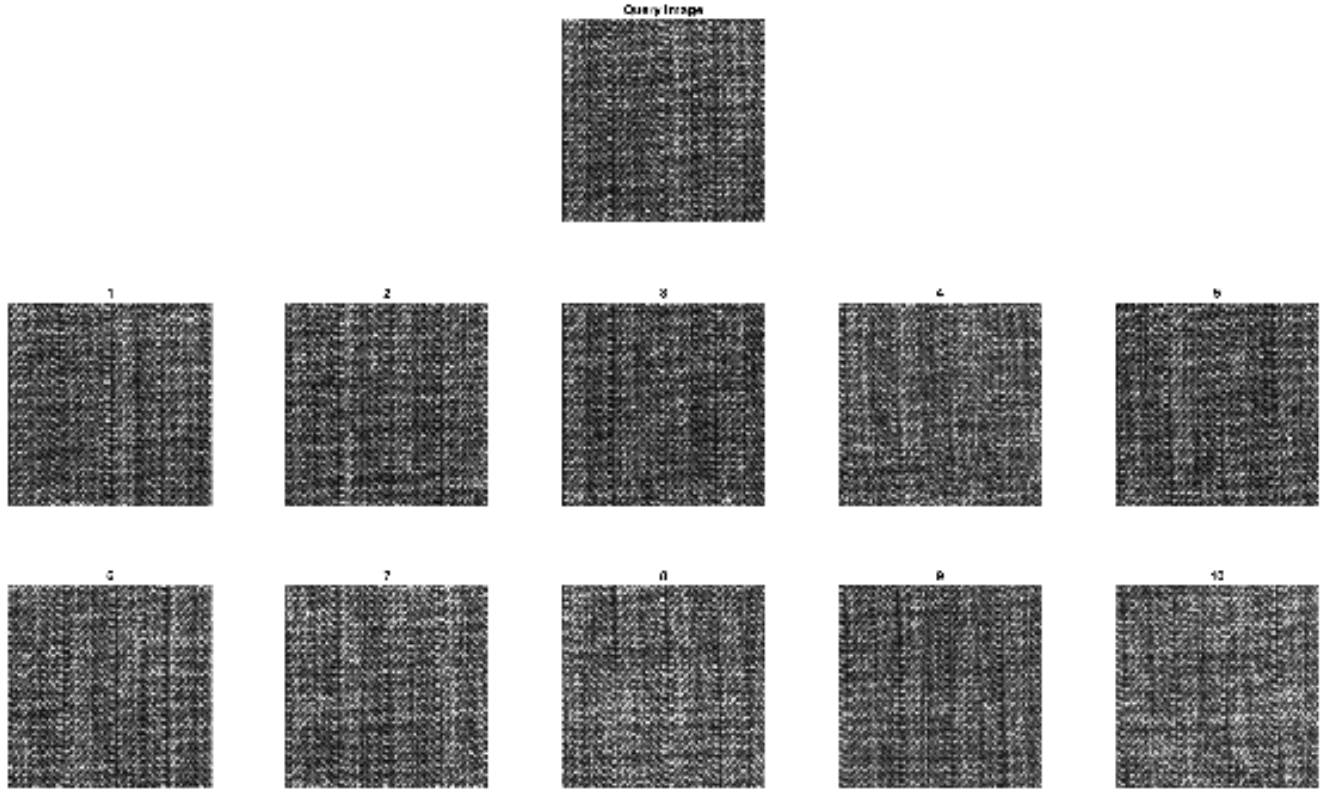
20% of the images for testing. For training purpose, we used 8 images from each category (i.e., $40 \times 8 = 320$ images in total), whereas for testing, we used 2 images from each category (i.e., $40 \times 2 = 80$ images in total). For classification, we employed the ESD classifier and tuned the following parameters: learning rate, number of learners, and subspace dimension. We selected the learning rate of 0.1, number of learners of 30, and the subspace dimension of 160 after extensive experimentation as we obtained the best results on these parameter settings. By using this experimentation protocol, we evaluated the performance of the proposed DMLHP and the DLGBD method. The proposed method achieves the classification accuracy of 95% while the DLGBD method [49] obtains an accuracy of 82.5%. The results of this experiment reveal that the proposed method provides superior detection performance by achieving 12.5% higher accuracy over the DLGBD method. We can conclude from this experiment that the proposed DMLHP descriptor is also capable of effectively representing the facial images to achieve remarkable performance for facial recognition.

2) PERFORMANCE COMPARISON AGAINST DIFFERENT CLASSIFIERS

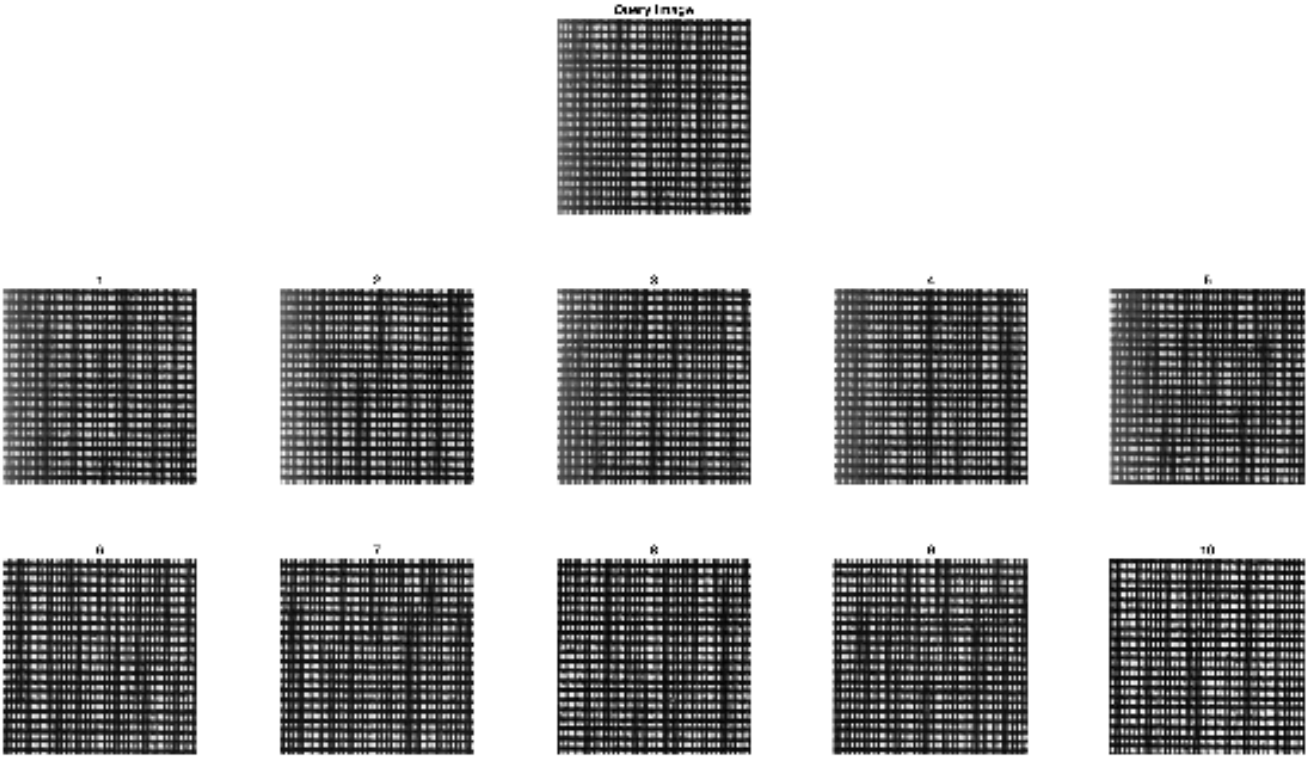
To evaluate the effectiveness of the ESD classifier with the proposed descriptor for CBIR, we compared the performance of ESD against conventional classifiers, including SVM, k-nearest neighbors (KNN), and Extreme gradient boosting (XGBoost). The results of this comparative analysis in terms of average precision, average recall, average specificity, and average accuracy rate are provided in Table 3. From the results shown in Table 3, we can conclude that the ESD classifier outperforms all the classifiers and achieved the highest average precision value of 95%, the highest average recall value of 96%, the highest average specificity value of 99%, and the highest average accuracy value of 95%. SVM performs second best by achieving an average precision, recall, specificity, and accuracy rate of 86%, 87%, 99%, and 86%, respectively, while KNN achieves an average precision, recall, specificity, and accuracy rate of 85%, 86%, 99%, and 85%, respectively. XGBoost performs the worst and achieved 81% average precision, 79% average recall, 99% average specificity, and 81% average accuracy rate. A sample process is shown in Fig. 11, where the query image is taken from the 11^{th} category, and all the retrieved images are relevant to that query image.

D. PERFORMANCE ANALYSIS ON THE BRODATZ TEXTURE IMAGE REPOSITORY

The Brodatz texture image repository is a combination of 112 grayscale textures with a resolution of 640×640. Each category $(D\_1, \ldots\ldots, D\_112)$ is divided into 25 non-overlapping sub-images. Thus, the BT image



(a) Category: D_16



(b) Category: D_21

FIGURE 14. Sample images are selected from ***D*_16** and ***D*_21** category, and the top 10 relevant images are retrieved from the Brodatz Texture image repository.

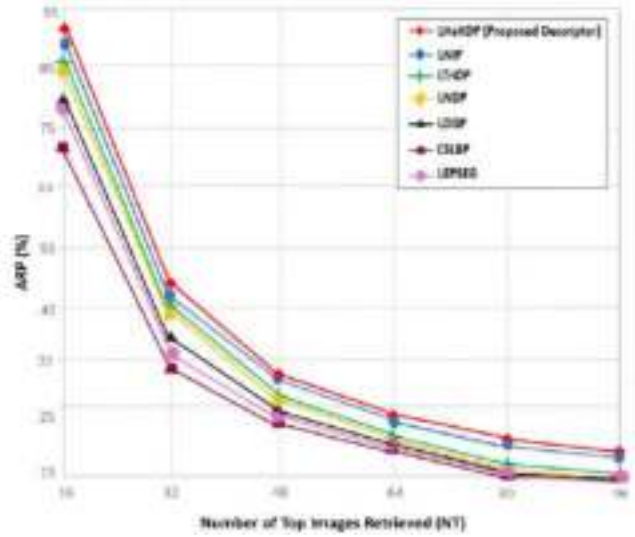


FIGURE 15. Average retrieval precision with a varying number of NT for the MIT Vistex image repository.

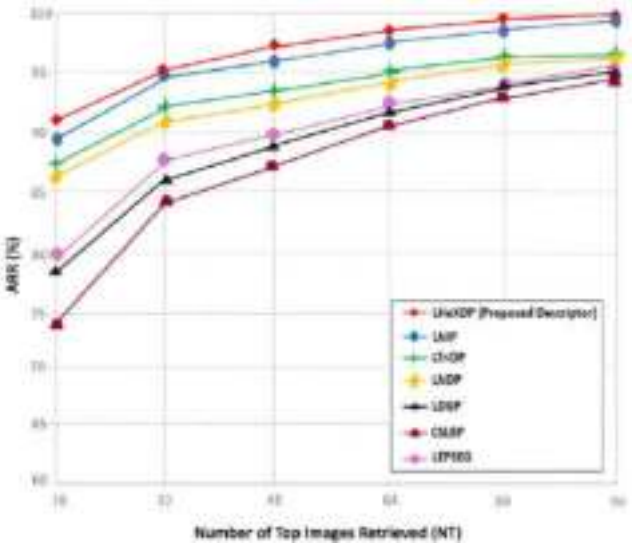


FIGURE 16. Average retrieval recall with a varying number of NT for the MIT Vistex image repository.

repository contains 2800 images in the form of 112 textural categories and each category contains 25 images with a resolution of 128 × 128. Few sample images from the BT image repository are depicted in Fig. 8(b).

For the similarity-based approach, the retrieval performance of our descriptor is evaluated by randomly selecting images from each category. Therefore, 25 images are retrieved initially followed by increasing the retrieved images in a group of 5. Thus, an overall 70 images are retrieved in this process. Hence, the proposed descriptor achieves high retrieval accuracy of 83% using weighted Manhattan distance. As a learning-based approach, a set of 2240 images are used to train the classifier, and a test set of 560 images are used for the evaluation of the proposed descriptor using the ESD classifier. We obtained the average precision rate of 92% that indicates the superior performance of the proposed descriptor for classification.

1) PERFORMANCE COMPARISON AGAINST DIFFERENT STATE-OF-THE-ART DESCRIPTORS

A performance comparison of the proposed descriptor based on texture orientation and magnitude in terms of ARP and ARR with other state-of-the-art methods on BT image repository is shown in Figs. 12 and 13, respectively. When comparing our proposed descriptor with the recent texture-based descriptors, we can see that the retrieval accuracy of the proposed texture descriptor is significantly higher than those of LNIP, LTriDP, LNDP, LDGP, LEPSEG, and CSLBP descriptors by up to 4%, 7%, 8%, 19%, 20%, and 30%, respectively, as shown in Table 2. These results signify the effectiveness of our descriptor over comparative descriptors for CBIR on the Brodatz Texture image repository.

2) PERFORMANCE COMPARISON AGAINST DIFFERENT CLASSIFIERS

Due to the high impact of classifiers on the semantic gap problem, we performed a comparative analysis using different classifiers with the proposed descriptor. Apart from the ESD classifier, we employed the SVM, KNN, and XGBoost-based classifiers for this experiment. The results of the proposed method are compared against the SVM, KNN, and XGBoost classifiers. From the results presented in Table 3, we can observe that the ESD classifier achieved the best results when used with the proposed descriptor. More specifically, we achieved an average precision of 92%, an average recall of 94%, an average specificity of 99%, and an average accuracy of 92%. SVM performed second best and achieved an average precision of 87%, an average recall of 90%, an average specificity of 99%, and an average accuracy of 87%. KNN achieved an average precision of 85%, an average recall of 87%, an average specificity of 99%, and an average accuracy of 85%. The proposed descriptor with the XGBoost classifier performed the lowest by achieving an average precision, recall, specificity, and accuracy of 89%, 88%, 99%, and 89% respectively, as shown in Table 3.

In Figs. 14 and (b), the single image shown in the first row is the query image, while the remaining 10 images are the retrieved images in response to the query image. It can be seen from Fig. 14 that the retrieval results for categories D_16 and also for D_21 are quite effective. The smoothness of the sample images may look similar in terms of the spatial arrangement of colors or intensities; despite this resemblance, the proposed texture descriptor can recognize the texture of images accurately with the same visual appearance.

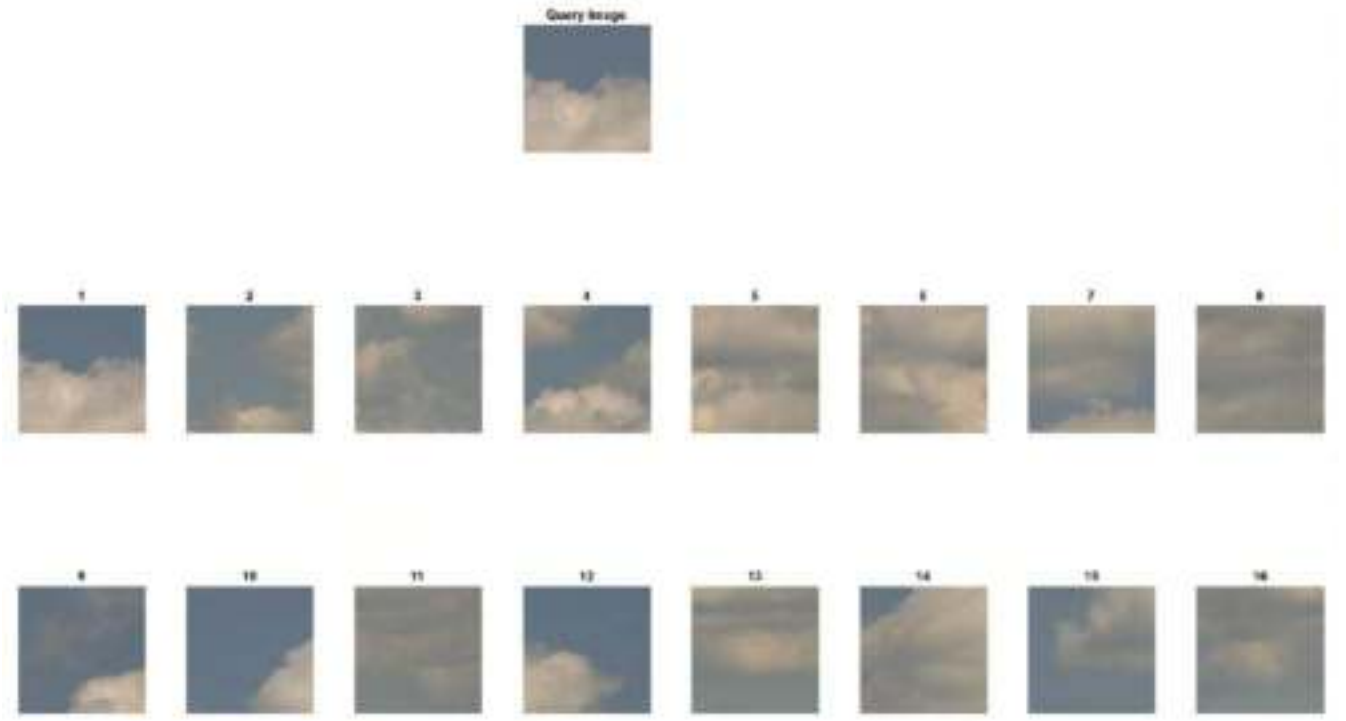


FIGURE 17. Retrieved images show a reduction of the semantic gap in response to the inquiry image taken from the semantic category "Clouds" of the MIT VisTex image repository.

E. PERFORMANCE ANALYSIS ON THE MIT VISTEX IMAGE REPOSITORY

The third image repository, MIT Vistex contains 30 visual texture categories of some natural scenes, including MtValley, ValleyWater, GrassLand, GroundWaterCity, and GrassPlantsSky, with a resolution of 512 × 512. It also contains some random categories, e.g., clouds, food, fabric, and buildings. Each category is divided into 16 sub-images with a resolution of 128 × 128, so there are a total of 30 categories with 16 images in each category. Sample images are presented in Fig. 8(c). For retrieval purposes, the images are retrieved in a group of 16, 32, 48, …, 96. For this experiment, we randomly selected the inquiry images from each category of the MIT Vistex image repository, and we achieved an ARR of 92% on an NT value of 16, using a similarity matching-based approach. For the learning-based approach, we split the image repository into a training set of 360 images and a testing set of 120 images. We observed a significant performance improvement over the similarity matching-based approach. More specifically, we obtained an average precision rate of 99%, which indicates a clear winner between these two approaches.

1) PERFORMANCE COMPARISON AGAINST DIFFERENT STATE-OF-THE-ART DESCRIPTORS

To evaluate the retrieval capabilities of the proposed descriptor, Figs. 15 and 16 present graphical plots of the proposed descriptor with comparative CBIR descriptors in terms of ARP and ARR on different values of NT, i.e., from the top 16 to 96 images. From the results (Table 2), we can easily observe that the proposed texture descriptor outperforms the comparative CBIR descriptors in terms of ARR. Our descriptor achieves better retrieval performance over LNIP by 2%, LTriDP by 6%, LNDP by 7%, LDGP by 12%, LEPSEG by 14%, and CSLBP by 19%.

2) PERFORMANCE COMPARISON AGAINST DIFFERENT CLASSIFIERS

We performed the same comparative analysis experiment on different classifiers with the proposed descriptor for the MIT Vistex repository. The average precision, recall, specificity, and accuracy rate comparison is illustrated in Table 3. The proposed (DMLHP) feature descriptor achieves the best average precision, average recall, average specificity, and average accuracy rates with the ESD, i.e., 99%, 98%, 99%, and 98%, respectively. SVM performs second best and achieves the average precision rate, average recall rate, average specificity rate, and average accuracy rate of 98%, 98%, 99%, and 98%, which is approximately similar to the ESD. Similarly, the average precision, recall, specificity, and accuracy rates for the KNN classifier are 92%, 94%, 99%, and 92%, respectively, while the XGBoost classifier gives an average precision of 85%, an average recall of 86%, an average specificity of 99%, and an average accuracy rate of 85%. Similar to earlier experiments, XGBoost performs the worst for CBIR.

TABLE 4. Classification performance against state-of-the-art deep learning methods in terms of average precision rate, average recall rate, average specificity rate, and average accuracy rate on the MIT Vistex image repository.

Image Repository	Deep learning-based methods								Proposed Method			
MIT Vistex	DBN & SAID [42]				CNN [41]							
	Prec %	Rec %	Sp %	Acc %	Prec %	Rec %	Sp %	Acc %	Prec %	Rec %	Sp %	Acc %
	98.45	98.15	99	98.45	95.28	94.11	99	95.28	99	98	99	98

TABLE 5. Average time complexity analysis of the proposed and comparative descriptors.

Feature descriptors	Feature length	Average features extraction time (second)	Average retrieval time (second)
CSLBP	16	0.1817	0.0623
LEPSEG	512	31.68	2.35
LDGP	64	31.62	2.29
LTriDP	768	31.71	2.38
LNIP	512	31.68	2.35
LTrP (Local Tetra Pattern)	80	31.63	2.30
Proposed	320	31.66	2.33

Moreover, we evaluate the accuracy of each semantic class of the MIT Vistex image repository, and we observe that ESD gets a 100% precision ratio on all the categories that are semantically more enriched because of their overlapped background, color, and texture, while ESD obtains a 60% precision rate on just one category (WheresWaldo). Similarly, ESD achieves a 100% recall ratio on more complex and semantically enriched categories (Clouds, ValleyWater, GraoundWaterCity, Grass, Leaves, etc.), while the recall ratio is 75% and 80% on just two categories, i.e., GrassPlantsSky and MtValley, respectively. Moreover, our method achieves 100% specificity on all categories except on WheresWaldo, where we obtained 98% specificity. Additionally, the accuracy ratio is 100% on almost all the categories except for three: GrassPlantsSky, MtValley, and WheresWaldo. The accuracy ratio on these categories is 99%, 99%, and 98%, respectively. However, the average precision (99%), average recall (98%), average specificity (99%), and average accuracy (98%) rate of the ESD classifier on the MIT Vistex image repository are high compared to those of other selected classifiers such as SVM, KNN, and XGBoost. To show the semantic robustness of the proposed descriptor, the results of the top 16 image retrievals, by taking the target image from the semantic category ''Clouds'' of the MIT Vistex image repository, are shown in Fig. 17. From the visual samples shown in Fig. 17, all the retrieved images belong to the same query class of Clouds, which clearly shows that the proposed descriptor provides accurate results. The most related image class, GroundWaterCity, may be retrieved as the retrieval output because both classes of images (Clouds and GroundWaterCity) are related due to common visual properties such as texture and color, and image semantics such as sky and water. However, both images belong to two different classes. These results demonstrate that our proposed descriptor is capable of retrieving the images of a relevant class even in the presence of images belonging to other semantically similar classes such as Clouds and GroundWaterCity class. Thus, we can argue that the proposed descriptor successfully addresses the issue of the semantic gap in CBIR.

F. PERFORMANCE COMPARISON AGAINST STATE-OF-THE-ART DEEP LEARNING METHODS

The objective of this experiment is to compare the performance of the proposed method against state-of-the-art deep learning methods. For this purpose, we compared the performance of our method (DMLHP - ESD) against the deep learning systems, i.e., DBN & SAID [42], and CNN [41] on the MIT Vistex image repository, and the results are reported in Table 4. From these results, we can observe that the CNN model [41] achieves the lowest accuracy of 95.28%, whereas the proposed method performs best and obtain the highest accuracy of 98%. This comparative analysis illustrates the effectiveness of the proposed method over deep learning models for CBIR.

G. TIME COMPLEXITY ANALYSIS

The objective of this experiment is to compare the computational cost of the proposed method over state-of-the-art CBIR methods based on feature descriptors. The response time of the CBIR system can be determined by the time it takes to extract the features and retrieve the images. For this purpose, we calculated the average computational complexity from two perspectives: average features computation time and average retrieval time. We have computed these times for

the proposed and baseline models, and the results are shown in Table 5. We calculated the average features extraction time of random query images for all the three image repositories individually such as 33.8 seconds for AT&T, 31.29 seconds for Brodatz Texture, and 29.9 seconds for MIT Vistex. Similarly, we computed the average retrieval time of AT&T, Brodatz texture, and MIT Vistex for the retrieval of 10, 25, and 16 images against multiple query images, i.e., 2.1 seconds for AT&T, 2.21 seconds for Brodatz texture, and 2.7 seconds for MIT Vistex. Then we have taken the average of these query features extraction times, i.e., 31.66 seconds, and retrieval times, i.e., 2.33 seconds, and reported the results of the proposed and comparative methods in Table 5. For fair performance comparison, we ensured to compute the average features computation time and average retrieval time of the same query images for all of the comparative feature descriptors. From this time complexity analysis, we can observe that the CSLBP descriptor achieves the lowest time for both the features computation and retrieval, whereas LTriDP achieves the highest time for both the features extraction and retrieval. The proposed (DMLHP) descriptor achieves 4th place in terms of efficiency among the seven descriptors used in this experiment. It is worth mentioning that the feature dimensions of the CSLBP, LDGP, and LTrP descriptors are 16, 64, and 80, respectively, while the proposed descriptor has a feature dimension of 320. Although our descriptor has large dimension and requires numerous spatial computations to obtain the gradient directions, LDGP and LTrP are just 0.03 and 0.04 seconds faster than our proposed descriptor. This negligible difference in the computational cost of the proposed descriptor over the LDGP and LTrP descriptors is compensated with our method yielding the best average retrieval rate performance. Moreover, the extraction and retrieval time of the proposed descriptor is less than those of the LEPSEG, LTriDP, and LNIP descriptors. This time complexity comparative analysis demonstrates that the proposed system achieves high efficiency as well as effectiveness for the CBIR task.

It is to be noted that all the experiments in our implementation were executed on MATLAB R2018a version, running on the computer system with the following specifications: Intel(R) Core (TM) i3-8130U CPU @2.21 GHz processor and 8 GB RAM. The features computation and retrieval time of the proposed method can be further improved by using a high-performance GPU. Our method has the potential to become more suitable for real-time CBIR applications.

V. CONCLUSION

This paper has presented a novel local texture descriptor that captures the internal structure of an image, based on texture orientation and magnitude, for effective image retrieval. The proposed approach uses 16 directions to represent the visual contents of the image in a robust way due to the formation of orientation and magnitude patterns. Additionally, we employed a learning-based approach to reduce the semantic gap problem. The performance of the proposed descriptor is measured on three standard image repositories that are diverse in terms of pose variations, noise, occlusions, and a variety of natural and artificial regular textures. The experimental results signify the effectiveness of the proposed CBIR system. To measure the retrieval performance of the proposed descriptor, we compared our method with relevant state-of-the-art descriptors. The experimental results indicate the superiority of the proposed method over comparative approaches for image retrieval. Moreover, we have compared the results of both the similarity matching-based approach and the learning-based approach. The comparative results show that the classification-based approach outperforms the conventional similarity matching-based approach by a clear margin, i.e., AT&T 30%, BT 11%, and MIT Vistex 7%. These results demonstrate the effectiveness of the proposed descriptor with a classification-based approach. As compared to some existing feature descriptors, the proposed feature descriptor takes more time to extract a feature vector. Therefore, there is room to improve the computational efficiency of our method. In the future, we plan to enhance the efficiency of the proposed method. Additionally, we will also explore other similarity metric techniques to improve the retrieval results.

REFERENCES

[1] K. R. Dinnes and S. Hariharan, "Approaches and trends in content based image retrieval," in *Proc. 24th Int. Conf. Commun., Netw. Comput. (CNC)*. 2014, pp. 473–477.

[2] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surveys*, vol. 40, no. 2, pp. 1–60, Apr. 2008.

[3] H. Bannour, L. Hlaoua, and B. Ayeb, "Survey of the adequate descriptor for content-based image retrieval on the web: Global versus local features," in *Proc. CORIA*, 2009, pp. 445–456.

[4] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2548–2555.

[5] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[6] S. A. Bakar, M. S. Hitam, and W. N. J. H. Wan Yussof, "Content-based image retrieval using SIFT for binary and greyscale images," in *Proc. IEEE Int. Conf. Signal Image Process. Appl.*, Oct. 2013, pp. 83–88.

[7] R. Desai and B. Sonawane, "Gist, hog, and DWT-based content-based image retrieval for facial images," in *Proc. Int. Conf. Data Eng. Commun. Technol.* Singapore: Springer, 2017, pp. 297–307.

[8] N. Gopal and R. S. Bhooshan, "Content based image retrieval using enhanced SURF," in *Proc. 5th Nat. Conf. Comput. Vis., Pattern Recognit., Image Process. Graph. (NCVPRIPG)*, Dec. 2015, pp. 1–4.

[9] S. R. Dubey, S. K. Singh, and R. K. Singh, "Multichannel decoded local binary patterns for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4018–4032, Sep. 2016.

[10] M. Martolia, N. Dhanore, A. Singh, V. Shahare, and N. Arora, "A modified local binary pattern (LBP) for content-based image retrieval," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 1, pp. 1630–1644, 2020.

[11] M. Fachrurrozi, "Multi-object face recognition using content based image retrieval (CBIR)," in *Proc. Int. Conf. Electr. Eng. Comput. Sci. (ICECOS)*, Aug. 2017, pp. 193–197.

[12] M. Khare, P. Srivastava, J. Gwak, and A. Khare, "A multiresolution approach for content-based image retrieval using wavelet transform of local binary pattern," in *Proc. Asian Conf. Intell. Inf. Database Syst.* Cham, Switzerland: Springer, 2018, pp. 529–538.

[13] P. Srivastava and A. Khare, "Utilizing multiscale local binary pattern for content-based image retrieval," *Multimedia Tools Appl.*, vol. 77, no. 10, pp. 12377–12403, May 2018.

[14] X. Qian, X.-S. Hua, P. Chen, and L. Ke, "PLBP: An effective local binary patterns texture descriptor with pyramid representation," *Pattern Recognit.*, vol. 44, nos. 10–11, pp. 2502–2515, Oct. 2011.
[15] S. Fadaei, R. Amirfattahi, and M. R. Ahmadzadeh, "Local derivative radial patterns: A new texture descriptor for content-based image retrieval," *Signal Process.*, vol. 137, pp. 274–286, Aug. 2017.
[16] A. Nazir, R. Ashraf, T. Hamdani, and N. Ali, "Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor," in *Proc. Int. Conf. Comput., Math. Eng. Technol. (iCoMET)*, Mar. 2018, pp. 1–6.
[17] L. Li, L. Feng, S. L. Liu, M. X. Sun, J. Wu, and H. B. Wang, "Intensity-based co-occurrence local ternary patterns for image retrieval," *J. Comput.*, vol. 29, no. 4, pp. 12–30, 2018.
[18] M. Agarwal, A. Singhal, and B. Lall, "Multi-channel local ternary pattern for content-based image retrieval," *Pattern Anal. Appl.*, vol. 22, no. 4, pp. 1585–1596, Nov. 2019.
[19] J.-X. Zhou, X.-D. Liu, T.-W. Xu, J.-H. Gan, and W.-Q. Liu, "A new fusion approach for content based image retrieval with color histogram and local directional pattern," *Int. J. Mach. Learn. Cybern.*, vol. 9, no. 4, pp. 677–689, Apr. 2018.
[20] S. Unar, X. Wang, C. Wang, and M. Wang, "New strategy for CBIR by combining low-level visual features with a colour descriptor," *IET Image Process.*, vol. 13, no. 7, pp. 1191–1200, May 2019.
[21] V. P. Singh and R. Srivastava, "Improved image retrieval using color-invariant moments," in *Proc. 3rd Int. Conf. Comput. Intell. Commun. Technol. (CICT)*, Feb. 2017, pp. 1–6.
[22] P. Srivastava and A. Khare, "Content-based image retrieval using local ternary wavelet gradient pattern," *Multimedia Tools Appl.*, vol. 78, no. 24, pp. 34297–34322, Dec. 2019.
[23] B. Kakde and M. Okade, "A novel technique for fast content-based image retrieval using dual-cross patterns," in *Proc. 3rd Int. Conf. Converg. Technol. (I CT)*, Apr. 2018, pp. 1–5.
[24] R. Biswas, S. Roy, and D. Purkayastha, "An efficient content-based medical image indexing and retrieval using local texture feature descriptors," *Int. J. Multimedia Inf. Retr.*, vol. 8, no. 4, pp. 217–231, Dec. 2019.
[25] G. V. S. Kumar and P. G. K. Mohan, "Local mean differential excitation pattern for content based image retrieval," *Social Netw. Appl. Sci.*, vol. 1, no. 1, pp. 1–10, Jan. 2019.
[26] T. G. S. Kumar and V. Nagarajan, "Local contourlet tetra pattern for image retrieval," *Signal, Image Video Process.*, vol. 12, no. 3, pp. 591–598, Mar. 2018.
[27] F. Baji and M. Mocanu, "Uniform extended local ternary pattern for content based image retrieval," in *Proc. 22nd Int. Conf. Syst. Theory, Control Comput. (ICSTCC)*, Oct. 2018, pp. 391–396.
[28] G. M. Galshetwar, P. W. Patil, A. B. Gonde, L. M. Waghmare, and R. P. Maheshwari, "Local directional gradient based feature learning for image retrieval," in *Proc. IEEE 13th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Dec. 2018, pp. 113–118.
[29] D. Giveki, M. A. Soltanshahi, and G. A. Montazer, "A new image feature descriptor for content based image retrieval using scale invariant feature transform and local derivative pattern," *Optik*, vol. 131, pp. 242–254, Feb. 2017.
[30] C. Joshi and S. Mukherjee, "Empirical analysis of SIFT, Gabor and fused feature classification using SVM for multispectral satellite image retrieval," in *Proc. 4th Int. Conf. Image Inf. Process. (ICIIP)*, Dec. 2017, pp. 1–6.
[31] D. Patil, S. Krishnan, and S. Gharge, "Medical image retrieval by region based shape feature for CT images," in *Proc. Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput. (COMITCon)*, Feb. 2019, pp. 155–159.
[32] P. Vikhar and P. Karde, "Improved CBIR system using edge histogram descriptor (EHD) and support vector machine (SVM)," in *Proc. Int. Conf. ICT Bus. Ind. Government (ICTBIG)*, 2016, pp. 1–5.
[33] AT&T Laboratories Cambridge. *The Database of Faces*. Accessed: Sep. 16, 2021. [Online]. Available: http://www.cl.cam.ac.uk/research/dtg/attarchive/face database.html
[34] D. C. He and S. Abdelmounaime. *Multiband Texture Database*. Accessed: Sep. 16, 2021. [Online]. Available: http://multibandtexture.recherche.usherbrooke.ca/original_brodatz.html
[35] *MIT VisTex Database*. Accessed: Sep. 16, 2021. [Online]. Available: http://vismod.media.mit.edu/pub/
[36] A. Alzu'bi, A. Amira, and N. Ramzan, "Content-based image retrieval with compact deep convolutional features," *Neurocomputing*, vol. 249, no. 2, pp. 95–105, Aug. 2017.
[37] R. R. Saritha, V. Paul, and P. G. Kumar, "Content based image retrieval using deep learning process," *Cluster Comput.*, vol. 22, no. S2, pp. 4187–4200, Mar. 2019.
[38] S. Hamreras, B. Boucheham, M. A. Molina-Cabello, R. Benítez-Rochel, and E. López-Rubio, "Content based image retrieval by ensembles of deep learning object classifiers," *Integr. Comput.-Aided Eng.*, vol. 27, no. 3, pp. 317–331, May 2020.
[39] A. Shah, R. Naseem, Sadia, S. Iqbal, and M. A. Shah, "Improving CBIR accuracy using convolutional neural network for feature extraction," in *Proc. 13th Int. Conf. Emerg. Technol. (ICET)*, Dec. 2017, pp. 1–5.
[40] A. Raza, H. Dawood, H. Dawood, S. Shabbir, R. Mehboob, and A. Banjar, "Correlated primary visual texton histogram features for content base image retrieval," *IEEE Access*, vol. 6, pp. 46595–46616, 2018.
[41] M. Benco, P. Kamencay, M. Radilova, R. Hudec, and M. Sinko, "The comparison of color texture features extraction based on 1D GLCM with deep learning methods," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Jul. 2020, pp. 285–289.
[42] S. Bhardwaj, G. Pandove, and K. Dahiya, "An analysis of two novel and efficient deep learning models for fast and accurate image retrieval," *J. Inf. Sci. Eng.*, vol. 37, no. 1, pp. 185–201, 2021.
[43] P. Banerjee, A. K. Bhunia, A. Bhattacharyya, P. P. Roy, and S. Murala, "Local neighborhood intensity pattern—A new texture feature descriptor for image retrieval," *Expert Syst. Appl.*, vol. 113, pp. 100–115, Dec. 2018.
[44] M. Verma and B. Raman, "Local tri-directional patterns: A new texture feature descriptor for image retrieval," *Digit. Signal Process.*, vol. 51, pp. 62–72, Apr. 2016.
[45] M. Verma and B. Raman, "Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval," *Multimedia Tools Appl.*, vol. 77, no. 10, pp. 11843–11866, May 2018.
[46] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Local directional gradient pattern: A local descriptor for face recognition," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 1201–1216, Jan. 2017.
[47] C.-H. Yao and S.-Y. Chen, "Retrieval of translated, rotated and scaled color textures," *Pattern Recognit.*, vol. 36, no. 4, pp. 913–929, Apr. 2003.
[48] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with center-symmetric local binary patterns," in *Computer Vision, Graphics and Image Processing*. Berlin, Germany: Springer, 2006, pp. 58–69.
[49] N. Cevik and T. Cevik, "DLGBD: A directional local gradient based descriptor for face recognition," *Multimedia Tools Appl.*, vol. 78, no. 12, pp. 15909–15928, Jun. 2019.
[50] R. Ashraf, M. Ahmed, U. Ahmad, M. A. Habib, S. Jabbar, and K. Naseer, "MDCBIR-MF: Multimedia data for content-based image retrieval by using multiple features," *Multimedia Tools Appl.*, vol. 79, nos. 13–14, pp. 8553–8579, Apr. 2020.
[51] R. Ashraf, K. B. Bajwa, and T. Mahmood, "Content-based image retrieval by exploring bandletized regions through support vector machines," *J. Inf. Sci. Eng.*, vol. 32, no. 2, pp. 245–269, 2016.
[52] S. R. Dubey, S. K. Singh, and R. K. Singh, "Rotation and scale invariant hybrid image descriptor and retrieval," *Comput. Electr. Eng.*, vol. 46, pp. 288–302, Aug. 2015.
[53] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Centre symmetric quadruple pattern: A novel descriptor for facial image recognition and retrieval," *Pattern Recognit. Lett.*, vol. 115, pp. 50–58, Nov. 2018.
[54] S. R. Dubey, S. K. Singh, and R. K. Singh, "Rotation and illumination invariant interleaved intensity order-based local descriptor," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5323–5333, Dec. 2014.

AYESHA KHAN received the B.Sc. degree in computer engineering from UET Taxila, Pakistan, in 2017, where she is currently serving as an M.S. Scholar with the Department of Software Engineering. Her areas of interests include image processing, computer vision, and machine learning.

ALI JAVED (Member, IEEE) received the B.Sc. degree (Hons.) in software engineering and the M.S. and Ph.D. degrees in computer engineering from UET Taxila, Pakistan, in 2007, 2010, and 2016, respectively. He served as an Assistant Professor for the Department of Software Engineering, UET Taxila, where he is currently serving as an Associate Professor. He has also served as the HOD of the Department of Software Engineering, UET Taxila, in 2014. He has served as a Postdoctoral Scholar with the SMILES Laboratory, Oakland University, USA, in 2019, and as a Visiting Ph.D. Scholar with the ISSF Laboratory, University of Michigan, USA, in 2015. His areas of interests include digital image processing, computer vision, video content analysis, multimedia forensics, medical image processing, machine learning, and multimedia signal processing. He was a recipient of various research grants from Qassim University, KSA, HEC Pakistan, National ICT R&D Fund, NESCOM, and UET Taxila. He got selected as an Ambassador of the Asian Council of Science Editors from Pakistan, in 2016. He has also been a member of the Pakistan Engineering Council, since 2007. He received the Chancellor's Gold Medal in M.S. degree in computer engineering.

MUHAMMAD TARIQ MAHMOOD (Senior Member, IEEE) received the M.S. degree in computer science from Blekinge Institute of Technology, Sweden, in 2006, and the Ph.D. degree in informatics and mechatronics from Gwangju University of Science and Technology, Republic of Korea, in 2011. He is currently an Associate Professor with the School of Computer Science and Engineering, Korea University of Technology and Education, South Korea. His research interests include image processing, machine learning, and pattern recognition.

MUHAMMAD HAMZA ARIF KHAN received the B.Sc. degree in computer engineering from the University of Engineering and Technology Taxila, Pakistan, in 2017. His areas of interests include computer vision and machine learning.

IK HYUN LEE (Member, IEEE) received the B.S. degree in control and instrument engineering from Korea University, South Korea, in 2004, and the M.S. and Ph.D. degrees from the School of Information and Mechatronics, Gwangju Institute of Science and Technology, South Korea, in 2008 and 2013, respectively. He was a Postdoctoral Researcher with the Media Lab, Massachusetts Institute of Technology, and a Senior Researcher with Korea Aerospace Research Institute. He is currently an Assistant Professor with the Department of Mechatronics Engineering, Korea Polytechnic University, South Korea. His research interests include image registration, image fusion, depth estimation, and medical image processing.