

2D-LCoLBP: A Learning Two-Dimensional Co-Occurrence Local Binary Pattern for Image Recognition

Xiuli Bi^{ID}, Yuan Yuan, Bin Xiao^{ID}, Weisheng Li^{ID}, *Member, IEEE*, and Xinbo Gao^{ID}, *Senior Member, IEEE*

Abstract—The rotation, scale and translation invariance of extracted features have a high significance in image recognition. Local binary pattern (LBP) and LBP-based descriptors have been widely used in image recognition due to feature discrimination and computational efficiency. However, most of the existing LBP-based descriptors have been designed to achieve rotation invariance while fail to achieve scale invariance. Moreover, it is usually difficult to achieve a good trade-off between the feature discrimination and the feature dimension. In this work, a learning 2D co-occurrence LBP termed 2D-LCoLBP is proposed to address these issues. Firstly, a weighted joint histogram is constructed in different neighborhoods and scales of an image to represent the multi-neighborhood and multi-scale LBP (2D-MLBP) and achieve the rotation invariance. A feature learning strategy is then designed to learn the compact and robust descriptor (2D-LCoLBP) from LBP pattern pairs across different scales in the extracted 2D-MLBP to characterize the most stable local structures and achieve the scale invariance, as well as decrease the feature dimension and improve the noise robustness. Finally, a linear SVM classifier is employed for recognition. We applied the proposed 2D-LCoLBP on four image recognition tasks—texture, object, face and food recognition with ten image databases. Experimental results show that 2D-LCoLBP has obviously low feature dimension but outperforms the state-of-the-art LBP-based descriptors in terms of recognition accuracy under noise-free, Gaussian noise and JPEG compression conditions.

Index Terms—Image recognition, co-occurrence LBP, two-dimensional, multi-scale space, feature learning.

I. INTRODUCTION

FEATURE extraction is a fundamental issue in computer vision, e.g., face recognition [1]–[5], object detection [6], and texture classification [2], [7]–[16]. The discrimination of extracted features directly determines the final performance of

Manuscript received April 9, 2021; revised June 29, 2021; accepted August 2, 2021. Date of publication August 17, 2021; date of current version August 20, 2021. This work was supported in part by the National Key Research and Development Project under Grant 2019YFE0110800 and Grant 2016YFC1000307-3, in part by the National Natural Science Foundation of China under Grant 61806032 and Grant 61976031, in part by the National Major Scientific Research Instrument Development Project of China under Grant 62027827, and in part by the Scientific and Technological Key Research Program of Chongqing Municipal Education Commission under Grant KJZD-K201800601. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mai Xu. (*Corresponding author: Bin Xiao.*)

The authors are with the Department of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: bixl@cqupt.edu.cn; s190201069@stu.cqupt.edu.cn; xiaobin@cqupt.edu.cn; liws@cqupt.edu.cn; gaobx@cqupt.edu.cn).

Digital Object Identifier 10.1109/TIP.2021.3104163

real-world applications suffered from large intra-class variations. According to different feature extraction methods, image features can be categorized into global features [17], [18] and local features [7], [19], [20]. Global features describe holistic representations of images from color, texture, and shape visual cues. The representative ones include the co-occurrence matrix method [17], the filtering-based method [18] and so on. Local features are extracted from regions of interest, which characterizes local corner, edge, and line structures. Compared with global features, local features are more robust to image transformations such as scale, occlusion, and uneven illumination changes. Local features mainly include SIFT [19], LBP [7], SURF [20] and so on. Among those local features, LBP has aroused extensive attention due to its discrimination and computational efficiency.

LBP was proposed by Ojala *et al.* [7] and is a non-parametric local descriptor that encodes the differences between a central pixel and a rectangular neighborhood of 3×3 pixels. Thus, a texture image can be characterized by a probability distribution of 2^8 LBP patterns. In [15], Ojala *et al.* used a circular neighborhood (i.e., $8r$ neighboring pixels evenly distributed on a circle of radius r) instead of a rectangular neighborhood to calculate LBP patterns and further proposed the rotation invariant LBP (LBP^r), the uniform LBP (LBP^{u2}), and the rotation invariant uniform LBP (LBP^{riu2}). Since then, LBP has motivated a large family of LBP-based descriptors, which can be broadly summarized into two categories: individual occurrence LBP [1]–[3], [8]–[13], [16], [21]–[23] that encodes each pattern independently, and the co-occurrence LBP [4], [14], [24]–[27] that utilizes the spatial position relationship to extract strong correlations between adjacent patterns.

To improve the discrimination of individual occurrence LBP, some works have been focused on extracting discriminative information of local regions as much as possible. Guo *et al.* [8] proposed the completed LBP (CLBP) that combines three complementary components (i.e., the sign and magnitude of local differences as well as the central pixels) with building a 3D histogram for texture representation. The local n -ary pattern (LNP) was introduced by Wang *et al.* [9], which explores the discrimination of patterns and formulates the feature extraction process as an integer decomposition problem. While enhancing the discrimination of features, the robustness has also been taken into consideration by

some researchers. Tan and Triggs [2] developed the local ternary pattern (LTP), which quantifies local differences by the ternary pattern. Due to the user-specified threshold, LTP is more discrimination and less sensitive to noise. Liu *et al.* [10] introduced the median robust extended LBP (MRELBP) to capture both microstructure and macrostructure texture information and enhance the noise robustness. Song *et al.* [11] proposed the locally encoded transform feature histogram (LETRIST), which establishes transform features with rotation invariance based on Gaussian derivative filters. Song *et al.* [12] combined the local grouped order pattern (LGOP) and the non-local binary pattern (NLBP) via central pixel encoding to construct discriminative histograms as the texture descriptor (LGONBP). Recently, several learning-based descriptors [3], [13], [16], [21]–[23] have been proposed for dimensionality reduction. Liao *et al.* [13] presented the dominant LBP (DLBP) to establish histograms by selecting the most frequently occurring patterns. Guo *et al.* [16] proposed the scale selective LBP (SSLBP) based on dominant LBP in scale space to achieve scale invariance for texture classification. Duan *et al.* [22] presented the rotation invariant local binary descriptor (RI-LBD) by jointly learning orientations for local patches and hash functions for feature projection. They also proposed a context-aware local binary feature learning (CA-LBFL) [23] for face recognition. The compressive binary pattern (CBP) [3] was designed by replacing the local derivative filters with RF eigenfilters. Although the individual occurrence LBP can obtain high discriminative ability, these features generally have a huge dimension. For example, the feature dimension of SSLBP, RI-LBD and DLBP is up to 2400, 6000 and 14150 respectively. Moreover, most of the individual occurrence LBP cannot keep a stable performance under image scale transformation.

Compared with individual occurrence LBP, co-occurrence LBP captures the spatial contextual information between LBP patterns and provides higher-order statistical information. Therefore, co-occurrence LBP has higher discrimination than individual occurrence LBP. Louis and Plataniotis [4] proposed to use multiple instances of rotational LBP patterns as features instead of histogram bins of LBP patterns, and the multiple features called CoLBP are selected by using the sequential forward selection algorithm. Nosaka *et al.* [24] introduced the co-occurrence of adjacent LBP (CoALBP) by taking auto-correlation matrices calculated from two considered LBP patterns. They further presented the rotation invariant co-occurrence pattern of LBP (RIC-LBP) [25] by incorporating the concept of rotation equivalence class into CoALBP. Qi *et al.* [26] developed the pairwise rotation invariant co-occurrence LBP (PRiCoLBP), which uses a pairwise transform invariance principle. They also introduced the multi-scale co-occurrence LBP (MCLBP) [27] by capturing the correlations among different scales. Xiao *et al.* [14] proposed the 2D local binary pattern (2DLBP) that considers the spatial contextual information between LBP patterns called LBP pattern pairs to describe more subtle structures and performed better than traditional methods. However, the two-stages classifier proposed by 2DLBP requires $r_{max}+1$ SVM classifiers, where r_{max} denotes the number of a circular

neighborhood, and its classification process is very complicated. According to the concept of co-occurrence LBP $[A,B]_{CO}$, its feature dimension is the product of the dimension of feature A and the dimension of feature B , which leads to a sharp increase in feature dimension. Accordingly, the high dimension of co-occurrence LBP is an important issue that needs to be considered during classification. Besides, the scale invariance is also a common issue that was not achieved in these methods.

In summary, discrimination, rotation invariance, scale invariance, and feature dimension are the essential properties for an effective descriptor in image recognition. Both existing individual occurrence LBP and co-occurrence LBP have achieved higher recognition performance by increasing the dimension of features, which brings much redundant information in these features. In this paper, we can capture image information from both local and global perspectives by concatenating histograms of descriptors, which will also be accompanied by an increase in feature dimensionality. Although the above features can completely express image information, the contribution of each pattern of the feature is distinct. Thus, we can adopt a feature learning strategy to preserve useful features. It can maintain or even improve discrimination of descriptors while reducing the feature dimension. Next, scale invariance of the extracted features is also not well considered in existing methods, especially in the co-occurrence LBP. In fact, the characteristic scale varies with the scale of the image. We can also adopt the feature learning strategy to obtain stable scale information, which can be achieved through finding the characteristic scale of local structures of the image.

Based on the aforementioned analyses, this paper proposes a learning 2D co-occurrence LBP (2D-LCoLBP) for image recognition. The major contributions of this paper can be summarized as follows:

- The proposed 2D-LCoLBP considers the description of images in multi-scale space and multi-neighborhood to achieve the scale invariance and enhance the feature discrimination.
- The proposed 2D-LCoLBP utilizes a simple and efficient feature learning strategy on the co-occurrence LBP to obtain the stable local structures for decreasing feature dimension, which makes a well-balanced trade-off between the discriminative ability and feature dimension.
- The proposed 2D-LCoLBP has the complementarity of individual occurrence LBP and co-occurrence LBP, and it owns the robustness of individual occurrence LBP and the discriminative ability of co-occurrence LBP.

This paper is organized as follows. The 2DLBP is firstly briefly reviewed in Section II. In Section III, we described the definition of 2D-LCoLBP. In Section IV, we validated the effectiveness of the proposed 2D-LCoLBP and compared it with the state-of-the-art LBP-based descriptors. Section V concludes this paper.

II. BRIEF REVIEW OF 2DLBP

Co-occurrence LBP has a larger receptive field than individual occurrence LBP and hence can depict more subtle

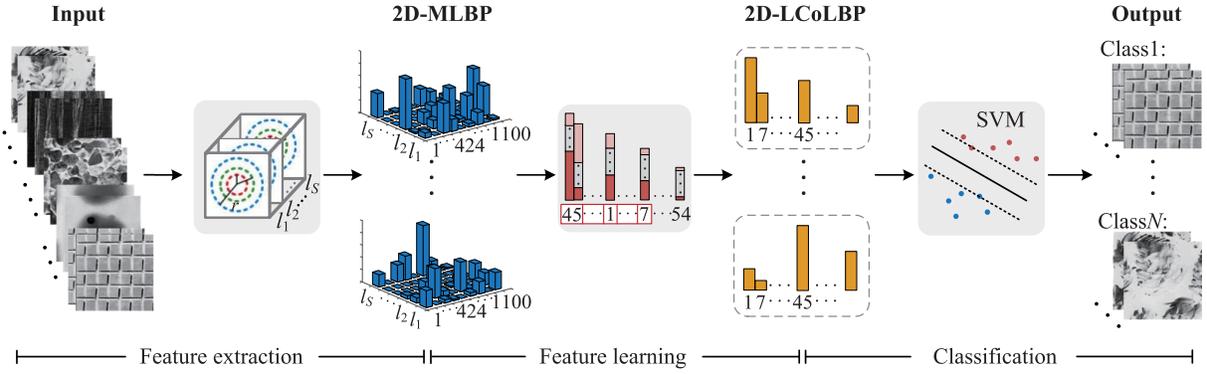


Fig. 1. The framework of the proposed image recognition method.

and complex structures of images, which is more suitable for practical image recognition applications. 2DLBP [14] is a co-occurrence LBP. It is designed for gray-scale texture images and considers the spatial contextual information between LBP patterns to describe more refined and complex structures of images. To obtain 2DLBP with rotation invariance, the rotation invariant uniform LBP (LBP^{riu2}) with $8r+2$ patterns is chosen to extract LBP feature maps:

$$LBP_r^{riu2}(v_c) = \begin{cases} \sum_{n=0}^{8r-1} Q(v_n - v_c), & \text{if } U(LBP_r(v_c)) \leq 2 \\ 8r+1, & \text{otherwise,} \end{cases} \quad (1)$$

where v_c is gray value of the central pixel, v_n is gray value of neighboring pixel that is evenly distributed on a circle of radius r . $Q(\cdot)$ is the sign function, which is defined as

$$Q(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0. \end{cases} \quad (2)$$

$U(\cdot)$ counts bit-wise transitions from “0” to “1” or vice versa, which can be formulated as

$$U(LBP_r(v_c)) = |Q(v_{8r-1} - v_c) - Q(v_0 - v_c)| + \sum_{n=1}^{8r-1} |Q(v_n - v_c) - Q(v_{n-1} - v_c)|. \quad (3)$$

For a given image I with size $M \times N$, firstly, the LBP pattern pair is defined to extract the spatial contextual information of LBP feature map:

$$P_{CO}(t, \Delta t) = (LBP_r^{riu2}(v_t), LBP_r^{riu2}(v_{t+\Delta t})), \quad (4)$$

where $P_{CO}(t, \Delta t) \in \{(0, 0), \dots, (8r+1, 8r+1)\}$, $t = (x, y)$ denotes the coordinates of pixel in the image I with $x \in [r+1, M-r]$ and $y \in [r+1, N-r]$. $\Delta t = (\Delta x, \Delta y)$ represents the spatial position relationship between the LBP patterns in the local area $a \times a$, with $\Delta x \in (-a/2, a/2)$ and $\Delta y \in (-a/2, a/2)$. Then, P_{CO} is globally counted in the LBP feature map:

$$2DLBP_r^I(P_1, P_2) = \sum_{(x,y) \in I} f(P_{CO}, P), \quad (5)$$

with $P = (P_1, P_2) \in \{0 \leq P_1 \leq (8r+1), 0 \leq P_2 \leq (8r+1)\}$, and $f(\cdot)$ is defined as

$$f(P_{CO}, P) = \begin{cases} 1, & \text{if } LBP_r^{riu2}(v_t) = P_1, LBP_r^{riu2}(v_{t+\Delta t}) = P_2 \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

On this basis, $2DLBP_r^I$ can be reshaped as a feature vector $\mathbb{R}^{1 \times (8r+2)^2}$ for representing the image. However, for improving the discriminative ability of 2DLBP, as same as the previous researches [4], [8], [10], [12], [14]–[16], 2DLBP also adopts the multi-neighborhood strategy by concatenating histograms of 2DLBP with different circular neighborhoods ($r=1, 2, 3$):

$$2DLBP^I = [2DLBP_1^I, 2DLBP_2^I, 2DLBP_3^I], \quad (7)$$

where $[\cdot]$ denotes concatenate operation along the neighborhood direction. Therefore, the dimension of 2DLBP is $1 \times (10^2 + 18^2 + 26^2) = 1 \times 1100$ for the image I .

III. THE PROPOSED 2D-LCoLBP

The multi-neighborhood 2DLBP not only captures information from different circular neighborhoods but also extracts the spatial contextual information between LBP patterns. However, 2DLBP is sensitive to image scale transformation. Moreover, it is usually difficult to get a good trade-off between the feature discrimination and feature dimension. For addressing the above issues, a learning 2D co-occurrence LBP (2D-LCoLBP) is proposed. The framework of an image recognition method based on the 2D-LCoLBP is shown in Fig. 1, and it mainly contains three stages. Firstly, a weighted joint histogram is constructed in different neighborhoods and scales of an input image to represent the multi-neighborhood and multi-scale LBP (2D-MLBP). Then, a feature learning strategy is designed on 2D-MLBP to obtain the scale-invariant, compact and robust descriptor (2D-LCoLBP). Finally, a linear SVM classifier is employed for final image recognition.

A. The Multi-Neighborhood and Multi-Scale LBP (2D-MLBP)

The multi-neighborhood strategy combines the responses of multiple descriptors with different radius, which improves the discriminative ability from the local perspective. From the

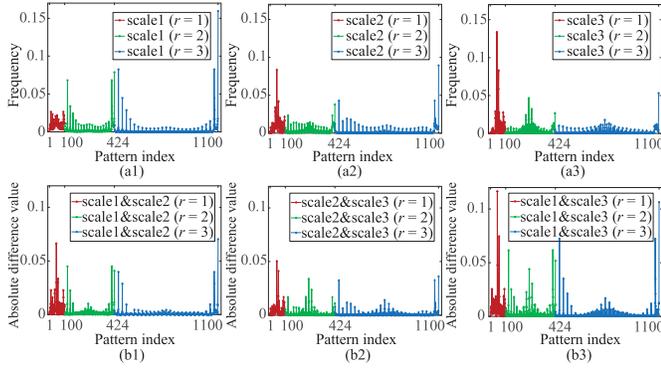


Fig. 2. Illustration of multi-scale analysis on KTH-TIPS. (a1)-(a3) are histograms of 2DLBP. (b1) is absolute differences of histograms between (a1) and (a2). (b2) is absolute differences of histograms between (a2) and (a3). (b3) is absolute differences of histograms between (a1) and (a3).

global perspective, image scaling resizes an image to simulate the change of the distance between observer and object. Witkin [28] and Koenderink [29] proposed the scale-space theory to represent image structures at different scales. The main intention is to separate image structures from the original image, such that the fine-scale image structures that exist at the finest scale can be obtained to represent the image. In this paper, a multi-scale space defined in Eq. (8) is constructed by convoluting an image with the Gaussian kernel function $G(\cdot)$. Note that new structures are not created in the image scale transformation defined in Eq. (8) from a finer to a coarser scale.

$$l_i(x, y) = \begin{cases} I(x, y), & i = 1 \\ G(x, y; \sigma) * l_{i-1}(x, y), & 1 < i \leq S, \end{cases} \quad (8)$$

where l_i is the i -th scaled image of an image I in the multi-scale space, S represents the total number of scaled images, and “*” is the convolution operation. $G(\cdot)$ represents the Gaussian kernel function with the standard deviation σ , which is defined as

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \quad (9)$$

In the multi-scale space, we encoded 2DLBP at three scales. 2DLBP histograms under different scales are shown in Fig. 2 (a1)-(a3). “scale1” represents the original images, “scale2” represents scaled images with $\sigma=1$, and “scale3” represents scaled images with $\sigma=4$. From Fig. 2 (a1)-(a3), we observed that the parts with high frequency in 2DLBP histograms at different scales are distinctly different. The 2DLBP histogram of “scale1” emphasizes the pattern pairs on both sides, the histogram of “scale2” is relatively uniform, and the histogram of “scale3” highlights the pattern pairs in the middle area. The absolute differences of histograms between the scaled images with different σ are shown in Fig. 2 (b1)-(b3), they become more obvious as the difference of σ increase. Theoretically, more information is reflected in the most frequently occurring LBP pattern pairs. Thus, 2DLBP histograms at different scales can describe different representations of the image. We can utilize the complementarity of these pattern pairs to capture more detailed and stable texture information.

Inspired by the above observation, this paper combines multi-scale representations of the original image. On this basis, for encoding the difference information between LBP pattern pairs further, we designed a weight matrix that weights the histograms of 2DLBP in multi-scale space for generating a weighted joint histogram. This weighted joint histogram is denoted as 2D-MLBP in this paper. Fig. 3 illustrates the framework of the proposed 2D-MLBP. For a given image I with size $M \times N$, a multi-scale space (l_1, l_2, \dots, l_S) is firstly derived by Eq. (8), aiming to obtain global representations of an image by convoluting it with Gaussian kernel function while keeping the image size unchanged. Then, the rotation-invariant LBP^{riu2} defined in Eq. (1) is applied to extract LBP feature maps of image at different scales respectively. Thirdly, the multi-neighborhood 2DLBP defined in Eq. (7) is used to extract the contextual information of LBP feature maps in multi-scale space (l_1, l_2, \dots, l_S) :

$$H^{2D} = [2DLBP^{l_1}; \dots; 2DLBP^{l_i}; \dots; 2DLBP^{l_S}], \quad (10)$$

where $[\cdot]$ denotes concatenate operation along the scale direction, $1 \leq i \leq S$, and $H^{2D} \in \mathbb{R}^{S \times 1100}$. Finally, for acquiring the weighted frequency of pattern pairs in H^{2D} , the weighted joint histogram 2D-MLBP can be obtained as

$$2D\text{-MLBP} = W \circ H^{2D}, \quad (11)$$

where “ \circ ” is the Hadamard product, the weight matrix W is composed of weight matrix W_r with different radius ($r=1, 2, 3$), and W_r is defined as

$$W_r(P_1, P_2) = \begin{cases} 1, & \text{if } P_1 = P_2 \\ |P_1 - P_2|^\lambda, & \text{otherwise.} \end{cases} \quad (12)$$

On this basis, W_r is reshaped from $\mathbb{R}^{(8r+2) \times (8r+2)}$ to $\mathbb{R}^{1 \times (8r+2)^2}$, and $W = [W_1, W_2, W_3] \in \mathbb{R}^{1 \times 1100}$. Note that W is broadcasted along the scale direction to $\mathbb{R}^{S \times 1100}$. So, the dimension of 2D-MLBP is $S \times 1100$. From Eq. (12), we can find that pattern pairs with the same difference have the same weight value. Moreover, the weight matrix W_r can be controlled using the weight coefficient λ under the following conditions:

- When $\lambda=0$ and $W_r(P_1, P_2)=1$, only the occurring number of LBP pattern pairs is counted, which is the same as existing co-occurrence LBP.
- When $\lambda>0$ and $W_r(P_1, P_2) \geq 1$, the value of $W_r(P_1, P_2)$ increases with the difference between P_1 and P_2 increases. This situation not only counts the occurring number of pattern pairs but also highlights difference information between pattern pairs.
- When $\lambda<0$ and $0 < W_r(P_1, P_2) \leq 1$, the value of $W_r(P_1, P_2)$ decreases as the difference between P_1 and P_2 increases, which is utilized to weaken the influence of difference information.

The value of λ is universal, and it applies to all co-occurrence LBP. The experimental analysis on the setting of parameter λ will be discussed in details in subsection IV-B.

To further verify the improvement of multi-neighborhood strategy and multi-scale space performed on the 2D-MLBP, we analyzed the recognition accuracy of 2D-MLBP with

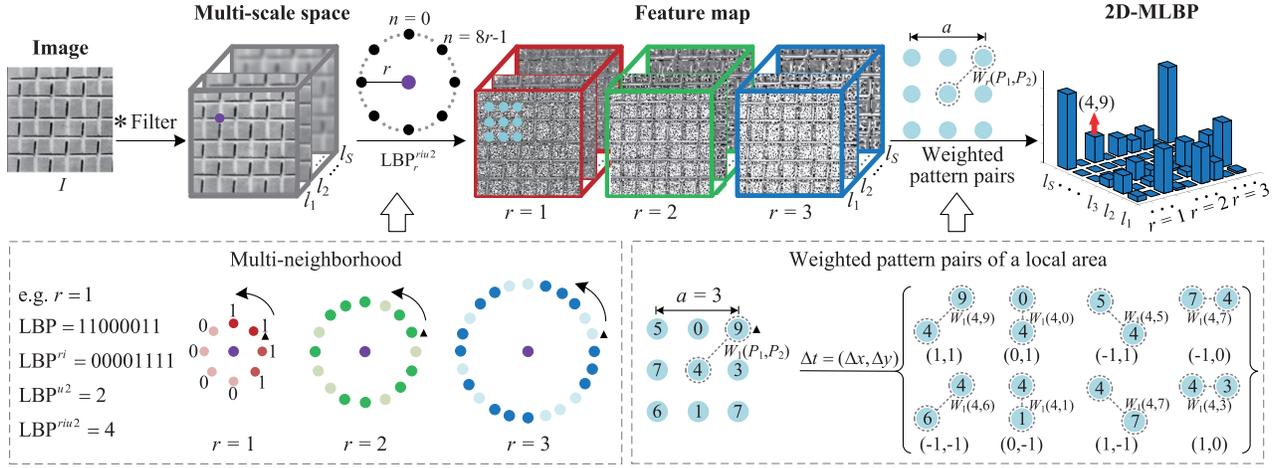


Fig. 3. Illustration of the proposed 2D-MLBP.

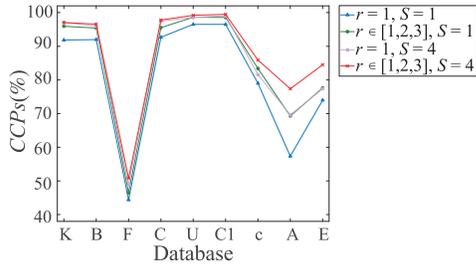


Fig. 4. The $CCPs$ of 2D-MLBP with different neighborhoods and scales on nine databases. The x -axis denotes the abbreviation of databases, i.e., KTH-TIPS(K), Brodatz(B), FMD(F), CURET(C), UMD(U), Coil-100(C1), corel1k(c), AR Face(A), Extended Yale B(E).

different neighborhoods ($r=1$ and $r \in [1, 2, 3]$) and different scales ($S=1$ and $S=4$) on nine image databases. As shown in Fig. 4, the $CCPs$ of 2D-MLBP with $r \in [1, 2, 3]$ and $S=1$ (i.e., multi-neighborhood and original image) is significantly better than that with $r=1$ and $S=1$ (i.e., single neighborhood and original image), which demonstrates the effective of multi-neighborhood strategy. Moreover, the $CCPs$ of 2D-MLBP with $r=1$ and $S=4$ (i.e., single neighborhood and the multi-scale space) is also better than that with $r=1$ and $S=1$, which proves that the multi-scale space is effective. It can also be seen that the $CCPs$ of 2D-MLBP with $r \in [1, 2, 3]$ and $S=4$ (i.e., multi-neighborhood and multi-scale space) has the highest recognition accuracy. These results show that the combination of multi-neighborhood strategy and multi-scale space can capture more stable texture information and further promote the recognition performance of 2D-MLBP.

B. A Learning 2D Co-Occurrence LBP (2D-LCoLBP)

For LBP-based descriptors, almost all methods improve the discriminative ability and robustness of descriptors by concatenating histograms of descriptors. We also adopted this idea in 2D-MLBP. However, this way will cause an increase in feature dimension. Meanwhile, are all pattern pairs the same roles when an image is scaled or rotated? And are all pattern pairs useful for representing an image? After exploring these

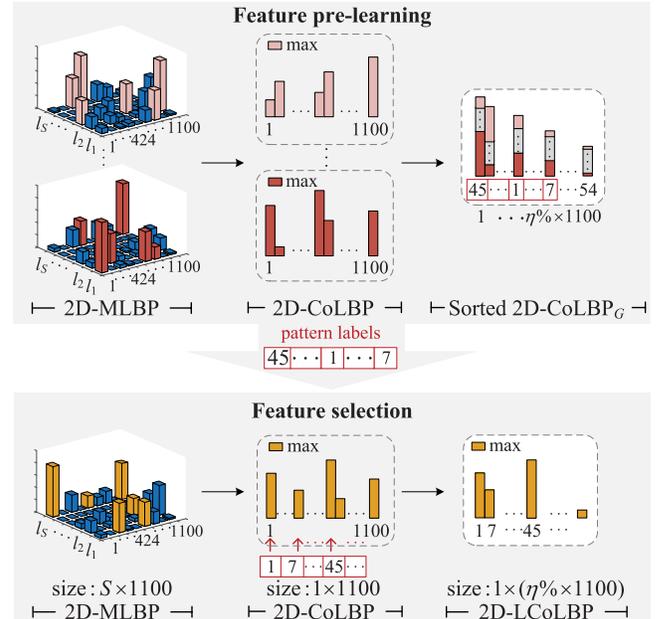


Fig. 5. Illustration of a feature learning strategy.

two doubts, we proposed a feature learning strategy as shown in Fig. 5, which contains two stages: feature pre-learning and feature selection. This strategy aims to learn the compact and robust descriptor (2D-LCoLBP) across different scales in 2D-MLBP and characterize the most stable local structures by 2D-LCoLBP for achieving the scale invariance, a smaller feature dimension and the robustness to noise.

For the first doubt mentioned above, Mikolajczyk [30] demonstrated that the scale selection technique based on the extrema of a scale space representation is the most reliable method for determining the characteristic scale of a local structure because the descriptor computed at this scale conveys more information comparing to descriptors at other scales. Thus, we can choose the extrema with more robustness against image scale, noise and illumination transformations, which can be accomplished by searching for stable features across all possible scales. Inspired by this idea, in the feature

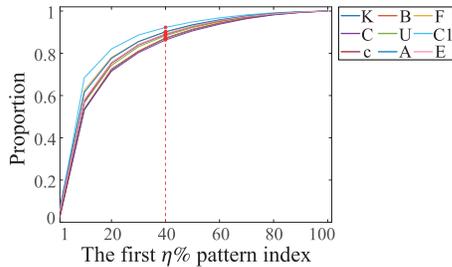


Fig. 6. Proportion of the first $\eta\%$ LBP pattern pairs among all LBP pattern pairs of images on nine image databases, i.e., KTH-TIPS(K), Brodatz(B), FMD(F), CURET(C), UMD(U), Coil-100(C1), corel1k(c), AR Face(A), Extended Yale B(E).

pre-learning stage, for 2D-MLBP of an image $I \in T$ (T is a learning set), a cross-scale co-occurrence LBP (2D-CoLBP) can be computed as

$$2D-CoLBP(b) = \max(2D-MLBP(1, b), \dots, 2D-MLBP(i, b), \dots, 2D-MLBP(S, b)), \quad (13)$$

where $1 \leq b \leq 1100$. In this way, the stable LBP pattern pairs of image $I \in T$ across all scales are kept.

For the second doubt, as pointed out by [13], [15], occurring frequencies of different patterns vary greatly, and some of the patterns rarely occur in an image. The proportion of these patterns are too small and inadequate to provide a reliable estimation for occurring possibilities of these patterns. To discuss whether LBP pattern pairs with tiny proportions are negligible, we calculated the proportion of pattern pairs occupied by different $\eta\%$ of most frequently occurring pattern pairs in the stable features. As illustrated in Fig. 6, the proportion increases with η increases on nine image databases, which reaches 0.85-0.95 at $\eta = 40$, indicating that this part is a preponderance of the overall pattern pairs.

Based on this analysis, in the feature pre-learning stage, after generating 2D-CoLBP of each image in a learning set, a global cross-scale co-occurrence LBP (2D-CoLBP_G) is counted by accumulating the 2D-CoLBP of each image in this learning set, which can be defined as

$$2D-CoLBP(b) = \sum 2D-CoLBP(b). \quad (14)$$

2D-CoLBP_G is sorted in descending order. The first $\eta\%$ LBP pattern pairs in sorted 2D-CoLBP_G are retained as pattern labels $PL_T[1, 2, \dots, \eta\% \times 1100]$ that contain a majority of image information.

In the feature selection stage, for 2D-MLBP of an image E , we firstly extract 2D-CoLBP that is learned across different scales in 2D-MLBP using Eq. (13). Then, the pattern labels PL_T established in the feature pre-learning stage are performed on 2D-CoLBP for selecting the pattern pairs to obtain the learning 2D co-occurrence LBP (2D-LCoLBP) with scale invariance and relatively low feature dimension. The feature dimension is reduced from $S \times 1100$ to $1 \times (\eta\% \times 1100)$.

We then discussed the recognition performance of 2D-MLBP, 2D-CoLBP, and 2D-LCoLBP on nine image databases. The feature dimensions of 2D-MLBP, 2D-CoLBP, and 2D-LCoLBP are 4400, 1100, and 440, separately. As illustrated in Fig. 7, 2D-CoLBP maintains the recognition

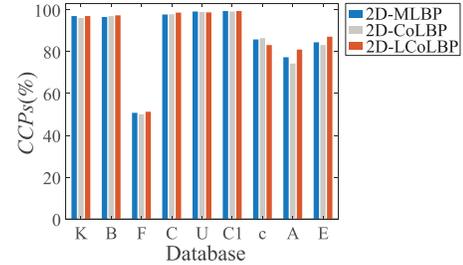


Fig. 7. The $CCPS$ of 2D-MLBP, 2D-CoLBP, and 2D-LCoLBP on nine image databases. The x -axis denotes the abbreviation of databases, i.e., KTH-TIPS(K), Brodatz(B), FMD(F), CURET(C), UMD(U), Coil-100(C1), corel1k(c), AR Face(A), Extended Yale B(E).

accuracy comparable to that of 2D-MLBP while greatly reducing the feature dimension. Moreover, 2D-LCoLBP with the lowest feature dimension has the highest recognition accuracy on seven of nine databases. Thus, each step of the feature learning strategy is effective. The discriminative ability of these descriptors is maintained or even enhanced while removing redundant information.

C. The Scale Invariance of 2D-LCoLBP

In this subsection, KTH-TIPS [31] with an available ground-truth of scales is used to illustrate that the proposed 2D-LCoLBP has scale invariance. KTH-TIPS exhibits texture images from ten different materials captured at nine different scales with nine samples per material. We divided KTH-TIPS into nine subsets (i.e., s1-s9) according to image scale parameters, so each subset includes 90 (10×9) images. In this experiment, we compared 2D-LCoLBP with LBP^{riu2} [15], SSLBP [16], MCLBP [27], and 2DLBP [14]. For 2D-LCoLBP, the first 54, 440, 886, and 1100 LBP pattern pairs are selected to maintain consistency with the number of bins in histograms of LBP^{riu2}, SSLBP, MCLBP, and 2DLBP, separately. Then, we used the Euclidean distance $Dist$ to quantify the similarity of histograms.

$$Dist(H_1, H_2) = \sqrt{\sum_{k \in K} (H_1(k) - H_2(k))^2}, \quad (15)$$

where H_1 and H_2 are histograms, $H_1(k)$ and $H_2(k)$ ($1 \leq k \leq K$) are the k -th bin of them respectively and K denotes the total number of bins in the histogram.

Fig. 8 illustrates the $Dist$ of different methods. From Fig. 8, we observed that for all methods, their $Dist$ increase as the gap of image scale between two subsets increases. The $Dist$ of 2D-LCoLBP is much smaller than LBP, SSLBP, MCLBP, and 2DLBP. In addition, to show the scale invariance of the proposed 2D-LCoLBP more intuitively, the results of two images are given in Fig. 9. Absolute differences of histograms between Fig. 9 (a) and (b) of different methods are shown in Fig. 9 (c3), (d3), (e3), and (f3). We found that the absolute differences of 2D-LCoLBP histograms are much smaller than that of these methods. These results indicate that 2D-LCoLBP with scale invariance is more stable than LBP^{riu2}, SSLBP, MCLBP, and 2DLBP under image scale transformation.

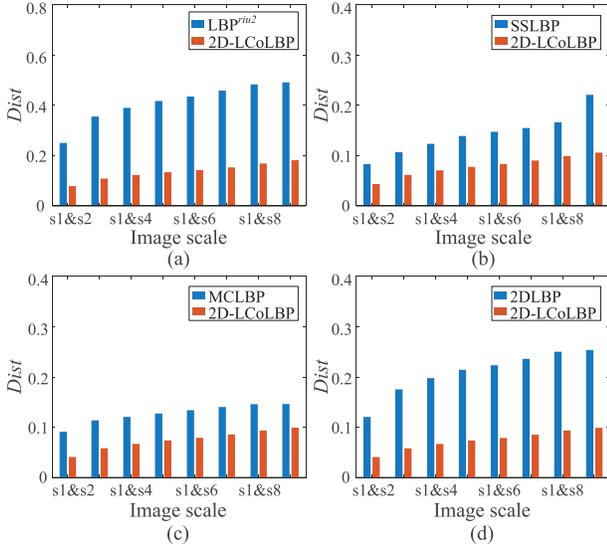


Fig. 8. The *Dist* of different methods on the nine subsets of KTH-TIPS. (a) is the *Dist* of LBP^{riu2} and 2D-LCoLBP. (b) is the *Dist* of SSLBP and 2D-LCoLBP. (c) is the *Dist* of MCLBP and 2D-LCoLBP. (d) is the *Dist* of 2DLBP and 2D-LCoLBP.

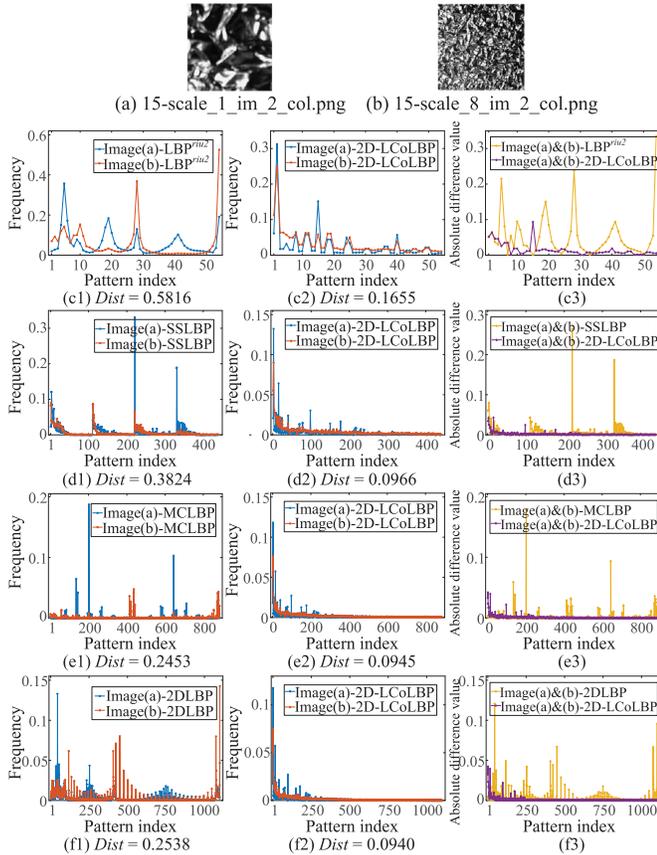


Fig. 9. The scale invariance of the proposed 2D-LCoLBP. (a) and (b) display two images of “aluminium_foil” class from KTH-TIPS with different image scale parameters. (c1)(c2), (d1)(d2), (e1)(e2), and (f1)(f2) are the histograms of different methods on these two images. (c3), (d3), (e3), and (f3) are absolute differences of histograms.

IV. EXPERIMENTAL RESULTS AND ANALYSES

To validate the effectiveness of 2D-LCoLBP, we carried out five groups of experiments in this section. In subsection IV-A,

TABLE I
THE BASIC INFORMATION OF TEN DATABASES

Type	Name	Image Size	Total Number of Images	Transformations		
				Rotation	Scale	Illumination
Texture	KTH-TIPS	200×200	810	✓	✓	✓
	Brodatz	213×213	999			
	FMD	512×384	1000	✓	✓	✓
	CUReT	200×200	5612	✓		✓
	UMD	1280×960	1000	✓	✓	✓
Object	Coil-100	128×128	7200	✓	✓	
	core11k	256×384	1000	✓	✓	
Face	AR Face	120×165	2600			✓
	Extended Yale B	168×192	2414			✓
Food	RawFoot	800×800	3128			✓

the image databases and experimental setup are introduced. In subsection IV-B, parameter settings of 2D-LCoLBP are discussed. In subsection IV-C, we analyzed the effectiveness of 2D-LCoLBP. In subsection IV-D, the geometric invariance of 2D-LCoLBP is verified. At last, we compared 2D-LCoLBP with the state-of-the-art LBP-based methods in terms of texture classification, object, face and food recognition in subsection IV-E and IV-F respectively.

A. Databases and Experimental Setup

1) *Databases*: There are ten databases employed for evaluation, which can be divided into four categories—texture, object, face and food databases. Five texture databases include KTH-TIPS [31], Brodatz [32], FMD [33], CUReT [34], and UMD [35]. Two object databases include Coil-100 [6] and core11k [36]. Two face databases include AR Face [37] and Extended Yale B [38]. The food database is RawFoot [39]. For clarity, we summarized the basic information of ten databases in Table I.

KTH-TIPS is an image database containing ten kinds of materials, and each material was captured by nine different scales, three different poses, and three different illuminations. FMD contains ten classes, including fabric, foliage, glass, etc. Each image in FMD is associated with a binary human-labeled mask, describing the location of an object. We only extracted the features within the masked regions. For CUReT, we utilized the same subset of images as [40], which contains 61 classes with 92 images in each class. These images were captured under different illuminations and viewpoint directions. AR Face contains 126 individuals and over 4000 frontal images. We used images of 100 subjects (i.e., 50 males and 50 females). For Extended Yale B, we adopted the same subset of images as [5], which contains a total of 38 subjects with severe illumination variations. All the frontal face images have already been manually aligned and cropped to 168×192 . In our experiments, each image of face databases was divided into four blocks with the same size. The size of each block in AR Face is 60×82 , and that in Extended Yale B is 84×96 . Note that the feature dimension of the partitioned face database is four times higher than that of the original face database. RawFoot consists of 68 texture classes of raw food, with each class having 46 images acquired under 46 lighting conditions.

TABLE II
IMAGE DATABASES USED IN THE EXPERIMENTAL SECTION

Name	Parameters			Texture Databases					Object Databases		Face Databases	
	Variable	Range	Step	KTH-TIPS	Brodatz	FMD	CUReT	UMD	Coil-100	corel1k	AR Face	Extended Yale B
Original	-	-	-	810	999	1000	5612	1000	7200	1000	2600	2414
Rotated	<i>deg</i>	[0°, 360°]	random	810	999	1000	5612	1000	7200	1000	-	-
Scaling	β	[0.6, 3.0]	0.2	810	999	1000	5612	1000	7200	1000	-	-
Gaussian noise	<i>var</i>	[0.02, 0.1]	0.02	810×5	999×5	1000×5	5612×5	1000×5	-	-	-	-
JPEG compression	<i>QF</i>	[30, 90]	20	810×4	999×4	1000×4	5612×4	1000×4	-	-	-	-

In the experimental section, to analysis the discriminative ability, rotation invariance, scale invariance, and robustness under Gaussian noise and JPEG compression conditions of the proposed 2D-LCoLBP, we conducted the corresponding operations on the above databases as shown in Table II. The details are described as follows:

- Image rotation transformation: We first resized each image to 1.5 times of its original size through bicubic interpolation. Then, we rotated the image with a random angle $deg \in [0^\circ, 360^\circ)$, and followed by cropping to the same size as the original image from the center of the rotated image. This step guarantees that no black space is cropped out. Note that the rotating and cropping operations may lead to information loss.
- Image scale transformation: We resized each image to a random β time of its original size through bicubic interpolation with β varying from 0.6 to 3.0 with a step size of 0.2. The resizing operation is used to simulate different scales of the image, but it may lead to information loss.
- Gaussian noise: Gaussian noise is added to each image with the mean value of 0, and the variance *var* varies from 0.02 to 0.1 with a step size of 0.02.
- JPEG compression: each image was recompressed in JPEG format with quality factors (*QF*) varying from 90 to 30 with a decrement of 20.

We did not utilize face databases under image transformations, mainly because the resizing, rotating, and cropping operations in the process of making database cannot guarantee the integrity of face, resulting in serious loss of image content. We only added Gaussian noise and JPEG compression attacks on texture databases. Because texture images have inter-class ambiguities and large intra-class variations. After adding the above attacks, the difficulty of recognition task will be greatly increased.

2) *Evaluation Metrics*: For evaluating the recognition accuracy of different methods, the correct classification percentages *CCPs* is defined as

$$CCPs = \frac{\text{the number of correctly classified images}}{\text{the total number of classified images}} \times 100\%. \tag{16}$$

We randomly chose a half of images per class for training and the remaining images for testing. This split was implemented for 10 times randomly, and the average *CCPs* over the 10 splits was used as the final evaluation result.

TABLE III
IMPLEMENTATION DETAILS OF THE PROPOSED 2D-LCoLBP AND THE STATE-OF-THE-ART METHODS

Methods	Implementation Details
LBP ^{riu2} ₁₊₂₊₃	LBP ^{riu2} ₁₊₂₊₃ ; (<i>r, P</i>)={ (1, 8), (2, 16), (3, 24) }
MRELBP	MRELBP ^{num} ₂₊₄₊₆₊₈ ; (<i>r, P</i>)={ (2, 8), (4, 8), (6, 8), (8, 8) }
LETRIST	(σ, w)={ (1, 7), (2, 13), (4, 25) }; a scale factor <i>c</i> =1
LGONBP	(<i>r, P, w</i>) = { (3, 24, 1), (5, 24, 3), (7, 24, 5) }; anchors <i>J</i> =3
CBFD	<i>r</i> =3; $\lambda_1=0.001, \lambda_2=0.0001$; the codebook size as 500
SSLBP	(<i>r, P</i>)={ (3, 24), (9, 24) }; a feature length <i>K</i> =600
CA-LBFL	<i>r</i> =3; $\lambda_1=10^3, \lambda_2=10^2, \lambda_3=10^9$; the dictionary size as 500
CoALBP	($\Delta r, \Delta s$)={ (2, 1), (4, 2), (8, 4) }; auto-correlation matrix 2×2
MCLBP	MCLBP/C; (<i>r, P</i>)={ (1, 8), (3, 8) }
2DLBP	2DLBP ^{riu2} ₁₊₂₊₃ ; (<i>r, P</i>)={ (1, 8), (2, 16), (3, 24) }; <i>w</i> =3
2D-LCoLBP	<i>r</i> ∈ [1, 2, 3]; <i>S</i> =4; $\sigma=2^{1/4}, w=9; a=3, \lambda=0.5; \eta=40$

3) *Implementation Details*: The relevant parameters of 2D-LCoLBP is illustrated in Table III. We randomly selected 25% images per class as a learning set *T* for feature learning. For classification, we used the publicly available LibSVM library [41]. We compared 2D-LCoLBP that is a learning-based co-occurrence LBP with 10 state-of-the-art LBP-based methods which can be divided into three categories: non-learning individual occurrence LBP (LBP^{riu2} (2002, [15]), MRELBP (2016, [10]), LETRIST (2018, [11]), and LGONBP (2020, [12])), learning-based individual occurrence LBP (CBFD (2015, [21]), SSLBP (2016, [16]), and CA-LBFL (2018, [23])), and non-learning co-occurrence LBP (CoALBP (2011, [24]), MCLBP (2015, [27]), and 2DLBP (2019, [14])). The corresponding authors provide the source codes of the comparative LBP-based methods, and the default parameters and classifiers provided by the authors are adopted to keep consistency with the results given in the original papers. The specific experimental parameters are also shown in Table III. All methods are implemented on a PC with a 2.40 GHz CPU.

B. Parameter Evaluation of the Proposed 2D-LCoLBP

The proposed 2D-LCoLBP involves four main parameters that need to be evaluated, i.e., *S* in Eq. (8), σ and *w* in Eq. (9), *a* in Eq. (4), as well as λ in Eq. (12). The suitable

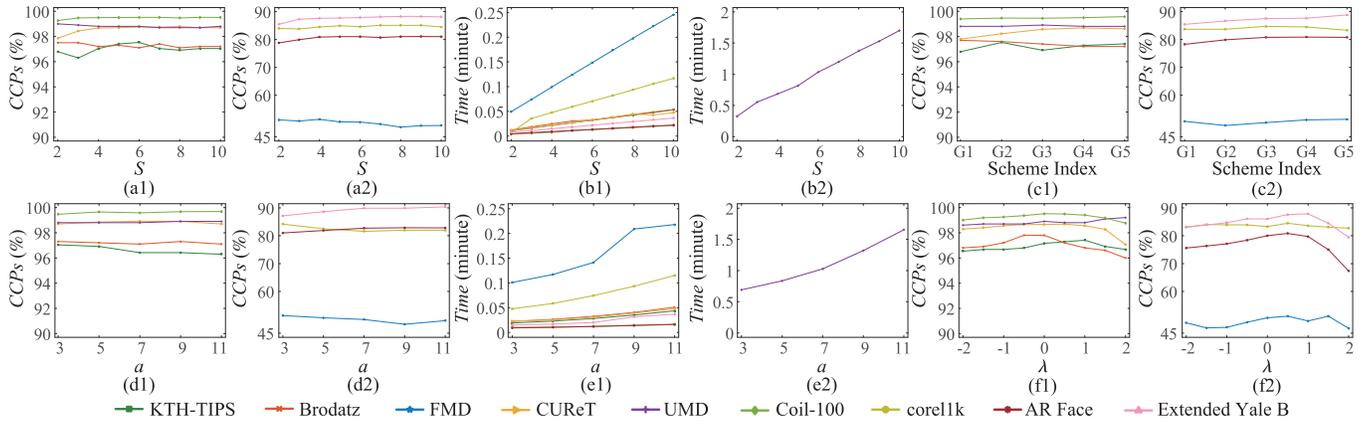


Fig. 10. Parameter evaluation of the proposed 2D-LCoLBP on nine image databases. (a1) and (a2) show the $CCPs$ of 2D-LCoLBP with S . (b1) and (b2) show the $Time$ of 2D-LCoLBP with S . (c1) and (c2) show the $CCPs$ of 2D-LCoLBP with schemes. (d1) and (d2) show the $CCPs$ of 2D-LCoLBP with $a \times a$. (e1) and (e2) show the $Time$ of 2D-LCoLBP with $a \times a$. (f1) and (f2) show the $CCPs$ of 2D-LCoLBP with λ .

parameters setting can make 2D-LCoLBP to be applicable for general image recognition applications. In this subsection, we evaluated these specific parameters on nine databases by fixing others as default values.

1) *Evaluation of Multi-Scale Space S* : Fig. 10 (a1) and (a2) illustrate the $CCPs$ of 2D-LCoLBP with S varying from 2 to 10 with step 1 on nine databases. From these figures, We observed that the $CCPs$ of 2D-LCoLBP on most databases firstly increases as S increases, then reaches a peak at $S=4$, and finally tends to stable gradually. As the number of scale S increases, the detailed information of the image gradually decreases after Gaussian filtering, and its computational complexity also increases significantly. Fig. 10 (b1) and (b2) show the average running time ($Time$: minute) of feature extraction process on each image by 2D-LCoLBP with different S . As shown in these figures, the $Time$ of 2D-LCoLBP increases as S increases. For KTH-TIPS, the $Time$ of 2D-LCoLBP with $S=6$ is 1.72 times of that with $S=4$. Therefore, choosing $S=4$ can effectively balance the relationship between recognition accuracy and computational efficiency.

2) *Evaluation of the Standard Deviation σ and Window Size $w \times w$* : The degree of image smoothing depends on the σ and $w \times w$. Increasing σ can enhance the influence of surrounding pixels on the central pixel and make images smoother, and vice versa. The value of w is often related to the value of σ . Fig. 10 (c1) and (c2) show the results of the proposed 2D-LCoLBP using five schemes (σ, w) : $G1=(2^{-1/8}, 3)$, $G2=(2^0, 5)$, $G3=(2^{1/8}, 7)$, $G4=(2^{1/4}, 9)$, and $G5=(2^{1/2}, 11)$. From these figures, we found that $G4$ can get a suitable smoothing effect and obtains the best recognition accuracy compared with $G1$, $G2$, $G3$ and $G5$.

3) *Evaluation of the Local Area $a \times a$* : Fig. 10 (d1) and (d2) show that the $CCPs$ of the proposed 2D-LCoLBP with different $a \times a$. For KTH-TIPS, Brodatz, FMD, and core11k, 2D-LCoLBP achieves the highest $CCPs$ when $a=3$, while the value of a has little effect on CURET, UMD, Coil-100, AR Face and Extended Yale B. Theoretically, the larger the local area $a \times a$, the more contextual information of P_{CO}

is involved in the proposed 2D-LCoLBP. However, the correlation of P_{CO} decreases with the increase of a , which causes the $CCPs$ to decline or remain unchanged. In addition, Fig. 10 (e1) and (e2) show the $Time$ of 2D-LCoLBP with different $a \times a$. From these figures, the $Time$ of 2D-LCoLBP increases with a increases. Therefore, the size of $a \times a$ is set to 3×3 by considering the balance between the recognition accuracy and computational efficiency.

4) *Evaluation of the Weight Coefficient λ* : As discussed in subsection III-A, the value of λ is universal. To determine the optimal weight coefficient, the value of λ is varied from -2 to 2 , with an increment of 0.5 . Fig. 10 (f1) and (f2) show the $CCPs$ of the proposed 2D-LCoLBP with different λ . As illustrated in Fig. 10 (f1) and (f2), with the increase of λ , the $CCPs$ increases and subsequently decreases. For all databases, the $CCPs$ of $\lambda \in [-2, 0]$ is generally lower than that of $\lambda \in [0, 1]$. It can further illustrate that 2D-LCoLBP not only counts the occurring number of pattern pairs but also highlights difference information between pattern pairs. Therefore, the weight coefficient λ is suggested as 0.5 in the following experiments.

C. The Effectiveness Analysis of the Proposed 2D-LCoLBP

In this experiment, we discussed the effect of different $\eta\%$ on the recognition performance of 2D-LCoLBP to demonstrate the effectiveness of a feature learning strategy. As shown in Fig. 11, the value of η varies from 10 to 100 with an increment of 10. From this figure, it can be found that, when η is up to 40, increasing the value of η does not improve the $CCPs$ of 2D-LCoLBP on all image databases. Therefore, the feature learning strategy can balance the relationship between the discriminative ability and feature dimension, and remove redundant information.

D. Geometric Invariance of the Proposed 2D-LCoLBP

In this subsection, we validated the rotation invariance and scale invariance of the proposed 2D-LCoLBP. Experiments are conducted on seven rotated databases and seven scaling databases separately.

TABLE IV
THE $CCPs$ OF THE PROPOSED 2D-LCoLBP AND THE STATE-OF-THE-ART METHODS ON SEVEN NEW ROTATED DATABASES

Methods	Feature Dimension	The Rotated Texture Databases					The Rotated Object Databases	
		KTH-TIPS	Brodatz	FMD	CUReT	UMD	Coil-100	corel1k
LBP ^{riu2}	54	79.88	76.08	32.60	74.04	83.50	81.71	57.20
MRELBP	1568	95.12	90.09	36.00	95.55	96.64	96.21	65.31
LETRIST	413	94.63	92.43	24.25	95.69	95.87	90.29	59.00
LGONBP	1404	95.24	89.98	23.70	92.85	95.90	91.22	54.50
CBFD	500	71.71	50.68	26.40	49.29	69.40	55.78	45.80
SSLBP	2400	94.76	89.75	29.80	95.65	93.70	96.49	60.40
CA-LBFL	500	73.17	54.28	23.40	54.13	71.60	59.33	46.80
CoALBP	3072	71.32	55.05	24.41	61.87	79.34	71.21	49.78
MCLBP	886	94.27	89.98	18.80	93.41	94.20	90.01	55.40
2DLBP	1100	93.52	86.04	27.79	86.86	94.71	95.04	51.91
2D-LCoLBP	440	95.56	94.00	37.10	96.10	96.70	97.21	66.30

TABLE V
THE $CCPs$ OF THE PROPOSED 2D-LCoLBP AND THE STATE-OF-THE-ART METHODS ON SEVEN NEW SCALING DATABASES

Methods	Feature Dimension	The Scaling Texture Databases					The Scaling Object Databases	
		KTH-TIPS	Brodatz	FMD	CUReT	UMD	Coil-100	corel1k
LBP ^{riu2}	54	79.88	76.08	32.60	74.04	83.50	81.71	57.20
MRELBP	1568	77.19	52.93	30.60	65.24	82.30	74.86	67.90
LETRIST	413	76.71	55.18	20.24	70.31	79.60	38.58	54.70
LGONBP	1404	74.39	50.56	16.90	64.58	71.80	37.94	51.50
CBFD	500	88.78	77.70	22.20	81.60	87.80	88.19	61.80
SSLBP	2400	86.83	76.46	28.30	79.62	94.20	78.17	66.80
CA-LBFL	500	85.85	79.17	23.20	82.47	89.40	87.99	64.50
CoALBP	3072	71.38	54.93	24.69	61.58	79.15	71.04	49.62
MCLBP	886	83.78	71.40	24.60	69.96	86.40	87.49	59.30
2DLBP	1100	77.96	56.76	18.82	63.36	87.29	88.83	53.82
2D-LCoLBP	440	89.88	79.28	38.80	83.70	94.30	91.11	75.60

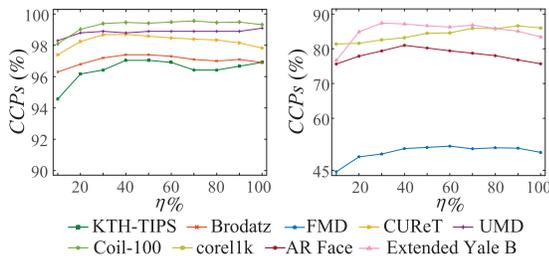


Fig. 11. The $CCPs$ of the proposed 2D-LCoLBP with different $\eta\%$.

1) *Rotation Invariance*: Because PCO is composed of two LBP^{riu2}, it is easy to deduce that 2D-LCoLBP has rotation invariance. Table IV shows the experimental results by different methods on seven rotated databases. It can be clearly seen that the $CCPs$ of the proposed 2D-LCoLBP is significantly higher than the state-of-the-art methods. Especially on rotated Brodatz, the $CCPs$ of 2D-LCoLBP outperforms most recently LETRIST method that has rotation invariance with 1.57% improvement. These experimental results illustrate the rotation invariance of 2D-LCoLBP.

2) *Scale Invariance*: In this experiment, we validated the scale invariance of 2D-LCoLBP. Table V shows the

experimental results by different methods on seven scaling databases. We observed that the $CCPs$ of 2D-LCoLBP is significantly higher than other comparative methods. The $CCPs$ of 2D-LCoLBP on scaling corel1k reaches 75.60%, which outperforms MRELBP with 7.7%. The main reason is that 2D-LCoLBP can capture scale information.

E. Texture Classification

For evaluating the performance of 2D-LCoLBP on five texture databases, three experiments are carried out in this subsection. The first experiment focuses on texture classification by 2D-LCoLBP, and compares it with 10 state-of-the-art methods. In the second experiment, we compared 2D-LCoLBP with these methods for texture classification under both noise-free and Gaussian noise conditions. In the third experiment, we evaluated the robustness of 2D-LCoLBP under JPEG compression condition.

1) *Comparison With the State-of-the-Art Methods*: Table VI lists the $CCPs$ of 2D-LCoLBP and the state-of-the-art methods on five texture databases. We observed that 2D-LCoLBP with relatively low feature dimension achieves the best results on four of five databases, i.e., Brodatz, CUReT, UMD, and FMD. On Brodatz, the $CCPs$ of 2D-LCoLBP outperforms

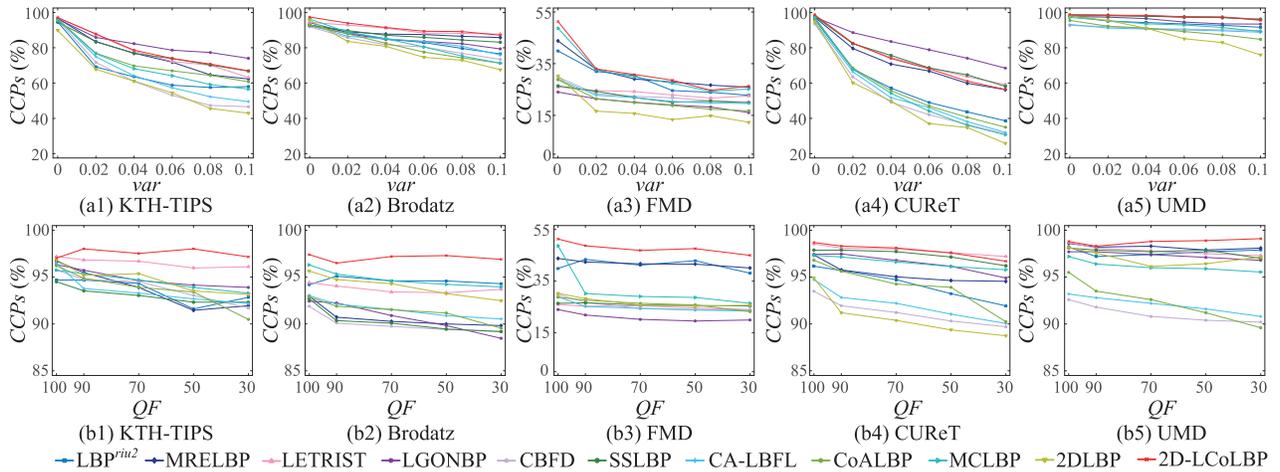


Fig. 12. Comparison of classification performance by different methods under Gaussian noise and JPEG compression on five texture databases. (a1)-(a5) KTH-TIPS, Brodatz, FMD, CUREt and UMD with different levels of Gaussian noise. (b1)-(b5) KTH-TIPS, Brodatz, FMD, CUREt and UMD with different quality factors of JPEG compression.

TABLE VI

THE $CCPs$ OF THE PROPOSED 2D-LCoLBP AND THE STATE-OF-THE-ART METHODS ON FIVE TEXTURE DATABASES

Methods	Feature Dimension	Texture				
		KTH-TIPS	Brodatz	FMD	CUREt	UMD
LBP ^{riu2}	54	94.69	94.19	39.90	96.17	98.20
MRELBP	1568	96.71	92.97	43.80	97.36	98.60
LETRIST	413	97.20	94.41	26.00	98.54	98.70
LGONBP	1404	96.34	92.43	24.00	97.43	97.70
CBFD	500	95.75	91.89	29.80	93.48	92.60
SSLBP	2400	94.51	92.70	26.40	97.88	98.10
CA-LBFL	500	96.50	93.02	29.00	94.73	93.20
CoALBP	3072	96.22	92.88	28.80	96.81	95.50
MCLBP	886	95.73	96.31	48.70	97.33	97.20
2DLBP	1100	96.85	95.65	30.30	94.92	98.15
2D-LCoLBP	440	97.04	97.40	51.30	98.70	98.80

most recently LGONBP method with 4.97% improvement. Moreover, the $CCPs$ of 2D-LCoLBP is 51.3% on FMD, which is a significant improvement. The main reason is that the texture images in FMD database are close to the real scene. Thus, the classification performance of traditional methods on this database is generally poor. 2D-LCoLBP outperforms other comparative methods because its learning-based strategy delivers powerful discriminative capability and achieves the scale invariance.

2) *The Noise Robustness*: Fig. 12 (a1)-(a5) show the classification performance by the proposed 2D-LCoLBP and comparative methods under different levels of Gaussian noise on five texture databases. It can be observed that the $CCPs$ of all methods show a decreasing trend as the noise level increases. The proposed 2D-LCoLBP is more robust to noise than that of LBP^{riu2}, MCLBP, and 2DLBP at all noise levels on five texture databases. Moreover, for Brodatz, 2D-LCoLBP outperforms the state-of-the-art methods under different noise levels, while LGONBP works much better than 2D-LCoLBP when $var \geq 0.02$ for KTH-TIPS and CUREt. Besides, 2D-LCoLBP

and MRELBP perform competitively on FMD. The noise robustness of MRELBP is mainly due to the use of the median filter. These results show that 2D-LCoLBP can also perform well in complex databases such as FMD with large intra-class variations, despite the corruption of different noise levels.

3) *The Robustness of JPEG Compression*: Fig. 12 (b1)-(b5) illustrates the classification performance of all comparative methods under different QF on five texture databases. From these figures, we found that the $CCPs$ of all methods shows a decreasing trend as the QF decreases. The $CCPs$ of 2D-LCoLBP can maintain the classification performance even when QF is down to 30. For KTH-TIPS, Brodatz, FMD, and UMD, there are significant performance gaps between 2D-LCoLBP and other comparative methods. Especially on CUREt, 2D-LCoLBP and LETRIST perform competitively. The main reason is that the smaller QF means the higher degree of information loss, which leads to smaller differences between the materials of each image in CUREt, such as light reflections, shadows, and etc.

F. Evaluation on Object, Face and Food Recognition

In this subsection, we evaluated the image recognition performance of the proposed 2D-LCoLBP in terms of object, face and food recognition.

1) *Object Recognition*: In this experiment, two object databases are employed for evaluation. The experimental results are shown in Table VII. For Coil-100, the recognition performance by 2D-LCoLBP is significantly better than that of other methods, especially for LETRIST and LGONBP, which shows 26.41% and 31.36% improvements in $CCPs$ respectively. The main reason is that both Coil-100 and corel1k have prominent shape areas, showing that 2D-LCoLBP can capture the texture information well and emphasize shape information.

2) *Face Recognition*: Table VII shows the experimental results for face recognition by different methods on AR Face and Extended Yale B. It can be clearly seen that for AR Face, the $CCPs$ of 2D-LCoLBP is 0.77% higher than that of CA-LBFL. The main reason is that this database has noticeable

TABLE VII

THE *CCPs* OF THE PROPOSED 2D-LCoLBP AND THE STATE-OF-THE-ART METHODS ON OBJECT, FACE, AND FOOD DATABASES

Methods	Feature Dimension	Object		Face		Food
		Coil-100	corel1k	AR Face	Extended Yale B	Raw FooT
LBP ^{riu2}	54	96.86	77.00	67.54	74.07	78.10
MRELBP	1568	99.46	82.60	78.77	75.21	93.61
LETRIST	413	73.06	70.20	69.15	50.79	93.55
LGONBP	1404	68.11	66.10	64.46	46.66	92.74
CBFD	500	94.61	67.20	79.69	87.48	94.24
SSLBP	2400	96.47	74.00	74.11	56.88	90.57
CA-LBFL	500	93.11	69.80	80.23	86.98	94.82
CoALBP	3072	96.10	72.10	78.92	68.32	90.03
MCLBP	886	98.78	71.80	72.46	53.21	93.93
2DLBP	1100	98.90	74.39	79.56	81.15	94.81
2D-LCoLBP	440	99.47	83.20	81.00	87.16	95.11

illumination changes, and the face is partially occluded. These results show that 2D-LCoLBP is not sensitive to the lack of information in the occluded part of the face, and the occluded part does not affect the overall distribution of features in 2D-LCoLBP. Moreover, 2D-LCoLBP and CBFD perform competitively on Extended Yale B that has heavy illumination changes. Thus, 2D-LCoLBP has shown a desirable performance in face recognition applications.

3) *Food Recognition*: In the previous subsections, we have demonstrated the effectiveness of 2D-LCoLBP on classic databases. To further prove that our method have more practical applicability, we conducted the experiment on RawFooT that is a relatively new database. The experimental results are shown in Table VII. We observed the *CCPs* of 2D-LCoLBP is significantly higher than other comparative methods.

V. CONCLUSION

In this paper, a learning 2D co-occurrence LBP named 2D-LCoLBP is proposed for image recognition. First of all, the proposed 2D-LCoLBP considers the description of the image in the multi-scale space and the contextual information of LBP pattern pairs in the multi-neighborhood, which achieves the scale invariance and enhances the feature discrimination. Then, a feature learning strategy balances the relationship between the feature discrimination and feature dimension. The low-dimensional 2D-LCoLBP is obtained across different scales in 2D-MLBP, which characterizes the most stable local structures. Finally, experimental results demonstrate an advanced performance of 2D-LCoLBP under noise-free, Gaussian noise and JPEG compression conditions. It can be verified in experimental results that the proposed 2D-LCoLBP outperforms the state-of-the-art LBP-based descriptors on four image recognition tasks, i.e., texture, object, face and food recognition. Combining the proposed method with a deep learning framework is the central issue in our future work to derive an optimized descriptor with high feature discrimination, robustness, and low-dimensionality.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valued comments and constructive suggestions that significantly improved the quality of this paper.

REFERENCES

- [1] C. Ding, J. Choi, D. Tao, and L. S. Davis, "Multi-directional multi-level dual-cross patterns for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 518–531, Jul. 2015.
- [2] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.
- [3] W. Deng, J. Hu, and J. Guo, "Compressive binary patterns: Designing a robust binary face descriptor with random-field eigenfilters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 758–767, Mar. 2018.
- [4] W. Louis and K. N. Plataniotis, "Co-occurrence of local binary patterns features for frontal face detection in surveillance applications," *EURASIP J. Image Video Process.*, vol. 2011, pp. 1–17, Dec. 2011.
- [5] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 533–544, Feb. 2009.
- [6] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (COIL-100)," Columbia Univ., New York, NY, USA, Tech. Rep. CUCS-006-96, 1996.
- [7] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.
- [8] Z. Guo and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jan. 2010.
- [9] S. Wang, Q. Wu, X. He, J. Yang, and Y. Wang, "Local *N*-ary pattern and its extension for texture classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1495–1506, Feb. 2015.
- [10] L. Liu, S. Lao, P. W. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Median robust extended local binary pattern for texture classification," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1368–1381, Mar. 2016.
- [11] T. Song, H. Li, F. Meng, Q. Wu, and J. Cai, "LETRIST: Locally encoded transform feature histogram for rotation-invariant texture classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 7, pp. 1565–1579, Jul. 2017.
- [12] T. Song, J. Feng, L. Luo, C. Gao, and H. Li, "Robust texture description using local grouped order pattern and non-local binary pattern," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 189–202, Jan. 2020.
- [13] S. Liao, M. W. K. Law, and A. C. S. Chung, "Dominant local binary patterns for texture classification," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 1107–1118, May 2009.
- [14] B. Xiao, K. Wang, X. Bi, W. Li, and J. Han, "2D-LBP: An enhanced local binary feature for texture image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2796–2808, Sep. 2019.
- [15] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [16] Z. Guo, X. Wang, J. Zhou, and J. You, "Robust texture image representation by scale selective local binary patterns," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 687–699, Feb. 2015.
- [17] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
- [18] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
- [19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [20] 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, 2008.
- [21] J. Lu, V. E. Liong, X. Zhou, and J. Zhou, "Learning compact binary face descriptor for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 2041–2056, Oct. 2015.

- [22] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Learning rotation-invariant local binary descriptor," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3636–3651, Aug. 2017.
- [23] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Context-aware local binary feature learning for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1139–1153, May 2017.
- [24] R. Nosaka, Y. Ohkawa, and K. Fukui, "Feature extraction based on co-occurrence of adjacent local binary patterns," in *Proc. Pacific-Rim Symp. Image Video Technol.* Berlin, Germany: Springer, 2011, pp. 82–91.
- [25] R. Nosaka, C. H. Suryanto, and K. Fukui, "Rotation invariant co-occurrence among adjacent LBPs," in *Proc. Asian Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 15–25.
- [26] X. Qi, R. Xiao, C.-C. Li, Y. Qiao, J. Guo, and X. Tang, "Pairwise rotation invariant co-occurrence local binary pattern," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2199–2213, Nov. 2014.
- [27] X. Qi, L. Shen, G. Zhao, Q. Li, and M. Pietikäinen, "Globally rotation invariant multi-scale co-occurrence local binary pattern," *Image Vis. Comput.*, vol. 43, pp. 16–26, Nov. 2015.
- [28] A. P. Witkin, *Scale-Space Filtering*. Amsterdam, The Netherlands: Elsevier, 1987, pp. 329–332.
- [29] J. J. Koenderink, "The structure of images," *Biological*, vol. 50, no. 5, pp. 363–370, 1984.
- [30] K. Mikolajczyk, "Detection of local features invariant to affines transformations," Ph.D. dissertation, Inst. Nat. Polytechnique de Grenoble-INPG, Grenoble, France, 2002.
- [31] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh, "On the significance of real-world conditions for material classification," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2004, pp. 253–266.
- [32] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, vol. 66. New York, NY, USA: Dover, 1966.
- [33] L. Sharan, R. Rosenholtz, and E. Adelson, "Material perception: What can you see in a brief glance?" *J. Vis.*, vol. 9, no. 8, p. 784, 2009.
- [34] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real-world surfaces," *ACM Trans. Graph.*, vol. 18, no. 1, pp. 1–34, Jan. 1999.
- [35] Y. Xu, H. Ji, and C. Fermüller, "Viewpoint invariant texture description using fractal analysis," *Int. J. Comput. Vis.*, vol. 83, no. 1, pp. 85–100, 2009.
- [36] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 9, pp. 947–963, Sep. 2001.
- [37] A. M. Martínez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.
- [38] K.-C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.
- [39] C. Cusano, P. Napoletano, and R. Schettini, "Evaluating color texture descriptors under large variations of controlled lighting conditions," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 33, no. 1, pp. 17–30, 2016.
- [40] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2032–2047, Nov. 2008.
- [41] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.



Xiuli Bi received the B.Sc. and M.Sc. degrees from Shanxi Normal University, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from the University of Macau, in 2017. She is currently an Associate Professor with the College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, China. Her research interests include image processing and multimedia security and forensic.



Yuan Yuan received the B.S. degree in internet of things engineering from Chongqing Technology and Business University, China, in 2019. She is currently pursuing the M.S. degree in computer science and technology with Chongqing University of Posts and Telecommunications, China. Her research interest is image processing.



Bin Xiao received the B.S. and M.S. degrees in electrical engineering from Shanxi Normal University, Xi'an, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from Xidian University, Xi'an. He is currently a Professor with Chongqing University of Posts and Telecommunications, Chongqing, China. His research interests include image processing and pattern recognition.



Weisheng Li (Member, IEEE) received the B.S. degree from the School of Electronics and Mechanical Engineering, Xidian University, Xi'an, China, in July 1997, and the M.S. and Ph.D. degrees from the School of Electronics and Mechanical Engineering and the School of Computer Science and Technology, Xidian University, in July 2000 and July 2004, respectively. He is currently a Professor of Chongqing University of Posts and Telecommunications. His research focuses on intelligent information processing and pattern recognition.



Xinbo Gao (Senior Member, IEEE) received the B.Eng., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively. From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Postdoctoral Research Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. He was with the School of Electronic Engineering, Xidian University, Xi'an, from 2001 to 2020. Since 2020, he has been the President of Chongqing University of Posts and Telecommunications, Chongqing, China. His current research interests include multimedia analysis, computer vision, pattern recognition, machine learning, and wireless communications. He has published five books and approximately 200 technical articles in refereed journals and proceedings, including the *IEEE TRANSACTIONS ON IMAGE PROCESSING*, the *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, the *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, the *International Journal of Computer Vision*, and *Pattern Recognition* in the above areas. He is a fellow of the Institution of Engineering and Technology. He has served as the General Chair/Co-Chair, the Program Committee Chair/Co-Chair, or a PC Member for approximately 30 major international conferences. He is on the Editorial Boards of several journals, including *Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier).