Pairwise Rotation Invariant Co-Occurrence Local Binary Pattern

Xianbiao Qi, Rong Xiao, Chun-Guang Li, *Member, IEEE*, Yu Qiao, *Senior Member, IEEE*, Jun Guo, and Xiaoou Tang, *Fellow, IEEE*

Abstract—Designing effective features is a fundamental problem in computer vision. However, it is usually difficult to achieve a great tradeoff between discriminative power and robustness. Previous works shown that spatial co-occurrence can boost the discriminative power of features. However the current existing co-occurrence features are taking few considerations to the robustness and hence suffering from sensitivity to geometric and photometric variations. In this work, we study the Transform Invariance (TI) of co-occurrence features. Concretely we formally introduce a Pairwise Transform Invariance (PTI) principle, and then propose a novel *Pairwise Rotation Invariant Co-occurrence* Local Binary Pattern (PRICoLBP) feature, and further extend it to incorporate multi-scale, multi-orientation, and multi-channel information. Different from other LBP variants, PRICoLBP can not only capture the spatial context co-occurrence information effectively, but also possess rotation invariance. We evaluate PRICoLBP comprehensively on nine benchmark data sets from five different perspectives, e.g., encoding strategy, rotation invariance, the number of templates, speed, and discriminative power compared to other LBP variants. Furthermore we apply PRICoLBP to six different but related applications—texture, material, flower, leaf, food, and scene classification, and demonstrate that PRICoLBP is efficient, effective, and of a well-balanced tradeoff between the discriminative power and robustness.

Index Terms—Co-occurrence LBPs, rotation invariance, texture classification, material recognition, flower recognition, leaf recognition, food recognition, scene recognition

1 Introduction

Designing effective features is a fundamental issue in computer vision. It plays a significant role in a wide range of applications that includes static and dynamic texture classification [1], [2], [3], [4], [5], object and scene recognition [6], [7], [8], [9], [10], face detection and recognition [11], image retrieval [12], [13], stereo correspondence, 3D reconstruction and many more. It is well-known that real-world applications always suffer from large intra-class variations due to the differences in pose, illumination, and appearance. It therefore becomes crucial to design discriminative and robust features. It is commonly accepted that designing effective features has to take a great tradeoff between the discriminativeness and robustness.

Previous works have shown in texture classification [14], object classification [8], [9], [15], [16], [17], and image retrieval [12], [13] that the spatial co-occurrence among features could increase the discriminative power of features. The rationales behind this claim are that:

- The spatial co-occurrence of two features captures a strong correlation between them and hence provides more information than their individual occurrence.
- The spatial co-occurrence feature has larger supporting regions than a single feature and hence can depict more subtle and complex structures in an image. Contrarily, traditional single feature describes a smaller supporting region individually and ignores the spatial relationship among features.
- The spatial co-occurrence of features can provide higher order statistical information than their individual occurrence without spatial coherence. It has been shown in psychological study that higher order information is important for performing categorization task in primate visual cortex [18].

However, the real world vision applications often suffer from geometric and photometric variations. Unfortunately the current existing works on co-occurrence features mainly focus on improving discriminativeness, e.g., [8], [9], [12], [14], [15], [16], [21], [22]. To the best of our knowledge, very few work in literature investigates the transformation invariance in spatial co-occurrence features.

In this paper, we address the issue of transform invariance in spatial co-occurrence features. The main contributions are highlighted as follows:

- We introduce a formal definition of the Pairwise Transform Invariance (PTI) for co-occurrence features and present a PTI principle which can be used to construct co-occurrence features of PTI.
- We propose a novel *pairwise rotation invariant co-occurrence local binary pattern* (PRICoLBP) feature, in which we encode two LBP features collaboratively

- *X. Qi, C.-G. Li, and J. Guo are with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, P.R. China. E-mail: qixianbiao@gmail.com, {lichunguang, guojun}@bupt.edu.cn.*
- *R. Xiao is with the Microsoft Corporation, 1 Microsoft Way, Redmond, WA 98052. E-mail: rxiao@microsoft.com.*
- *Y. Qiao and X. Tang are with the Shenzhen Key Laboratory for Computer Vision and Pattern Recognition, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, P.R. China. E-mail: yu.qiao@siat.ac.cn, xtang@ie.cuhk.edu.hk.*

Manuscript received 29 Aug. 2013; revised 15 Mar. 2014; accepted 1 Apr. 2014. Date of publication 10 Apr. 2014; date of current version 9 Oct. 2014. Recommended for acceptance by T. Tuytelaars.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2014.2316826

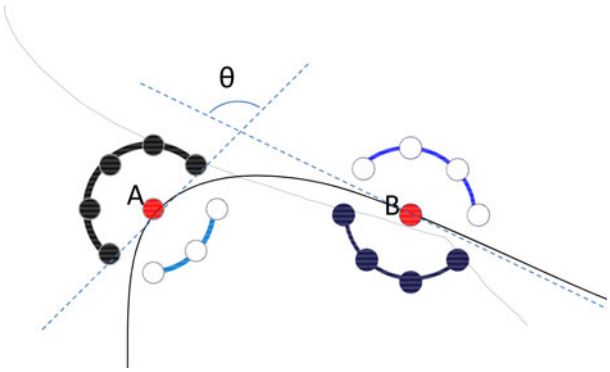


Fig. 1. Illustration of co-occurrence local binary pattern. Complex structure can be encoded into co-occurrence LBP, e.g., the relative angle θ, which captures the local curvature information and is rotation invariant.

while preserving the relative orientation angle between them as shown in Fig. 1.

- We further extend RPICoLBP to incorporate multi-scale, multi-orientation, and multi-channel color information.
- We evaluate PRICoLBP comprehensively on nine data sets from five different perspectives, including encoding strategy, rotation invariance, the number of templates, speed, and discriminative power compared with other LBP variants.
- We apply PRICoLBP to six different but related applications—texture, material, flower, leaf, food, and scene classification, and demonstrate that PRICoLBP is efficient, effective, and of a well-balanced tradeoff between the discriminativeness and robustness.

The remainder of this paper is organized as follows. We review the related works in Section 2. We introduce a PTI concept, present a PTI principle, and propose PRICoLBP feature in Section 3. In Section 4, we provide a comprehensive experimental evaluations of the PRICoLBP feature, and in Section 5, we apply the PRICoLBP feature to six real world applications and compare it with the state-of-the-art methods. We conclude this paper with a discussion in Section 6.

2 Related Works

2.1 Local Binary Pattern

The local binary pattern operator was firstly proposed by Ojala et al. [1] as a gray-scale invariant texture descriptor. For a pixel A in an image, its LBP code is computed by thresholding its circularly symmetric n neighbors in a circle of radius r with the pixel value of the central point and arranging the results as a binary string. For clarity, we denote the LBP of pixel A as $\mathrm{LBP}_{n,r}(A)$, which is defined as follows:

$$\mathrm{LBP}_{n,r}(A) = \sum_{i=0}^{n-1} s(g_i - g_c)2^i,$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0, \end{cases}$$

where A is a two-dimensional coordinates of a point in an image, g_c is the pixel value of point A, and g_i is the pixel value of point A's ith neighbor. LBP is gray-scale invariant because the function $s(g_i - g_c)$ is invariant to monotonic change of illumination.

Ojala et al. pointed out that the patterns with very few spatial transitions describe the fundamental image microstructures, and termed them as "uniform patterns". Precisely, the uniform patterns can be defined as those LBPs that $\mathcal{U}(\mathrm{LBP}_{n,r}(A)) \leq 2$, where $\mathcal{U}(\cdot)$ is a uniformity measure that is calculated as follows:

$$\mathcal{U}(\mathrm{LBP}_{n,r}(A)) = \sum_{i=1}^{n} |s(g_i - g_c) - s(g_{i-1} - g_c)|,$$

where g_n is equivalent to g_0. For example, "11110000" and "11000011" are uniform patterns, and "11110100" and "10100100" are non-uniform patterns.

To enhance the robustness to image rotation, rotation invariant LBP (LBP^{ri}) and rotation invariant uniform LBP (LBP^{ru}) are also introduced. The LBP^{ri} is defined as

$$\mathrm{LBP}^{ri}_{n,r}(\mathrm{A}) = \min\{\mathrm{ROR}(\mathrm{LBP}_{n,r}(\mathrm{A}), i) \mid \forall i \in [0, n-1]\},$$

where $\mathrm{ROR}(x, i)$ performs a circular bit-wise right shift for i times on n-bit number x. The $\mathrm{LBP}^{ru}_{n,r}(A)$ is defined as

$$\mathrm{LBP}^{ru}_{n,r}(A) = \begin{cases} \sum_{i=0}^{n-1} s(g_i - g_c), & \mathcal{U}(\mathrm{LBP}(A)) \leq 2 \\ n+1, & \text{otherwise}, \end{cases}$$

For simplicity, we limit our discussions to $\mathrm{LBP}_{8,1}$ and use LBP by default thereafter[1]. Note that our method can be easily extended to any general LBP configurations.

2.2 LBP Variants

After the pioneering LBP work [1], Tan and Triggs [21] proposed a local ternary pattern (LTP) for robust face recognition. Guo et al. [22] proposed a weighted LBP method, which is termed as LBPV. In LBPV, the variance that characterizes the local contrast information was used to weight the one-dimensional LBP histogram. Vu and Caplier proposed oriented edge patterns (POEM) [23] and patterns of dominant orientations (PDO) [24] for face recognition. Guo et al. [25] proposed a completed LBP (CLBP) feature to incorporate the sign and magnitude information into the final descriptor. Ahonen et al. [26] proposed an effective LBP Fourier histogram (LBPHF) to achieve rotation invariance. Zhao et al. [27] proposed a LBPHF_S_M, an extension of LBPHF, by combining the sign and magnitude information. However, all of the above-mentioned LBP variants are limited to optimize the encoding scheme on a single LBP point without capturing spatial co-occurrence information among the features.

2.3 Co-Occurrence Features

According to whether the spatial neighboring relationship among features is used in computing the co-occurrence

1. The uniform $\mathrm{LBP}_{8,1}$ includes 58 uniform patterns and one additional pattern (which is used to represent those non-uniform patterns). The rotation invariant uniform $\mathrm{LBP}_{8,1}$ includes 10 patterns.

statistics, existing co-occurrence features can be sorted into two categories: global co-occurrence features and local co-occurrence features. In [15], Yuan et al. proposed to mine co-occurrence statistics of SIFT words for visual recognition. Similarly, Rasiwasia and Vasconcelos [16] proposed to calculate the co-occurrence statistics on the whole image without considering the local spatial relationship among features. These works fall into the category of global co-occurrence features. In [8], [9], [12], [14], [17], [19], [20], the spatial co-occurrence are computed within locally adjacent neighbors instead of on the whole image. In [14], Haralick et al. proposed a Gray-Level Co-occurrence Matrices (GLCM) feature, in which the gray value co-occurrence is computed at some given offsets. In [8], Chang and Krumm explored color co-occurrence by representing co-occurrence statistics of color pixels in a fixed spatial distance. Unfortunately, GLCM is sensitive to gray-scale and rotation variation. Illumination variation leads to great changes in the co-occurrence matrix. The color co-occurrence [8] method is also sensitive to gray-scale and illumination variation. Recently, Ito et al. [9] proposed a co-occurrence of histogram of orientation gradient (CoHOG) method, which adopted the co-occurred gradient orientation under various positional offsets, and Yang et al. [17] proposed several pairwise co-occurrence features for food recognition. However, these features are all sensitive to image rotation. Recently a co-occurrence adjacent LBP (CoALBP) [28] is proposed for texture classification and face reocgnition. However CoALBP is built upon LBP with four neighboring pixels and hence suffers from the lack of discriminativeness. In addition, CoALBP is also sensitive to image rotation.

In this paper, we address the transformation invariance issue in spatial co-occurrence features. To be specific, we give a formal definition of the transformation invariance, introduce the principle for the design of transformation invariant co-occurrence features, and present a simple but effective PRICoLBP feature, which possesses the following merits:

- *Gray-scale invariance*. Monotonic gray-scale variation does not change the PRICoLBP pattern.
- *Rotation invariance*. The rotation invariance is obtained by firstly guaranteeing the correspondence of co-occurrence pair, and then using an effective pairwise rotation invariant encoding strategy.
- *Strong discriminativeness*. PRICoLBP encodes the relative orientation angle between the co-occurrence pair features. The relative orientation angle reflects the local curvature information, which is rotation invariant and discriminative.
- *Leveraging edge/coutour information*. We emphasize the shape regions by using local gradient magnitude to weight the co-occurrence patterns.
- *Computational efficiency*. The PRICoLBP can process about 30 images with image size 200×200 in one second.
- *Flexiblity to incorporate color information*. Color information is effective for many applications. Effectively fusing color information will benefit them.

3 Pairwise Transform Invariance and PRICoLBP

In this section, we start with a formal definition of pairwise transform invariance for co-occurrence features and then describe a novel pairwise rotation invariant co-occurrence LBP feature.

To begin with, let A and B be two points in an image, and $f_1(A)$ and $f_2(B)$ be two local features defined in the neighborhood of point A and B, respectively.[2] We then denote $[f_1(A), f_2(B)]_{co}$ as a co-occurrence feature defined on the neighborhoods of point A and point B by using local feature descriptors $f_1(\cdot)$ and $f_2(\cdot)$. Assume that points A and B are spatially adjacent, and point B can be obtained from point A by a mapping function $\varphi : A \rightarrow B$.

3.1 Pairwise Transform Invariance

Definition 1 (Transform Invariant). *Let $\tau(\cdot)$ as an in-plane transform (e.g., rotation, translation, and illumination change). The feature $f(A)$ is* transform invariant *(TI) with respect to (w.r.t.) a transform $\tau(\cdot)$ if and only if*

$$f(\tau(A)) = f(A), \tag{1}$$

in which A is an arbitrary point in an image, $f(\cdot)$ is any feature descriptor (e.g., LBP [1], SIFT [29]), and τ is any in-plane transform.

For example, LBP is transform invariant to monotonic illumination change. In this case, $f(\cdot)$ is LBP descriptor and $\tau(\cdot)$ is the transform induced by illumination change. Similarly we can also introduce the transform invariant property for co-occurrence features.

Definition 2 (Pairwise Transform Invariant). *Let $B = \varphi(A)$ and $[f_1(A), f_2(\varphi(A))]_{co}$ be a spatial co-occurrence features from feature $f_1(\cdot)$ and $f_2(\cdot)$. If*

$$[f_1(\tau(A)), f_2(\tau(\varphi(A)))]_{co} = [f_1(A), f_2(\varphi(A))]_{co}, \tag{2}$$

where $\tau(\cdot)$ is an in-plane transform, then we say the co-occurrence feature $[f_1(A), f_2(\varphi(A))]_{co}$ is pairwise transform invariant *(PTI) w.r.t. $\tau(\cdot)$.*

Based on the definitions above, we are in the position to introduce a *pairwise transform invariant* principle for designing co-occurrence feature.

Proposition 1 (PTI Principle). *Let $[f_1(A), f_2(\varphi(A))]_{co}$ be the co-occurrence features from features $f_1(\cdot)$ and $f_2(\cdot)$. If $f_1(\cdot)$ is transform invariant and $f_2(\varphi(A))$ is transform invariant w.r.t. point A, which means $f_2(\tau(\varphi(A))) = f_2(\varphi(\tau(A)))$, then the co-occurrence feature $[f_1(A), f_2(\varphi(A))]_{co}$ is pairwise transform invariant w.r.t. $\tau(\cdot)$.*

The PTI principle can be served as a theoretical guideline for designing PTI co-occurrence feature. According to the PTI principle, the co-occurrence feature $[f_1(A), f_2(\varphi(A))]_{co}$ of two TI features $f_1(\cdot)$ and $f_2(\cdot)$ is also transformation invariant. For instance, if $f_1(\cdot)$ and $f_2(\cdot)$ are both rotation invariant feature, then PTI feature $[f_1(A), f_2(\varphi(A))]_{co}$ is also rotation invariant.

2. $f_1(.)$ and $f_2(.)$ can be the same or different type of features.

According to the definition of co-occurrence, by using different types of feature $f_1(\cdot)$ and $f_2(\cdot)$, and $\varphi(\cdot)$, we can derive different co-occurrence feature. Here, we use LBP feature (i.e., $f_1(\cdot)$ and $f_2(\cdot)$ are both LBP features) and derive the following two types of co-occurrence features: UUCoLBP and RUCoLBP, in which

- UUCoLBP is defined as

$$\text{UUCoLBP}(A,B) = [\text{LBP}^u(A), \text{LBP}^u(B)]_{co}, \quad (3)$$

  where $\text{LBP}^u(\cdot)$ is the uniform LBP;
- RUCoLBP is defined as

$$\text{RUCoLBP}(A,B) = [\text{LBP}^{ru}(A), \text{LBP}^{ru}(B)]_{co}, \quad (4)$$

  where $\text{LBP}^{ru}(\cdot)$ is rotation invariant uniform LBP.

Note that the dimension of UUCoLBP is $59 \times 59 = 3{,}481$, and the dimension of RUCoLBP is $10 \times 10 = 100$. We can verify that RUCoLBP is pairwise rotation invariant but UUCoLBP is not. The two co-occurrence features are complementary. While UUCoLBP has richer patterns but is sensitive to rotation, RUCoLBP is robust to rotation but lacks of discriminativeness due to its limited patterns.

Co-occurrence features can capture more complex structures. However, the naive co-occurrence strategy, which simply encodes the two features individually, will lead to heavily information loss. To illustrate this point, let us see an example in Fig. 1 that, the co-occurrence feature can reveal the local curvature information, e.g., the relative orientation angle θ, which is discriminative and rotation invariant. If we encode the two co-occurrence features independently, the local curvature information will be ignored.

3.2 PRICoLBP

As an application of the PTI principle aforementioned, we propose an novel co-occurrence LBP encoding scheme, called *pairwise rotation invariant co-occurrence* LBP, in which the discriminative curvature information and pairwise invariance w.r.t. rotation are preserved properly.

Denote $\text{LBP}^u(A,i)$ as the uniform LBP of point A by using ith index as the start point of the binary sequence. To be concrete, we define PRICoLBP as follows:

$$\text{PRICoLBP}(A,B) = [\text{LBP}^{ru}(A), \text{LBP}^u(B, i(A))]_{co}, \quad (5)$$

where $i(A) \in \{0, n-1\}$ is an index, which can be determined by point A. In other words, for two points A and B, PRICoLBP *encode the feature at point A with rotation invariant uniform LBP, and then encodes the feature at point B with uniform LBP with respect to point A*. Precisely, we use point A to determine the start index point $i(A)$ that is used to encode point B, in which $i(A)$ is determined as follows

$$i(A) = \arg\max_{i \in \{0,n-1\}} \{\text{ROR}(\text{LBP}_{n,r}(A),\ i)\}. \quad (6)$$

Different from UUCoLBP and RUCoLBP, PRICoLBP can exactly encode the relative oriental information between two individual features. Consequently PRICoLBP not only possesses richer patterns but preserves the pairwise rotation invariance as well. We give an illustration in Fig. 2 to show that PRICoLBP yields exactly the same co-occurrence feature for the same point pair regardless of the image rotation.

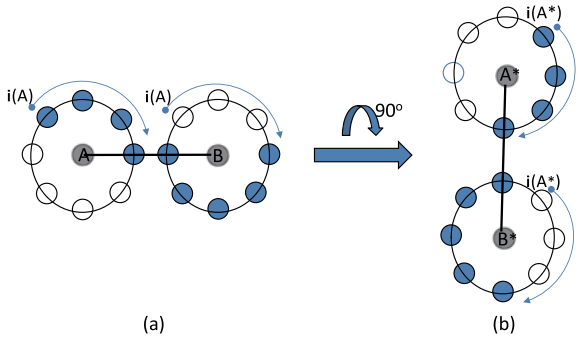


Fig. 2. An illustration of pairwise rotation invariance. For the panel (a), we first determine orientation $i(A)$ of the reference point A. We then compute the uniform pattern of B according to $i(A)$. Finally, we get co-patterns $[(11110000)^{ru}, (00011111)^u]$. If the image is rotated, i.e., the pair in the right panel, we can also yield the same co-occurrence features.

For implementation, we introduce a 2D histogram of co-patterns for image description. We densely sample each point in the image. For each point, we calculate its corresponding point according to pre-defined a and b, and compute its PRICoLBP pattern. The 2D co-occurrence histogram can be obtained by counting the number of points belonging to each PRICoLBP pattern. In practice, we use the gradient magnitudes of two points to weight the co-pattern.

In order to achieve global rotation invariance, the target point B needs to be uniquely identified from the point A in spite of image rotation. Let $g(A)$ and $n(A)$ be the unit gradient direction vector and the unit normal direction vector at point A. Then we can identify the point B from A as follows:

$$B \doteq \varphi(A) = a * g(A) + b * n(A) + A, \quad (7)$$

where a and b are coefficients, and $g(A)$ and $n(A)$ are the gradient and normal directions of point A. With abuse of notation, let A be the coordinates of point A, then we can see that $\varphi(A)$ is effectively as new coordinates defined with the local coordinates system in which $g(A)$ as the x-axis and $n(A)$ as the y-axis. With different choices of a and b, it will lead to different functions $\varphi(\cdot)$. For example, $a = 2$ and $b = 0$ mean that the point B is the point with 2 pixels distance away from A along the gradient direction.

Remarks: When rotation variation is not large, we define the function $\varphi(.)$ along the horizontal and vertical unit direction. In this case, the constructed PRICoLBP is no longer rotation invariant. For clarity, we define the non-rotation invariant PRICoLBP as PRICoLBP_0, and denote the rotation invariant PRICoLBP, which is defined along with unit gradient and normal direction, as PRICoLBP_g. In Section 4, we will compare the performance of PRICoLBP_0 and PRICoLBP_g.

3.3 Extension to Multi-Scale, Multi-Orientation, and Multi-Color Channels

We extend PRICoLBP to multi-scale and multi-orientation (MSMO PRICoLBP), in order to describe local image structures in multi-scale and multi-orientation way. For each reference point A, as shown in Fig. 3, we extract PRICoLBP patterns from point pairs (A, B_i) in multi-scale

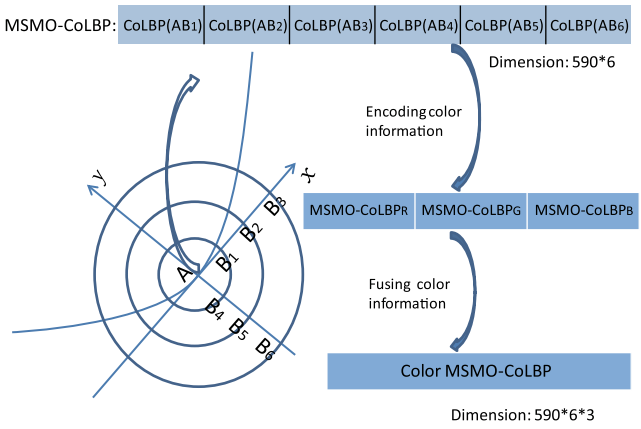


Fig. 3. An illustration for the color MSMO PRICoLBP feature of three scales and two orientations. PRICoLBP features that extracted on points pairs from all the combination of scales and orientations are concatenated to form a MSMO PRICoLBP feature. And then the MSMO PRICoLBP features from different color channels are concatenated to form the color MSMO PRICoLBP feature.

and multi-orientation. In order to achieve rotation invariance, we assume that $B_i = \varphi_i(A)$ where the function $\varphi_i(.)$ is defined with the local coordinates system using unit gradient direction as x-axis and unit normal direction as y-axis. In Fig. 3, we extract six co-occurrence patterns from three scales and two orientations, and thus yield to a $2 \times 3 \times 590 = 3{,}540$-dimensional feature vector.

Color information is crucial for many applications, such as flower recognition [30], [31], [32], food recognition [17]. In these cases, color provides a great amount of complementary information with texture and shape. To incorporate color information into PRICoLBP feature, we extract PRICoLBP feature from three color channels and concatenate them together. For example, in Fig. 3, when incorporating color information into MSMO PRICoLBP feature, the dimension of the obtained color MSMO PRICoLBP is $3540 \times 3 = 10{,}620$.[3]

3.4 Approximate Calculation

This section describes an approach to speed up the extraction of MSMO PRICoLBP. When multi-scale and multi-orientation strategy is applied, the direct interpolation-based method requires a heavy computational cost to calculate LBP values of target points, since we need to determine multiple target points and construct a points set $\{B_i\}$. Note that the positions of target points are usually not exactly on grid, we have to compute the pixel value of points B_i and its neighbors by many interpolation operations, according to Eq. (7). Meanwhile the LBP values cannot be pre-computed or reused, due to the fact that points $\{B_i\}$ are not exactly on the grid. As a result, in the multi-scale and multi-orientation extension of PRICoLBP, the computational cost is proportional to the number of scales and orientations. Fortunately, we can compute points $\{B_i\}$ approximately. To do so, we resize the image to its N times by bilinear interpolation, and then map the position of each point B_i at the original image to its nearest grid position at the enlarged images. In this way, we can pre-compute each LBP value in the resized image and hence reduce the computational cost significantly. The main computational cost in the approximation method comes from the computation of LBP values in resized images. In practice, we observed that $N = 3$ yields good tradeoff between the speed and accuracy.

In the approximation method, we pre-compute and reuse the B_is' LBP values. Although the mapped positions at resized image *not exactly coincide* with the positions at original image, the position offsets is small. In theory, the maximum distance of the position offsets between the original position and approximated position is $\frac{\sqrt{2}}{6} \approx 0.226$ if we set N to 3. In most cases, the approximation method does not change the points $\{B_i\}$'s LBP values, and the variation in gradient magnitude caused by the approximation is also small. We conducted experiments to compare LBP values between the non-grid points in original images and their approximated points in resized images, and found that more than 99 percent of them have the same LBP features.

Algorithm 1 Computing MSMO PRICoLBP

Input: Input image I, the number of templates T, and templates setting $\{(a_t, b_t)\}_{t=1}^{T}$;
Output: PRICoLBP histogram feature $\mathcal{H}$

Initiate a 3D co-occurrence histogram $\mathcal{H}$ of size $10 \times 59 \times T$ with zeros;
Smooth the image with a Gaussian filter.
Compute the gradient orientation matrix G and gradient magnitude matrix M from gradient images I_x and I_y;
Obtain image $I^{(3)}$ by enlarging image I three times;
Compute the gradient orientation matrix G_3 and magnitude matrix M_3, and normal orientation matrix N_3 from gradient images $I_x^{(3)}$ and $I_y^{(3)}$;
for each $A \in I$ **do**
  for each $t \in [1, T]$ **do**
    Obtain the gradient vector $G(A)$ of point A;
    Compute $\mathrm{RU}(A)$ of point A;
    Compute the position coordinates B_t according to

$$B^{(t)} = A + a_t \times G(A) + b_t \times N(A)$$

    if $B^{(t)} \in I$ **then**
      Project the position $B^{(t)}$ to $B_3^{(t)}$;
      Determine the index $i(A)$ according to Eq. 6;
      Compute $\mathrm{U}(B_3^{(t)})$ of point B with index $i(A)$;
      Construct PRICoLBP feature by

$$\mathcal{H}(\mathrm{RU}(A), \mathrm{U}(B_3^{(t)}), t) \leftarrow \mathcal{H}(\mathrm{RU}(A), \mathrm{U}(B_3^{(t)}), t) + M(A) + M_3(B_3^{(t)}).$$

    end if
  end for
end for
Return $\mathcal{H}$

For clarity, we summarize the approximation method for calculating PRICoLBP in Algorithm 1, in which $\mathrm{RU}(A)$

3. To reduce the feature dimension, we apply a standard dimensionality reduction technology, principal component analysis (PCA). However we also observed that the performance was decreased in some data sets when using PCA to reduce the dimensionality. This is caused by the fact that a part of discriminative information is removed by PCA. Because PCA is used to find the projection in the sense to best represent the data rather than to find the projection with the best discriminative power.

TABLE 1
Summary of Six Applications and Nine Databases Used in Our Experiments

Applications	Texture			Material	Flower	Leaf	Food	Scene	
Databases	Brodatz [33]	CUReT [34]	KTH-TIPS [35]	FMD [36]	Flower 102 [30]	Swed_Leaf [37]	PFID [38]	Scene-15 [39]	MIT Indoor 67 [71]
Classes	111	61	10	10	102	15	61	15	67
Total Samples	999	5,612	810	1,000	8,189	1,125	1,098	4,485	5,620
Train Samples	3	46	40	50	20	25	12	100	20

stands for the RU-LBP of point A and the $\mathrm{U}(B_3^{(t)})$ is for the uniform LBP of point B. Here we refer a scale or an orientation as a template, e.g., if we use three scales and two orientations, then the number of templates T is 6.

4 A Comprehensive Evaluation of PRICoLBP

In this section, we describe briefly the implementation details of the proposed PRICoLBP feature, and then present a comprehensive evaluation of its performance on nine benchmark data sets from five different perspectives: encoding strategy, rotation invariance, the number of templates, speed, and discriminative power compared with LBP variants.

For clarity, we summarize the basic information of the nine benchmark data sets in Table 1. A more detailed descriptions will be presented in Section 5.

4.1 Implementation Details

PRICoLBP settings. For some applications, e.g., Oxford Flower 102, PFID food, and MIT-Indoor 67, color information is important. In these data sets, we extracted PRICoLBP features from three color channels and concatenated them together. For other data sets, we used gray-scale images. Since that $\mathrm{LBP}_{8,2}$ and $\mathrm{LBP}_{8,1}$ shown similar performance, we used $\mathrm{LBP}_{8,2}$ on all data sets, except for Scene-15 and MIT-Indoor 67. On these two data sets, we observed that the combination of $\mathrm{LBP}_{8,1}$ and $\mathrm{LBP}_{8,2}$ could significantly outperform the results by only single feature, and thus we used two scales.

Classification method. In image classification, SVM classifier with χ^2 kernel is widely used and has been demonstrated its effectiveness [5]. Recently, Vedaldi et al. [40] proposed an efficient additive kernel approximation (AKA) method, which enables the fast training and test possible for nonlinear kernel and is scalable to data size. In some data sets, we use PCA to reduce the feature dimensionality and follows is a SVM with RBF kernel. In SVM, we use *one-versus-the-rest* strategy.[4]

4.2 Encoding Strategies Comparison

In this section, we evaluate three different encoding strategies—RUCoLBP, UUCoLBP, and PRICoLBP—which are proposed to encode co-occurrence LBP features. As mentioned in Section 3, PRICoLBP and RUCoLBP are PTI features, but UUCoLBP is not. We experimentally compare the effectiveness of the three encoding strategies. We use χ^2 kernel SVM. For data sets Brodatz, KTH-TIPS, and FMD, we use three, 40, and 50 samples respectively for training, and the rest for testing. Experimental results are listed in Table 2.

Table 2 shows that:

- The three features perform well on all data sets. This implies that co-occurrence LBP feature is effective.
- PRICoLBP outperforms RUCoLBP and UUCoLBP, e.g., PRICoLBP yields about 7 percent relative improvement on FMD data set. This suggests that capturing the relative orientation angle as shown in Fig. 1 is effective and can indeed enhance the discriminativeness of the descriptor.

The reason why UUCoLBP works well on the Brodatz and KTH-TIPS is that these two data sets do not have obvious rotation variation. Compared to RUCoLBP, while UUCoLBP has richer patterns, UUCoLBP is sensitive to rotation variations. Therefore UUCoLBP yields better performance than RUCoLBP on data sets without significant rotation variations. On FMD data set, RUCoLBP outperforms UUCoLBP, due to the exist of natural rotation variations. RUCoLBP gains a great robustness to rotation variation but loses the discriminative power of the descriptor. Contrarily, UUCoLBP owns great descriptive power but suffers from sensitivity to rotation variations. The encoding strategy in PRICoLBP possesses not only the rotation invariance but the discriminativeness as well.

4.3 Rotation Invariance

In this section, we validate the rotation invariance for three different encoding strategies: $\mathrm{PRICoLBP}_g$, $\mathrm{PRICoLBP}_0$, and CoALBP [28]. Experiments are conducted on KTH-TIPS, CUReT, and FMD data sets.

Since KTH-TIPS and CUReT data sets have only small rotation variations in image, we manually add arbitrary rotation variations[5] and produce two new data sets—the rotated versions of KTH-TIPS and CUReT. As FMD data set contains natural rotation variations, we use it directly. Therefore we conducte experiments on five data sets: KTH-TIPS, CUReT, the rotated KTH-TIPS, the rotated CUReT, and FMD. The same training samples and classification methods are used. Experimental results are shown in Table 3.

4. The one-versus-the-rest strategy outperforms the one-versus-one strategy [41].

5. For the KTH-TIPS, we first resize each image to 1.5 times of its original size and then randomly rotate the image with a random angle. We crop out an image of same size as the original image from the center of rotated image. This step guarantees that no black space is cropped out. For CUReT, each image is rotated randomly and then a sub-sampling image of size 141×141 is cropped out from the center of the rotated image. In this way, we create two manually rotated data sets with the same number of images, respectively. Note that the rotating and cropping operations may lead to information loss and rotation variations.

TABLE 2
Comparison of RUCoLBP, UUCoLBP, and PRICoLBP

Dataset	$RUCoLBP_g$	$UUCoLBP_g$	$PRICoLBP_g$
Brodatz(3)	93.9%	95.7%	**96.9%**
KTH-TIPS(40)	96.4%	97.9%	**98.4%**
FMD(50)	52.1%	50.6%	**57.1%**

(The numbers in the bracket after the data set name are the number of training samples.)

From Table 3, we observe that $\mathrm{PRICoLBP}_g$ outperforms $\mathrm{PRICoLBP}_0$ for 5.9 and 5.7 percent on the rotated data sets KTH-TIPS and CUReT. On FMD data set, $\mathrm{PRICoLBP}_g$ also outperforms $\mathrm{PRICoLBP}_0$ 3.7 percent. This result confirms that the rotation invariance in $\mathrm{PRICoLBP}_g$ is effective. On data sets KTH-TIPS and CUReT, $\mathrm{PRICoLBP}_g$ and $\mathrm{PRICoLBP}_0$ both work well and have similar performance.

Our $\mathrm{PRICoLBP}_g$ outperforms CoALBP about 23.6 percent on the rotated data set KTH-TIPS, and 30.1 percent on the rotated data set CUReT. The CoALBP encodes the co-occurrence of two LBP features with only four neighbors and hence weakens its discriminativeness. Even worse, CoALBP is extremely sensitive to rotation variations.

We also notice that the average performance on the rotated data sets is lower than the performance on the original data sets. This is due to the heavy information loss caused by rotating and cropping.

4.4 Templates Setting

In Section 3.3, we extended PRICoLBP to multi-scales and multi-orientations. Here we refer a scale or an orientation as a template. Since images in specific applications may have effective features at specific scales or orientations, PRICoLBP may yield different performance under different templates settings. It is interesting to examine the performance of PRICoLBP under different template setting in different applications. Therefore we conduct experiments on six applications (including nine data sets) to evaluate PRICoLBP with six templates settings. Note in Eq. (7) that (a, b) determines a scale or an orientation. We summarized the templates settings in Table 4 and presented the results in Table 5.

From Table 5, we conclude that: while a single template already works well on all data sets, using PRICoLBP with multiple templates can further improve the performance, especially when strong shape structures (e.g., strong edges or contours) exist in data.

Using more templates in PRICoLBP does significantly improve the performance on data sets with strong shape structures, e.g., Oxford Flower 102, Swedish Leaf, and MIT-Indoor 67. Remarkably, on the Oxford Flower 102, using six templates in PRICoLBP leads to over 9 percent accuracy improvement than the result of with a single template. In addition, on data sets, e.g., FMD, PFID, and Scene-15, using multiple templates in PRICoLBP also bring notable improvement. However, on Brodatz, CUReT, and KTH-TIP, using more templates in PRICoLBP *does not* bring significant improvement. This is because that these data have no strong shape structures.

TABLE 3
Comparison of $\mathrm{PRICoLBP}_0$, $\mathrm{PRICoLBP}_g$, and CoALBP

	Rotation	$\mathrm{PRICoLBP}_0$	$\mathrm{PRICoLBP}_g$	CoALBP
KTH-TIPS(40)	Without	**98.4%**	**98.4%**	97.0%
	With	90.2%	**96.1%**	72.5%
CUReT(46)	Without	**98.5%**	98.4%	98.0%
	With	88.6%	**94.3%**	64.2%
FMD(50)	With	53.4%	**57.1%**	45.2%

TABLE 4
Templates Settings

Number of Template	$\{(a_i, b_i)\}$
1	{(2, 0)}
2_a	{(2, 0), (0, 2)}
2_b	{(2, 0), (3, 0)}
3	{(2, 0), (3, 0), (6, 0)}
6	{(2, 0), (3, 0), (6, 0), (0, 2), (0, 3), (0, 6)}
10	{(2, 0), (3, 0), (6, 0), (0, 2), (0, 3) (0, 6), (2, 2), (4, 4), (-2, 2), (-4, 4)}

4.5 Speed

We evaluate the speed of computing different features: $\mathrm{PRICoLBP}_0$, $\mathrm{PRICoLBP}_g$, standard LBP, multi-scale LBP (MSLBP), and an accelerated Bag-of-Word model.

For LBP, we use $\mathrm{LBP}_{8,1}$. For MSLBP, we use three scales: $\mathrm{LBP}_{8,1}$, $\mathrm{LBP}_{12,2}$ and $\mathrm{LBP}_{16,3}$. For the Bag-of-Words (BOW) model, we use the widely used SIFT feature. Following the standard setup, we apply dense sampling with 5-pixels spaced grid. The extraction of Bag-of-SIFT consists of two steps: feature extraction and feature quantization. We use Vlfeat [42] to evaluate the running time of Bag-of-SIFT method. Vlfeat is currently the fastest bag-of-sift extraction toolbox. Note that Vlfeat uses Streaming SIMD Extensions (SSE) to speed up.

PRICoLBP is very fast, since the major computational cost in PRICoLBP is the computation of gradients and LBP patterns. We implemented PRICoLBP using C++ and evaluated the PRICoLBP on a desktop computer with dual-core 2.8G Intel CPU and 2 GB memory. We run the five methods on six groups of images. Each group have 100 images with the same size. We run the code in each group and report the average running time for each group. The running time is in Table 6.

LBP is extremely fast since it needs only a few interpolations and comparison to calculate a LBP pattern. From Table 6, we observed the speed of $\mathrm{PRICoLBP}_0$ and $\mathrm{PRICoLBP}_g$ is also very rapid. Notice that $\mathrm{PRICoLBP}_0$ takes only about double time of $\mathrm{LBP}(8, 1)$ to process an image, and is much faster than MSLBP.

Compared with $\mathrm{PRICoLBP}_0$, $\mathrm{PRICoLBP}_g$ takes about six times of computational time. As described in Section 4.5, we

TABLE 5
Performance of $\mathrm{PRICoLBP}_g$ with Different Templates Settings

Number of Templates	1	2_a	2_b	3	6	10
Brodatz(3)	96.59	96.86	96.71	96.73	97.08	97.13
CUReT(46)	98.14	98.43	98.23	98.16	98.34	98.28
KTH-TIPS(40)	98.12	98.40	98.36	98.32	98.42	98.45
FMD(50)	52.32	54.88	54.83	55.52	57.11	57.55
Flowers 102(20)	70.28	75.69	72.73	75.29	79.44	80.24
Swed_Leaves 15(25)	98.72	99.38	98.85	99.03	99.47	99.55
PFID(12)	40.16	43.08	41.71	44.08	45.36	44.44
Scene-15(100)	75.00	79.15	77.71	78.62	81.50	82.04
MIT-Indoor 67	39.43	44.62	42.82	43.26	47.87	48.05

TABLE 6
Comparison of Running Time (Seconds) of $\mathrm{PRICoLBP}$, LBP, MSLBP, and Bag-of-SIFT Model

Image Sizes	LBP	MSLBP	$\mathrm{PRICoLBP}_0$	$\mathrm{PRICoLBP}_g$	BoW(Vlfeat)
200×200	0.003	0.016	0.005	0.031	0.165
400×400	0.011	0.057	0.022	0.135	0.717
600×600	0.026	0.138	0.052	0.319	1.693
800×800	0.044	0.218	0.085	0.591	3.179
1000×1000	0.068	0.338	0.127	0.930	5.275
2000×2000	0.250	1.039	0.506	4.412	21.425

up-sample the original image to three times of the original size, and compute LBP for each point. When the LBP patterns for all points are obtained, we just need lookup table operator to get the $\mathrm{PRICoLBP}_g$ feature. The main computational cost comes from the computation of the LBP pattern for the up-sampled image. Compared with bag-of-sift method, $\mathrm{PRICoLBP}_g$ without using SSE speedups is more than five times faster.

4.6 Comparison with LBP Variants

In this section, we compare PRICoLBP with some powerful LBP variants, including LBPV, CoALBP, LBPHF_S, CLBP_S/M/C, and LBPHF_S_M. In literature, LBP variants were usually used for texture and material classification. Therefore we use texture and material classification task to evaluate the performance. Results are shown in Table 7.

Table 7 shows that:

- PRICoLBP outperforms CoALBP significantly on all data sets;
- PRICoLBP achieves better performance than VLBP, LBPHF_S, LBPHF_S_M on texture classification data sets Brodatz, CUReT, and KTH-TIPS, and yields comparable performance with CLBP_S/M/C;
- PRICoLBP outperforms significantly all LBP variants on material data set FMD.

Note that CLBP_S/M/C, LBPHF_S, LBPHF_S_M, and LBPV usually encode multi-scale information, e.g., using $\mathrm{LBP}(8,1)$, $\mathrm{LBP}(16,2)$, and $\mathrm{LBP}(24,3)$; whereas PRICoLBP in this case uses the spatial co-occurrence of a single-scale and hence runs much faster than CLBP_S/M/C and LBPHF_S_M.

5 Applications

In this section, to keep consistency, we will use $PRICoLBP_g$ for all experiments. In Brodatz, CUReT, KTH-TIPS, Leaf, and Scene-15 data sets, we used 2_a templates configurations. In FMD, Oxford Flower 102, PRID, and Mit-indoor 67 data sets, we used the 6 templates configurations. Consequently,

TABLE 7
Comparison with Other LBP Variants

Methods	Brodatz	KTH-TIPS	CUReT	FMD
$PRICoLBP_g$	**96.9**	**98.4**	98.4	**57.1**
CoALBP [28]	94.2	97.0	98.0	45.2
CLBP_S/M/C [25]	94.8	**98.4**	**98.9**	50.3
LBPHF_S_M [27]	95.3	97.8	97.6	49.2
LBPHF_S [26]	94.6	97.0	95.9	45.5
LBPV [22]	93.8	95.5	94.0	39.8
LBP [1]	91.6	92.2	96.3	43.5

TABLE 8
Texture Classification Results on Brodatz, CUReT, and KTH-TIPS

Methods	Brodatz(3)	CUReT(46)	KTH-TIPS(40)
$\mathrm{PRICoLBP}_g$	**96.9%**	98.4%	**98.4%**
CLBP_S/M/C [25]	94.8 %	**98.9%**	**98.4%**
LBPHF_S_M [27]	95.3%	97.6%	97.8%
LBPHF_S [26]	94.6	95.9	97.0
CoALBP [28]	94.2%	98.0%	97.0%
LBPV [22]	93.8%	94.0%	95.5%
MSLBP [1]	91.6%	96.3%	92.2%
Liu et al. [45]	94.2%	98.5%	—
Nguyen et al. [46]	96.1%	—	95.7%
Zhang et al. [5]	95.9%	95.3%	96.1%
VZ-Patch [43]	92.9%	98.0%	92.4%
Caputo et al. [44]	95.0%	98.5%	94.8%
Lazebnik et al. [3]	88.2%	72.5%	91.3%

the dimensionality of the $PRICoLBP_g$ on the Brodatz, CUReT, KTH-TIPS, and Leaf are $590 \times 2 = 1{,}180$, and the dimensionality for the FMD is $590 \times 6 = 3{,}540$. Color information is incorporated for Oxford Flower, PRID and MIT-indoor, the dimensionality for Oxford Flower and PRID is $3{,}540 \times 3 = 10{,}620$. Since we use two scales (LBP(8, 1) and LBP(8, 2)) on two scene data sets, the dimensionality for Gray $PRICoLBP_g$ on Scene-15 and MIT-indoor 67 is 2,360, 7,080 individually. When color information is used, the dimensionality increased to three times. When SPM (e.g., 1×1, 2×2 and 3×1) is applied, e.g., on Scene-15 and MIT-indoor, the dimensionality increased to eight times. When PCA is used, cross-validation was used to determine the dimension. In FMD, we found that 120-dimensional feature after PCA yielded the best performance.

5.1 Texture Classification

The *Brodatz* album [33] is a well-known texture classification benchmark data set. It contains 111 texture classes, with nine images in each class.

The *CUReT* data set [34] is a widely used texture classification data set. We use the same subset of images as [43], [44], which contains 61 texture classes with 92 images for each class. These images are captured under different illuminations and viewpoint directions.

The *KTH-TIPS* data set [35] consists of 10 texture classes. Each image is captured at nine scales, under three different illumination directions, with three different poses and hence each class contains 81 samples.

We use three, 46 and 40 training images per class individually for Brodatz, CUReT and KTH-TIPS and the rest for testing.

Lazebnik et al. [3] proposed to use Harris-affine corners and Laplacian-affine blobs to detect interest points, use SPIN and RIFT as descriptors, and use nearest neighbor classifier for texture classification. Caputo et al. [44], instead of nearest neighbor classifier, used kernel SVM as classifier and demonstrated that the SVM classifier could achieve better performance than the nearest neighbor classifier. In [5], Zhang et al. evaluated different features and kernels for object and texture classification. Besides of Bag of Words model, the LBP based methods, e.g., LBPV [22], LBPHF_S [26], LBPHF_S_M [27], CoALBP [28], and CLBP [25], also performed well on texture classification. We evaluated our $\mathrm{PRICoLBP}_g$ on the data sets Brodatz, CUReT, and

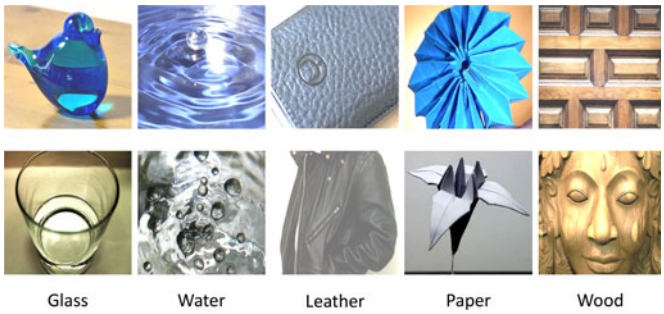


Fig. 4. Sample from five categories of FMD data set.

KTH-TIPS, and compared with all of above-mentioned the methods. For MSLBP, CLBP_S/M/C, LBPV, LBPHF_S, and LBPHF_S_M, we used three scales with the codes on the authors' websites. For CoALBP, we implemented it following [28]. Experimental results are shown in Table 8.

From Table 8, we observe that:

- PRICoLBP_g outperforms MSLBP, CoALBP, LBPV, LBPHF_S, and LBPHF_S_M significantly on all three data sets, and achieves comparable performance with CLBP_S/M/C.[6]
- PRICoLBP_g also exceeds the performance of several bag-of-words methods, [43], [44], [5], [45], and [46].

The remarkable performance of PRICoLBP_g on texture classification task validates that the encoding strategy—preserving the relative orientation angle information—is effective. Note that PRICoLBP_g performs well on data set KTH-TIPS, which contains relatively small variations in scales (i.e., nine scales, continually from 0.5 to 2). This result suggests that PRICoLBP_g is not sensitive to small scale variations.

5.2 Material Recognition

Flickr Material Database (FMD) is a newly published challenging material data set. FMD contains 10 classes, including fabric, foliage, glass, leather, metal, paper, plastic, stone, water, and wood. Each category contains 100 images, in which half of the images are in close-up views and the rest are of views at object-scale. We presented some samples from five categories in Fig. 4. Each image in FMD is associated with a binary human-labeled mask, describing the location of the object. We use just the gray-scale image and extract PRICoLBP_g within the masked regions.

Liu et al. [47] reported the first results by exploring four types of features, including color, texture and micro-texture, outline shape, and reflectance-based features. They divide these four types of features into seven features. Meanwhile, they proposed an augmented Latent Dirichlet Allocation (aLDA) method for classification. Recently, Sharan et al. [48] extended [47] by using a kernel SVM and reported that SVM outperforms the aLDA classifier significantly. In [49], Hu and Bo applied five kernel descriptors [6] with effective matching kernel (EMK) [50] to object and material recognition, and shown that kernel descriptors with EMK yielded superior performance. We conducted experiments to compare our PRICoLBP_g on FMD with all aforementioned methods. We follow the standard experimental setup—using 50 samples for training and the rest 50 samples for testing. Results are shown in Table 9.

TABLE 9
Experimental Results on Data Set FMD

Methods		50
Single Feature	Edge-slice [48]	27.2%
	Liao et al. CVPR 2013 [51]	28%
	Color [48]	33.8%
	SIFT [48]	41.2%
	MSLBP	43.5%
	CoALBP [28]	45.2%
	Kernel Descriptor in [49]	49%
	LBPHF_S_M	49.2%
	CLBP_S/M/C	50.3%
	PRICoLBP_g	**57.1%±1.8**
	PRICoLBP_g(PCA)	**55.7%±1.8**
	PRICoLBP_g(AKA)	**57.7%±1.7**
Multiple Features	Liu et al. CVPR 2010 [47]	44.6%
	Li et al. ECCV 2012 [52]	48.1%
	Hu et al. BMVC 2011 [49]	54% ±2.0

When using single feature only, PRICoLBP_g outperforms significantly other single features. Note that PRICoLBP_g improves the second-best feature relatively over 14 percent. Compare to the results of using multiple features, our results by using a single PRICoLBP_g also beat other methods. It should be mentioned that the best performance on FMD is 60.6 percent, reported recently by Sharan et al. [48], in which eight features were combined. Images in FMD data set have natural rotation and large intra-class variations. The remarkable performance on FMD data set confirms the well-balanced tradeoff between discriminativeness and robustness of PRICoLBP_g.

Note that when using PCA to reduce the dimension and using kernel SVM as classifier, the performance of PRICoLBP_g is degenerated about 1.4 percent. This result suggests that using PCA to reduce the dimension may remove some discriminative information. With an additive kernel approximation and using a linear SVM, the performance of PRICoLBP_g is slightly increased (about 0.6 percent).

5.3 Flower Recognition

Oxford Flowers 102 data set [30] has 8,189 images, including 102 categories, with $40-258$ samples per category. We show some samples in Fig. 5. Previous works, e.g., [30], [31], and [32] reported that segmentation could greatly boost the final performance. Here we use Grabcut [53] to segment the flower foreground from each image. It has been shown in

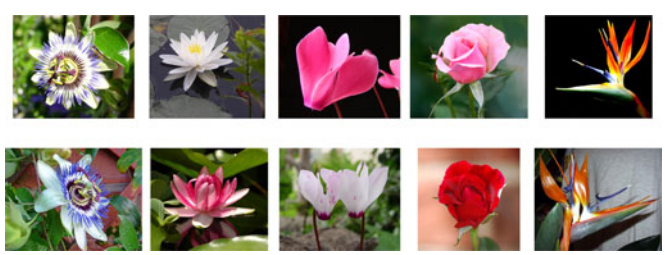


Fig. 5. Some sample from Oxford Flower 102 data set. (The two images in each column from the same category).

6. We used the best CLBP settings (i.e., CLBP_S/M/C) in this paper.

TABLE 10
Recognition Results on Oxford Flower 102 Data Set

Single Features	20	30	Multiple Features	20
SIFT_B [30]	32.0	–	Ito et al. [9]	53.2
HSV [30]	43.0	–	Nilsback et al. [30]	72.8
HOG [30]	49.6	–	Yuan et al. [55]	74.1
SIFT internal [30]	55.1	–	Nilsback's thesis [31]	76.8
CoHED [9]	48.2	–	Grabcut [32]	77.0
MSLBP	52.0	56.1	Chai et al. [32]	80.0
MSDS [32]	69.5	73.4	PRICoLBP$_g$+MSDS	84.2
LDC [56]	61.45	–	Chai et al. [57]	**85.6**
Kanan et al [10]	71.4	75.2		
PRICoLBP$_g$	**79.1**	**82.3**		

[30], [31], [32], [54] that color information could boost the recognition performance. Therefore we extract color PRICoLBP$_g$ features on the foreground region, in which the foreground images were resized into the minimum resolution 128 (while keeping the ratio between the width and height). We follow the same train/test protocol as in [32]—using 20 or 30 training samples per class and the rest for testing. We compare PRICoLBP$_g$ with the other works in single feature and multiple features settings. All results are originally reported, except for MSLBP, CoHED, and Multi-Scale Dense Sampling (MSDS). For MSLBP, we extracted the features from three channels and hence the dimension of MSLBP was $54 \times 3 = 162$. We re-implemented CoHED [9] and the bag-of-MSDS [54]. The results are shown in Table 10.

Table 10 shows that:

- When using a single feature, PRICoLBP$_g$ outperforms all other methods. Note that PRICoLBP$_g$ outperformed MSLBP remarkably over 27 percent and outperformed another co-occurrence features, CoHED, over 30 percent.
- When using a combination of multiple features, PRICoLBP$_g$ combining with MSDS achieves a recognition rate at 84.2 percent (only slightly worse than the best result reported recently by Chai et al. [57]), which significantly exceeded the results of using each of them alone. This fact suggests that PRICoLBP$_g$ is complementary with MSDS. The performance of using single feature in [57] is not reported.

The remarkable performance of PRICoLBP$_g$ on flower recognition data set can be accounted from the following facts. Flower images contain rich color, texture, and shape.

TABLE 11
Recognition Performance on Swedish Leaf

Methods	Input	Performance
Söderkvist [37]	Contour only	82.40%
SC + DP [59]	Contour only	88.12%
Spatial PACT [58]	Contour only	90.61%
IDSC+DP [59]	Contour only	94.13%
SPTC + DP [59]	Gray-scale Image	95.33%
Shape-Tree [60]	Contour only	96.28%
SLPA [61]	Gray-scale Image	96.33%
Spatial PACT [58]	Gray-scale Image	97.87%
PRICoLBP$_g$ + SVM	Gray-scale Images	**99.38%**

Fig. 6. Example images from Swedish Leaf with one sample per category. The final image is an example of mask image.

Fortunately the proposed PRICoLBP$_g$ can capture the texture information well, incorporate color information easily, and also emphasize the shape information.

5.4 Leaf Recognition

The Swedish leaf data set [37] consists of 15 species of leaves, with 75 images per species. Some samples are shown in Fig. 6. Following the standard protocol [37], [58], we randomly select 25 images per class for training and the rest for testing.

Swedish leaf data set has two characteristics: a) The leaf images are manually aligned well; b) The leaf images are in a good quality, without obvious damage. Thus the spatial layout prior of leaves can be used to boost the classification accuracy.

Shape-based methods are widely used for leaf classification, such as [59], [60]. However, these methods suffer from their own limitations. On one hand, accurate shape extraction is still a challenging problem, and shape descriptor can be affected heavily by the partial damage of leaf and footstalk. On the other hand, while some categories share the similar shapes, the shapes of leaves from the same category may vary greatly, due to geometric distortion and photometric variation, such as viewpoint change and folding of leaves.

Experimental results are presented in Table 11. We observe that PRICoLBP$_g$ with kernel SVM outperforms other methods significantly. Note that PRICoLBP$_g$ reduced the classification error of PACT [58] relatively over 70 percent. with the same classifier. Besides, our method with PRICoLBP$_g$ did not use the spatial layout prior information. However in the spatial PACT, each image was divided into three-layer pyramid, in which the first layer had one block (i.e., the original image), the second layer contained five block, and the third had $5 \times 5 = 25$ blocks. If the leaf images were well-aligned, as in Swedish leaf data set, the spatial prior information could provide important cues for classification. However, in practical cases, the condition can hardly be satisfied.

Fig. 7. Sample images from PRID food data set.

TABLE 12
Classification Accuracy on PFID Data Set and Dimensionality of the Representation Used in Each Method

Algorithm	ColorH	OM	BOW	SPMK	MSF	CH	PDK	TB	GPVP	ORSC	PRICoLBP$_g$
Accuracy(%)	11.3	28.2	30.6	33.4	32.7	34.9	35.3	38.4	35.9	42.9	**45.4**
Dimensionality	200	6,144	50	1,050	100	1,500	40,000	54,732	100,000	15,452	10,620

The results of BOW, SPMK, MSF, CH, PDK, TB, GPVP, and ORSC are taken from [62].

Different from shape-based methods, which are sensitive to the inaccurate shape extraction caused by partial damage, our PRICoLBP$_g$ is robust to the partial loss in leaf images or the clutter background.

5.5 Food Recognition

Pittsburgh Food Image Database (PFID) [38] is a recently released food data set. The PFID consists of fast food images and videos from 13 chain restaurants that are acquired in laboratory. As in [17], we use a subset of 61 categories of specific food items (e.g., McDonald's Big Mac) with masked background. Each food category consists of three different instances (bought in different days from different branches of the restaurant chain). Each instance contains six images in six different viewpoints. Fig. 7 shows some samples.

Following the standard experimental protocol proposed by Chen et al. [38], we perform three-fold cross-validation and use 12 images from the first two instances for training and six images from the third instance for testing. This protocol ensures that no images of any given food item appears in both the training and test sets. We report the averaged recognition accuracy.

We compare the proposed PRICoLBP$_g$ with the state-of-the-art methods in the literature, including color histogram (ColorH) [38], midpoint category pairwise features (OM) [17], the baseline Bag-of-Words representation, spatial pyramid matching kernel (SPMK) [39], Markov stationary features, contextualized histogram (CH), TextonBoost (TB), geometry-preserving visual phrases (GPVP) [12], and omni-range spatial contexts (ORSC) [62]. Experimental results are listed in Table 12.

We read from Table 12 that our PRICoLBP$_g$ achieves the highest accuracy, with a moderate dimensionality. Recently, Wang et al. [56] proposed to a Linear Distance Coding (LCD) for food recognition, in which Spatial Pyramid Matching was used to construct the final representation. The recognition rate of LCD with SPM is 48.45 percent. If we use PRICoLBP$_g$ with SPM, the recognition rate can reach 51.5 percent. Note also that our PRICoLBP$_g$ is very fast.

5.6 Scene Classification

Scene-15 data set was gradually built. The initial eight classes were collected by Oliva and Torralba [63], and then five categories were added by Fei-Fei and Perona [64]. Finally, two additional categories were introduced by Lazebnik et al. [39]. Scene-15 data set has 15 categories, including office, kitchen, living room, bedroom, store, industrial, tall building, inside cite, street, highway, coast, open country, mountain, forest, and suburb. Images are about 250×300 resolution, with 210 to 410 images per category. Some samples are shown in Fig. 8. In our experiments, we resized the images to have the minimum dimension of 256 pixels (while maintaining the aspect ratio).

Fig. 8. Samples from Scene 15 data set.

We compared the propose PRICoLBP$_g$ feature with several popular scene recognition methods. We use three-levels SPM with the settings of 1×1, 2×2, and 3×1. In experiments, we found that using the more subtle spatial division (e.g., 1×1, 2×2 and 4×4) was not necessary. Experimental results are shown in Table 13.

Table 13 shows that without using SPM, our PRICoLBP$_g$ feature outperforms other single feature methods significantly, e.g., GIST [63], Bag-of-SIFT [39], and CENTRIST [58]. Using SPM can improve the performance significantly. From Table 13, we observe that our method outperforms the traditional Bag-of-SIFT with SPM, CENTRIST with SPM, RBoW, Object Bank, LDC [56], and SPMSM, and yields comparable performance with LPR-RBF [70].

Scene 15 contains both indoor and outdoor scenes. Wu et al. [58] pointed out that GIST worked well on outdoor categories and performed poor on indoor categories. But our PRICoLBP$_g$ performs well on both indoor and outdoor categories.

MIT Indoor 67 data set was firstly introduced by [71]. It is a challenging data set, including 15,620 indoor scenes images in 67 different categories. Some samples are shown in Fig. 9. Following [71], we used 80 images per class for training and 20 images per class for testing. We used the same train/test configuration files provided by the authors of [71]. We resized the images to have the maximum dimension of 400 pixels (while maintaining the aspect ratio).

We evaluated four different implementations of our PRICoLBP, i.e., PRICoLBP on gray-scale images (gray-PRICoLBP), PRICoLBP on color images (Color-PRICoLBP),

TABLE 13
Classification Results on Scene 15 Data Set

Feature Types	Without SPM	With SPM
16 channel weak features [39]	45.3 ± 0.5	66.8 ± 0.6
GIST [63]	71.2 ± 0.8	$--$
Bag of SIFT(200 Codebooks) [39]	72.2 ± 0.6	81.1 ± 0.3
Bag of SIFT(800 Codebooks) [39]	74.8 ± 0.3	81.4 ± 0.5
CENTIST [58]	73.29 ± 0.96	83.9 ± 0.8
BSC [16]	72.5 ± 0.3	–
Kernel Codebook [65]	75.6 ± 0.7	–
$PRICoLBP_g$	$\mathbf{79.2\% \pm 0.7}$	84.3 ± 0.7
RBoW [66]	Object Bank [67]	LDC [56]
78.5	80.9	82.6
SCP + SPM [68]	SPMSM [69]	LPR-RBF [70]
80.4	82.3	**85.8**

Fig. 9. Samples from MIT Indoor 67 data set.

TABLE 14
Average Accuracy (Percent) for MIT Indoor Scene Data Set

HOG	22.8	DPM + GIST-Color	39.0
GIST-grayscale [63]	22.0	DPM + SPM	40.5
GIST-color [63]	29.7	DPM + GIST-Color + SPM	43.1
Spatial Pyramid [39]	34.4	RBoW [66] ECCV 2012	37.9
GIST-color + SPM	38.5	SPMSM [69] ECCV 2012	44.0
ROI + GIST [71] CVPR 2009	26.5	LPR-RBF [70] ECCV 2012	44.8
MM-scene [73] NIPS 2010	28.0	Gray-PRICoLBP	43.4
DPM [72] ICCV 2011	30.4	Color-PRICoLBP	47.9
CENTRIST [58] PAMI 2012	36.9	SPM-gray-PRICoLBP	47.9
Object Bank [67] NIPS 2010	37.6	SPM-Color-PRICoLBP	**51.0**

The results of HOG, GIST-grayscale, GIST-color, SPM and DPM related methods are taken from [72].

PRICoLBP on gray-scale images with SPM (SPM-gray-PRICoLBP), PRICoLBP on color images with SMP (SMP-Color-PRICoLBP), and compared them with the well-known methods, e.g., HOG, GIST, GIST-color, Spatial pyramid, Object Bank, DPM, and several recently published methods [58], [69], [66], [72]. For the color PRICoLBP feature, we extracted PRICoLBP features from each channel, and then concatenated them together. When use SPM, we use three-levels SPM with the settings of 1×1, 2×2, and 3×1. We use linear SVM with square root feature. Results are shown in Table 14.

We observe from Table 14 that:

- Gray-scale PRICoLBP obtains 43.4 percent which significantly outperforms its counterparts, e.g., HOG, GIST and DPM.
- Color-PRICoLBP improves the gray-scale PRICoLBP by 4.5 percent. Similar improvement is also observed from gray-scale GIST to GIST-color. This confirms that the color information in MIT-indoor data set is effective.
- SPM-PRICoLBP improves Gray-PRICoLBP by 4.5 percent, and significantly outperforms traditional SPM + bag-of-sift and CENTRIST. This result suggests that the spatial layout information in MIT-indoor data set is useful.
- SPM-Color-PRICoLBP reaches 51.0 percent which is remarkably higher than the state-of-the-art methods, e.g., Object Bank [67], SPMSM [69], and LPR-RBF [70].

Note in Table 15, our PRICoLBP achieves the best performance on 54 categories out of all 67 categories, compared with RBoW, DPM, and Gist-color. Unlike DPM based method, our method does not need detect any specific "objects" or "parts" explicitly, and its computational cost is extremely low.

6 Conclusion and Discussion

In this paper, we addressed the transform invariance issue for co-occurrence feature. Specifically, we presented a pairwise transform invariance principle, proposed an effective and efficient co-occurrence encoding scheme, *pairwise rotation invariant co-occurrence* LBP, and extended it to incorporate multi-scales, multi-orientations, and multi-color channels information. Different from other LBP variants, the PRICoLBP can not only capture the spatial context co-occurrence information effectively, but also possess rotation invariance. We evaluated the performance of the PRICoLBP comprehensively on nine benchmark data sets from five different perspectives. Moreover we also applied the PRICoLBP to six different but related applications and demonstrated that PRICoLBP was efficient, effective, and of a well-balanced tradeoff between the discriminativeness and robustness.

The PRICoLBP has many potential applications, such as medical image analysis (e.g., Indirect immunofluorescence (IIF) Cell Classification, Breast Cancer Tumor Tissue Classification, Lung Tissue Classification, Hepatic Tissues Classification), Biometrics (e.g., Palm Recognition, Iris Recognition), Remote Sensing, and Face Detection and Recognition, Dynamic Texture and Scene Recognition, and Dynamic Facial Expression Recognition, and other texture related applications. Owing to its low computational cost, it is convenient to combine it with other feature for some real world challenging problem. For example, PRICoLBP

TABLE 15
Category-Wise Accuracy (Percent) for $\mathrm{PRICoLBP}_g$, RBoW [66], DPM [72], and GIST-Color [63] on MIT Indoor 67

Category	Ours	RBoW	DPM	GC	Category	Ours	RBoW	DPM	GC	Category	Ours	RBoW	DPM	GC	Category	Ours	RBoW	DPM	GC
buffet	**90**	65	75	50	cloth. store	**67**	44	33	11	hos. room	**50**	40	5	15	lobby	**30**	10	**30**	**30**
greenhouse	**90**	75	65	55	comp. room	**67**	56	22	28	subway	**48**	33	38	38	artstudio	**30**	15	5	10
elevator	**86**	62	52	67	nursery	**65**	55	60	50	grocerystore	**48**	33	19	43	lab. wet	**27**	**27**	5	9
Ins. subway	**86**	81	62	10	corridor	**62**	43	57	48	fastfoodrest	**47**	24	12	18	jewel. shop	**27**	0	5	5
cloister	80	80	**90**	80	garage	**61**	44	56	28	winecellar	**48**	29	14	43	toystore	**27**	14	9	14
chu. inside	**79**	74	63	74	auditorium	**61**	44	11	22	meeting r.	45	41	**75**	45	deli	**26**	0	5	16
casino	**79**	47	32	32	tv studio	**61**	22	6	33	bathroom	44	33	**50**	33	oper. room	**26**	5	5	**26**
inside bus	**78**	**78**	43	48	poolinside	**60**	10	0	55	den. office	43	**48**	24	33	shoeshop	**26**	21	16	11
pantry	**75**	55	75	40	stairscase	**60**	35	35	35	bakery	**42**	5	11	37	museum	26	**39**	13	4
bowling	75	**85**	35	45	rest. kitchen	**57**	26	4	17	air. inside	**40**	20	5	5	gameroom	25	**45**	40	10
florist	74	**84**	79	63	library	**55**	50	0	35	living r.	**40**	5	20	10	mall	**25**	**25**	**25**	20
stu. music	**74**	37	32	42	kitchen	**52**	38	29	43	child. room	**39**	28	6	17	waiting r.	**24**	**24**	5	14
classroom	**72**	**72**	67	39	gym	**50**	28	22	11	bedroom	**38**	19	5	0	warehouse	**24**	19	24	24
closet	**72**	61	44	50	movie the.	50	**55**	45	25	laundromat	36	36	**45**	18	bar	22	**44**	11	11
kin. garden	**70**	40	15	25	videostore	**50**	41	18	18	prisoncell	35	**45**	40	35	restaurant	**20**	10	5	0
concert hall	**70**	65	65	60	bookstore	**50**	30	45	20	locker r.	**33**	19	19	5	office	**14**	10	10	10
trainstation	**70**	45	35	55	din. room	**50**	28	28	**50**	hairsalon	33	19	**43**	29					

combined with Bag-of-SIFT method won the ICIP 2013 Competition on Cells Classification by Fluorescent Image Analysis [74]. Meanwhile, on some other medical applications, the proposed feature also shows superior performance, such as Tissue Classification. We hope that the PRICoLBP can become a *de facto standard* tool for texture relevant classification and retrieval applications.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for constructive comments. We also want to thank Dr. Guoying Zhao for giving some useful comments. X. Qi, C.-G. Li, and J. Guo are supported by the Natural Science Foundation of China (NSFC) under Grant nos. 61175011, 61273217, and 61171193, and the 111 Project under Grant no. B08004. Y. Qiao and X. Tang are supported by NSFC (91320101), Shenzhen Basic Research Program (JC201005270350A, JCYJ20120903092050890, JCYJ20120617114614438), 100 Talents Programme of CAS.

REFERENCES

[1] 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.

[2] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.

[3] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.

[4] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *Int. J. Comput. Vis.*, vol. 62, pp. 61–81, 2005.

[5] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," in *Proc. IEEE Workshop Comput. Vis. Pattern Recognit. Workshop*, 2007, pp. 213–238.

[6] L. Bo, "Kernel descriptors for visual recognition," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 244–252.

[7] O. Boiman, E. Shechtman, and M. Irani, "In defense of nearest-neighbor based image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[8] P. Chang and J. Krumm, "Object recognition with color co-occurrence histograms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1999.

[9] S. Ito and S. Kubota, "Object classification using heterogeneous co-occurrence features," in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 209–222.

[10] C. Kanan and G. Cottrell, "Robust classification of objects, faces, and flowers using natural image statistics," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2472–2479.

[11] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[12] Y. Zhang, Z. Jia, and T. Chen, "Image retrieval with geometry-preserving visual phrases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 809–816.

[13] O. Chum and J. Matas, "Unsupervised discovery of co-occurrence in sparse high dimensional data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 3416–3423.

[14] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Systems, Man and Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[15] J. Yuan, M. Yang, and Y. Wu, "Mining discriminative co-occurrence patterns for visual recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2777–2784.

[16] N. Rasiwasia and N. Vasconcelos, "Holistic context modeling using semantic co-occurrences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1889–1895.

[17] S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar, "Food recognition using statistics of pairwise local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2249–2256.

[18] G. A. Orban, "Higher order visual processing in macaque extrastriate cortex," *Physiological Rev.*, vol. 88, no. 1, pp. 59–89, 2008.

[19] Y. Yang and S. Newsam, "Spatial pyramid co-occurrence for image classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 1465–1472.

[20] X. Qi, R. Xiao, J. Guo, and L. Zhang, "Pairwise rotation invariant co-occurrence local binary pattern," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 158–171.

[21] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Proc. 3rd Int. Conf. Anal. Model. Faces Gestures*, 2007, pp. 168–182.

[22] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using lbp variance (lbpv) with global matching," *Pattern Recognit.*, vol. 43, pp. 706–719, 2010.

[23] N.-S. Vu and A. Caplier, "Face recognition with patterns of oriented edge magnitudes," in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 313–326.

[24] N.-S. Vu and A. Caplier, "Mining patterns of orientations and magnitudes for face recognition," in *Proc. IEEE Int. Joint Conf. Biometrics*, 2011, pp. 1–8.

[25] Z. Guo, L. Zhang, 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, Jun. 2010.

[26] T. Ahonen, J. Matas, C. He, and M. Pietikäinen, "Rotation invariant image description with local binary pattern histogram fourier features," in *Proc. 16th Scandinavian Conf. Image Anal*, 2009, pp. 61–70.

[27] G. Zhao, T. Ahonen, J. Matas, and M. Pietikainen, "Rotation-invariant image and video description with local binary pattern features," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1465–1477, Apr. 2012.

[28] R. Nosaka, Y. Ohkawa, and K. Fukui, "Feature extraction based on co-occurrence of adjacent local binary patterns," in *Proc. 5th Pacific Rim Conf. Adv. Image Video Technol.*, 2012, pp. 82–91.

[29] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, pp. 91–110, 2004.

[30] M. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *Proc. 6th Indian Conf. Comput. Vis., Graph. Image Process.*, 2008, pp. 722–729.

[31] M. Nilsback, "An automatic visual flora—Segmentation and classification of flowers images," in Ph.D. thesis, Department of Engineering Science, University of Oxford, Oxford, U.K., 2009.

[32] Y. Chai, V. Lempitsky, and A. Zisserman, "Bicos: A bi-level co-segmentation method for image classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 2579–2586.

[33] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*. New York, NY, USA: Dover, 1999.

[34] K. Dana, B. Van Ginneken, S. Nayar, and J. Koenderink, "Reflectance and texture of real-world surfaces," *ACM Trans. Graph.*, vol. 18, pp. 1–34, 1999.

[35] E. Hayman, B. Caputo, M. Fritz, and J. Eklundh, "On the significance of real-world conditions for material classification," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 253–266.

[36] L. Sharan, R. Rosenholtz, and E. Adelson, "Material perception: What can you see in a brief glance?" *J. Vis.*, vol. 9, p. 784, 2009.

[37] O. Söderkvist, "Computer vision classification of leaves from swedish trees," Master's thesis, Department of Electrical Engineering, Linköping University, Linköping, Sweden, 2001.

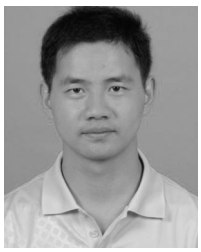
[38] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar, and J. Yang, "PFID: Pittsburgh fast-food image dataset," in *Proc. 16th IEEE Int. Conf. Image Process.*, 2009, pp. 289–292.

[39] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 2169–2178.

[40] A. Vedaldi and A. Zisserman, "Efficient additive kernels via explicit feature maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 480–492, 2012.

[41] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, "Sun database: Large-scale scene recognition from abbey to zoo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 3485–3492.

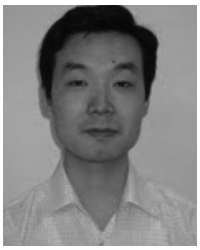
[42] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *Proc. ACM Int. Conf. Multimedia*, 2010, pp. 1469–1472.
[43] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2032–2047, Nov. 2008.
[44] B. Caputo, E. Hayman, M. Fritz, and J. Eklundh, "Classifying materials in the real world," *Image Vis. Comput.*, vol. 28, pp. 150–163, 2010.
[45] L. Liu and P. Fieguth, "Texture classification from random features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 574–586, Mar. 2012.
[46] H. Nguyen, R. Fablet, and J. Boucher, "Visual textures as realizations of multivariate log-Gaussian Cox processes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2945–2952.
[47] C. Liu, L. Sharan, E. Adelson, and R. Rosenholtz, "Exploring features in a Bayesian framework for material recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 239–246.
[48] L. Sharan, C. Liu, R. Rosenholtz, and E. H. Adelson, "Recognizing materials using perceptually inspired features," *Int. J. Comput. Vis.*, vol. 103, pp. 348–371, 2013.
[49] D. Hu and L. Bo, "Toward robust material recognition for everyday objects," in *Proc. Brit. Mach. Vis. Conf.*, 2011, pp. 48.1–48.11.
[50] L. Bo and C. Sminchisescu, "Efficient match kernels between sets of features for visual recognition," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 1730, 2009, p. 1731.
[51] Z. Liao, J. Rock, Y. Wang, and D. Forsyth, "Non-parametric filtering for geometric detail extraction and material representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 963–970.
[52] W. Li and M. Fritz, "Recognizing materials from virtual examples," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 345–358.
[53] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts," in *ACM Trans. Graph*, 2004, vol. 23, pp. 309–314.
[54] Y. Chai, "Recognition between a large number of flower species," Masters' thesis, Department of Electrical Engineering, Swiss Federal Inst. Technol. Zurich, Switzerland, 2011.
[55] X. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 3493–3500.
[56] Z. Wang, J. Feng, S. Yan, and H. Xi, "Linear distance coding for image classification," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 537–548, Feb. 2013.
[57] Y. Chai, E. Rahtu, V. Lempitsky, L. Van Gool, and A. Zisserman, "Tricos: A tri-level class-discriminative co-segmentation method for image classification," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 794–807.
[58] J. Wu and J. Rehg, "Centrist: A visual descriptor for scene categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1489–1501, Aug. 2011.
[59] H. Ling and D. Jacobs, "Shape classification using the inner-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 286–299, Feb. 2007.
[60] P. Felzenszwalb and J. Schwartz, "Hierarchical matching of deformable shapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.
[61] S. Zhang, Y. Lei, T. Dong, and X.-P. Zhang, "Label propagation based supervised locality projection analysis for plant leaf classification," *Pattern Recognit.*, vol. 46, pp. 1891–1897, 2013.
[62] B. Ni, M. Xu, J. Tang, S. Yan, and P. Moulin, "Omni-range spatial contexts for visual classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 3514–3521.
[63] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, pp. 145–175, 2001.
[64] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 524–531.
[65] J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, and J. M. Geusebroek, "Visual word ambiguity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1271–1283, Jul. 2009.
[66] S. N. Parizi, J. G. Oberlin, and P. F. Felzenszwalb, "Reconfigurable models for scene recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2775–2782.
[67] L.-J. Li, H. Su, E. P. Xing, and L. Fei-Fei, "Object bank: A high-level image representation for scene classification and semantic feature sparsification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1378–1386.
[68] L. Wang, Y. Li, J. Jia, J. Sun, D. Wipf, and J. M. Rehg, "Learning sparse covariance patterns for natural scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2767–2774.
[69] R. Kwitt, N. Vasconcelos, and N. Rasiwasia, "Scene recognition on the semantic manifold," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 359–372.
[70] F. Sadeghi and M. F. Tappen, "Latent pyramidal regions for recognizing scenes," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 228–241.
[71] A. Quattoni and A. Torralba, "Recognizing indoor scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 413–420.
[72] M. Pandey and S. Lazebnik, "Scene recognition and weakly supervised object localization with deformable part-based models," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 1307–1314.
[73] J. Zhu, L.-J. Li, L. Fei-Fei, and E. P. Xing, "Large margin learning of upstream scene understanding models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 2586–2594.
[74] P. Hobson, G. Percannella, M. Vento, and A. Wiliem, (2013). "Competition on Cells Classification by Fluorescent Image Analysis," *Proc. IEEE 20th Int. Conf. Image Process.*, [Online]. Available: http://nerone.diiie.unisa.it/contest-icip-2013/index.shtml.



Xianbiao Qi received the BE degree in information engineering from the Beijing University of Posts and Telecommunications (BUPT) in 2008. He is currently working toward the PhD degree at BUPT. He visited the Web Search and Mining Group in Microsoft Research Asia (MSRA) as a visiting student from January 2011 to May 2012. His research interests include texture-relevant computer vision applications, including discriminative texture feature design, and texture and material recognition.

Rong Xiao received the PhD degree from Nanjing University, China, in 2001. He joined Microsoft Research China as an associate researcher in July 2001. He is currently a senior research software engineer at Microsoft Bing, Redmond. He has published approximately 30 papers at leading conferences such as ICCV, CVPR, ECCV, and ACM Multimedia. His research interests include statistical machine learning, face detection and recognition, object detection and tracking, and local feature design.

Chun-Guang Li received the BE degree from Jilin University in 2002 and the PhD degree from the Beijing University of Posts and Telecommunications (BUPT) in 2007. Currently, he is a lecturer with the School of Information and Communication Engineering, BUPT. From July 2011 to April 2012, he visited the Visual Computing Group, Microsoft Research Asia. From December 2012 to November 2013, he visited the Vision, Dynamics and Learning lab, Johns Hopkins University. His research interests include statistical machine learning, compressive sensing, and pattern recognition. He is a member of the IEEE.



Yu Qiao received the PhD degree from the University of Electro-Communications in Japan, in 2006. He was a JSPS fellow, and then a project assistant professor with the University of Tokyo from 2007 to 2010. He is currently a professor with the Shenzhen Institutes of Advanced Technology at the Chinese Academy of Science. His research interests include pattern recognition, computer vision, multimedia, image processing, and machine learning. He has published more than 90 papers in these fields. He received the Lu Jiaxi Young Researcher Award from the Chinese Academy of Science in 2012. He is a senior member of the IEEE.

Jun Guo received the BE and ME degrees from the Beijing University of Posts and Telecommunications (BUPT), China, in 1982 and 1985, respectively, and the PhD degree from the Tohuku-Gakuin University, Japan, in 1993. At present, he is a professor and vice president of BUPT. His publications cover CVPR, ICCV, ECCV, and the *IEEE Transactions on Pattern Analysis and Machine Intelligence*. His research interests include pattern recognition theory and application, information retrieval, content-based information security, and network management.

Xiaoou Tang received the BS degree from the University of Science and Technology of China, Hefei, in 1990, and the MS degree from the University of Rochester, New York, in 1991, and the PhD degree from the Massachusetts Institute of Technology, in 1996. He is a professor in the Department of Information Engineering at the Chinese University of Hong Kong. He worked as the group manager of the Visual Computing Group at Microsoft Research Asia from 2005 to 2008. He was a program chair of the IEEE International Conference on Computer Vision (ICCV) 2009 and is an associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and the *International Journal of Computer Vision*. His research interests include computer vision, pattern recognition, and video processing. He received the Best Paper Award at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2009. He is a fellow of the IEEE.