SIViP
DOI 10.1007/s11760-015-0804-2

ORIGINAL PAPER

Land-use scene classification using multi-scale completed local binary patterns

Chen Chen[1] · **Baochang Zhang**[2] · **Hongjun Su**[3] · **Wei Li**[4] · **Lu Wang**[4]

Received: 25 April 2015 / Revised: 8 June 2015 / Accepted: 15 July 2015

Abstract In this paper, we introduce the completed local binary patterns (CLBP) operator for the first time on remote sensing land-use scene classification. To further improve the representation power of CLBP, we propose a multi-scale CLBP (MS-CLBP) descriptor to characterize the dominant texture features in multiple resolutions. Two different kinds of implementations of MS-CLBP equipped with the kernel-based extreme learning machine are investigated and compared in terms of classification accuracy and computational complexity. The proposed approach is extensively tested on the 21-class land-use dataset and the 19-class satellite scene dataset showing a consistent increase on performance when compared to the state of the arts.

Keywords Land-use scene classification · Multi-scale analysis · Local binary patterns · Extreme learning machine

✉ Baochang Zhang
bczhang@buaa.edu.cn

✉ Hongjun Su
hjsurs@163.com

✉ Wei Li
liw@mail.buct.edu.cn

Chen Chen
chenchen870713@gmail.com

[1] Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX, USA

[2] School of Automation Science and Electrical Engineering, Beihang University, Beijing, China

[3] School of Earth Science and Engineering, Hohai University, Nanjing, China

[4] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China

1 Introduction

Land-use scene classification aims to assign a semantic label, e.g., urban and forest, to an image according to its content. With the technological development of various satellite sensors, the volume of high-resolution remote sensing image data is increasing rapidly. It is necessary to develop effective and efficient classification methods to annotate the large remote sensing images for building intelligent databases.

It is of great interest in exploiting computer vision techniques for classifying aerial or satellite images. For example, the bag-of-visual-words (BOVW) model [1] establishes a set of visual words (i.e., a visual vocabulary) by clustering the local features extracted from a collection of images. Then, an image is represented by a histogram of the frequency of visual words. The BOVW model has been successfully applied to remote sensing land-use scene classification and achieved good performance [2,3]. However, the model only considers the occurrences of the visual words and ignores the spatial information in the images. Since land-use remote sensing images contain rich texture information, the performance of the BOVW model can be improved by incorporating spatial information. In [3], a spatial co-occurrence kernel was used to explore spatial extensions to the BOVW model for land-use scene classification. In [4], a 2D wavelet decomposition-based BOVW model was proposed to exploit the texture structures in land-use images. The spatial pyramid matching (SPM) model [5] is another approach to address the lack of spatial information in the BOVW representation. In the SPM model, an image is partitioned into increasingly fine subregions and histograms of local features computed from each subregion are concatenated. In [6], multi-resolution analysis and dual modalities for horizontal and vertical partitions were incorporated into the SPM model to further improve the land-use scene classification performance. How-

Published online: 31 July 2015

ever, one limitation of the SPM model is that it captures the absolute spatial information with ordered block partitions of an image, which is sensitive to rotation variations of the image scenes. To achieve rotation invariance, a concentric circle-based partition strategy [7] was proposed to represent the spatial information in remote sensing images. A pyramid-of-spatial-relatons (PSR) model [8] was developed to capture both absolute and relative spatial relationships of local features leading to a translation and rotation invariance representation for land-use scene images.

Although the aforementioned variants of the original BOVW model are able to capture the spatial layout information of scene images, they mostly use the scale-invariant feature transform (SIFT) descriptor [9] to extract local features and have not exploited other effective local features to capture the texture and structure information in remote sensing images. There are some studies that put efforts on evaluating various features and combinations of features for land-use scene classification. For instance, local structural texture similarity descriptor was applied to image blocks to achieve structural texture representation for aerial image classification [10]. An enhanced Gabor texture descriptor (EGTD) [11] was a global image descriptor based on cross-correlation between spatial frequency subbands of Gabor image decomposition. The descriptor achieved good results in classification of high-resolution remote sensing images. In [12], three different feature extraction strategies including raw pixel intensity values, oriented filter responses and dense SIFT descriptors were investigated in the land-use scene classification framework. Combining multiple features such as SIFT, GIST [13], DAISY [14] was also explored to enhance the discriminative power of features, e.g., [15,16]. Although classification performance can be improved by fusing different features, parameter tuning for each feature is necessary. Moreover, computational complexity and feature dimensionality may be increased significantly.

In this study, we focus on the development of an effective and efficient feature extraction method for land-use scene classification. Local binary patterns (LBP) [17] descriptor, as a texture descriptor, has been successfully applied to a variety of image processing and machine vision applications [18,19]. Completed local binary patterns (CLBP) [20] descriptor is a completed modeling of the LBP operator designed to achieve significant improvement over the LBP for rotation invariant texture classification. Due to the effectiveness of the CLBP descriptor, we propose a texture-based classification paradigm using a multi-scale CLBP (MS-CLBP) extraction method. Two different implementations of the MS-CLBP are investigated and compared in terms of classification accuracy and computational complexity. For scene classification, kernel-based extreme learning machine (KELM) [21] is employed for its efficient computation and good classification performance. We evaluate the proposed MS-CLBP feature extraction approach using two publicly available land-use scene datasets to demonstrate its effectiveness for land-use scene classification.

There are two main contributions of our work:

1. We introduce the CLBP operator for the first time on remote sensing land-use scene classification. We propose an effective texture feature extraction method using MS-CLBP and carry out extensive experiments to demonstrate its superior performance over the multi-resolution LBP described in [17] and other state-of-the-art methods for land-use scene classification.
2. We present two implementations of the proposed MS-CLBP descriptor. The performance of the two implementations is compared in terms of classification accuracy and computational complexity. Furthermore, the entire classification framework is devised to be computationally efficient.

This paper is organized as follows. Section 2 describes details of the proposed feature extraction method. Section 3 presents land-use scene classification using KELM. Section 4 provides the experimental results with two land-use scene datasets. Section 5 is the conclusion.

2 Feature extraction

2.1 Completed local binary patterns

LBP [17] is an effective measure of spatial structure information of local image texture. Given a center pixel t_c, its neighboring pixels are equally spaced on a circle of radius r ($r > 0$) with the center at t_c. If the coordinates of t_c are $(0, 0)$ and m neighbors $\{t_i\}_{i=0}^{m-1}$ are considered, the coordinates of t_i are $(-r\sin(2\pi i/m), r\cos(2\pi i/m))$. The LBP is computed by thresholding the neighbors $\{t_i\}_{i=0}^{m-1}$ with the center pixel t_c to generate an m-bit binary number. The resulting LBP for t_c can be expressed in decimal form as follows:

$$\mathrm{LBP}_{m,r}(t_c) = \sum_{i=0}^{m-1} s(t_i - t_c)2^i = \sum_{i=0}^{m-1} s(d_i)2^i, \quad (1)$$

where $d_i = (t_i - t_c)$ is the difference between each neighbor and the center pixel, $s(d_i) = 1$ if $d_i \geq 0$ and $s(d_i) = 0$ if $d_i < 0$. The LBP only uses the sign information of d_i while ignoring the magnitude information. However, the sign and magnitude are complementary, and they can be used to exactly reconstruct the difference d_i. In the CLBP [20] scheme, the image local differences are decomposed into two complementary components: the signs and magnitudes (absolute values of d_i, i.e., $|d_i|$). Figure 1 shows an example of the sign and magnitude components of the CLBP extracted

67	12	73
15	34	26
38	54	40

(a)

33	-22	39
-19		-8
4	20	6

(b)

1	-1	1
-1		-1
1	1	1

(c)

33	22	39
19		8
4	20	6

(d)

Fig. 1 **a** 3×3 sample block. **b** The local differences. **c** The sign component of CLBP. **d** The magnitude component of CLBP

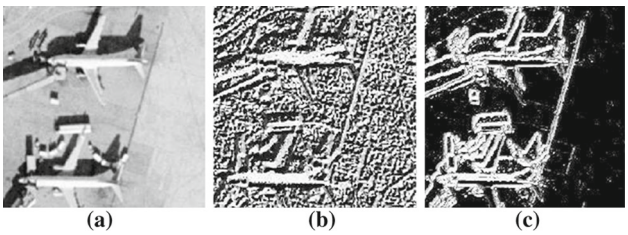

Fig. 2 **a** Input image. **b** CLBP_S coded image. **c** CLBP_M coded image

from a sample block. Note that "0" is coded as "−1" in CLBP (see Fig. 1c). Two operators, CLBP-Sign (CLBP_S) and CLBP-Magnitude (CLBP_M), are used to code these two components. CLBP_S is equivalent to the traditional LBP operator. The CLBP_M operator is defined as follows:

$$\mathrm{CLBP}_{m,r} = \sum_{i=0}^{m-1} p(|d_i|, c)2^i, \quad p(u, c) = \begin{cases} 1 & u \geq c \\ 0 & u < c \end{cases}, \tag{2}$$

where c is a threshold that is set to the mean value of $|d_i|$ from the whole image. It should be noted that there is also the CLBP-Center part which codes the values of the center pixels in the original CLBP. Here, we only consider the CLBP_S and CLBP_M operators for computational efficiency.

In Fig. 2, an example of the CLBP_S and CLBP_M coded images corresponding to an input aerial scene (airport scene) is illustrated. The pixel values of the CLBP-coded images are CLBP codes in decimal form. It can be observed that CLBP_S and CLBP_M operators both can capture the spatial pattern and the contrast of local image texture, such as edges and corners. The CLBP_S operator is able to provide more detailed texture information than the CLBP_M operator.

2.2 Two implementations of multi-scale completed local binary patterns

In [17], a multi-resolution representation based on LBP was developed to cope with the limitation that LBP features computed from a single scale may not be able to detect the dominant texture features since they characterize the image texture only at a particular resolution. This motivates us to use a multi-scale representation for CLBP as well. In this paper, we develop two implementations for the MS-CLBP.

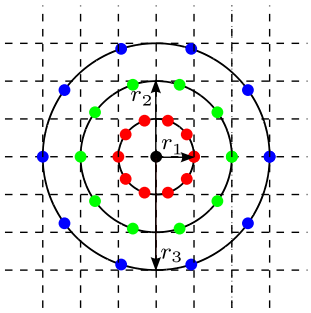

Fig. 3 An example of a 3-scale CLBP operator ($m = 10$, $r_1 = 1$, $r_2 = 2$, and $r_3 = 3$)

In the first implementation, we follow the same approach in [17] where the radius of a circle r is altered to change the spatial resolution. The multi-scale analysis is accomplished by combining the information provided by multiple operators of varying (m, r). For simplicity, we fix the number of neighbors m and tune different values of r to achieve the optimal combination. An example of a 3-scale (three r values) CLBP operator is illustrated in Fig. 3. The CLBP_S and CLBP_M histogram features extracted from each scale are concatenated to form an MS-CLBP representation. One disadvantage of this multi-scale analysis implementation is that the computational complexity increases due to multiple resolutions. For an image with size of $I_x \times I_y$ pixels, a total of $I_x \times I_y \times m$ thresholding operations are required for all the pixels in each scale. Moreover, the CLBP_S or the CLBP_M histogram is calculated based on $I_x \times I_y$ binary strings (CLBP_S codes or CLBP_M codes). Therefore, we also develop a second multi-scale analysis implementation to reduce the computational complexity.

In the second implementation, we down-sample the original image using the bicubic interpolation to obtain multiple images of different scales with respect to the original image. Figure 4 shows a down-sampled image having half of the size of the original image, i.e., scale = 1/2. The scale is between 0 and 1 with 1 being the original image. Then, the

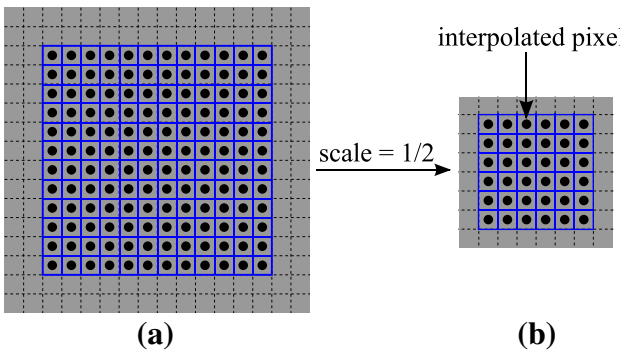

Fig. 4 **a** Original image. **b** Down-sampled image with 1/2 of the size (scale = 1/2) of the original image

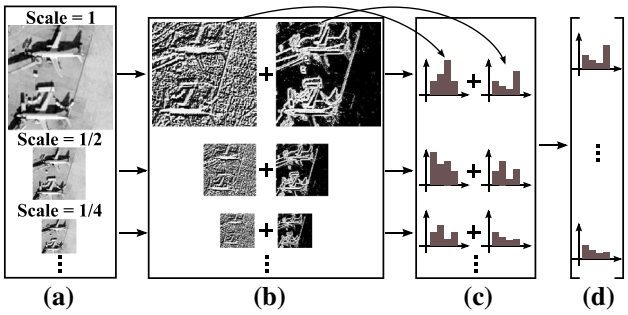

Fig. 5 MS-CLBP feature extraction method (the second implementation). **a** Images of various scales. **b** The corresponding CLBP_S (*left*) and CLBP_M (*right*) coded images. **c** Histograms calculated from the CLBP_S and CLBP_M coded images. **d** The final composite feature vector by stacking all the histograms

CLBP_S and CLBP_M operators of fixed radius and number of neighbors are applied to the images of different scales. For image of each scale, two CLBP-coded images are generated and histogram features are computed from the two images. Note that the rotation invariant pattern in CLBP is used to achieve image rotation invariance. Finally, all the histograms are concatenated to form a composite feature vector as the MS-CLBP representation for the input image. The feature extraction procedure is illustrated in Fig. 5.

Based on the two implementations, the second approach is more computationally efficient than the first one. This is because the images of different scales in the first approach all have the same size as the original image, while the images in the second approach have smaller sizes than the original image leading to much fewer pixels for the CLBP operator. This is also the reason for the scale chosen to be between 0 and 1 in the second implementation in order to facilitate computational efficiency.

3 Scene classification using kernel-based extreme learning machine

In our land-use scene classification framework, KELM is employed due to its efficient computation and good classification performance. ELM [22] is an efficient learning algorithm for single hidden layer feed-forward neural networks (SLFNs). The hidden layer weights and biases in ELM are randomly generated leading to low computational cost.

Consider n distinct training samples $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{n}$ from C classes with $\mathbf{x}_i \in \mathbb{R}^M$ and $\mathbf{y}_i \in \mathbb{R}^C$ indicating the class label of $\mathbf{x}_i$, the model of a single hidden layer neural network having L hidden nodes can be written as

$$\sum_{j=1}^{L} \boldsymbol{\beta}_j h(\mathbf{w}_j \cdot \mathbf{x}_i + e_j) = \mathbf{y}_i, \quad i = 1, \ldots, n, \tag{3}$$

where $h(\cdot)$ is a nonlinear activation function (e.g., sigmoid function), $\boldsymbol{\beta}_j \in \mathbb{R}^C$ is the output weight vector connecting the jth hidden node to the output nodes, $\mathbf{w}_j \in \mathbb{R}^M$ denotes the weight vector connecting the jth hidden node to the input nodes, and e_j is the bias of the jth hidden node. The compact form of model (3) can be rearranged using a matrix notation

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{Y}, \tag{4}$$

where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T \cdots \boldsymbol{\beta}_L^T]^T \in \mathbb{R}^{L\times C}$, $\mathbf{Y} = [\mathbf{y}_1^T \cdots \mathbf{y}_n^T]^T \in \mathbb{R}^{n\times C}$, and $\mathbf{H}$ is usually referred to as the hidden layer output matrix:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x}_1) \\ \vdots \\ \mathbf{h}(\mathbf{x}_n) \end{bmatrix} = \begin{bmatrix} h(\mathbf{w}_1 \cdot \mathbf{x}_1 + e_1) & \cdots & h(\mathbf{w}_L \cdot \mathbf{x}_1 + e_L) \\ \vdots & \ddots & \vdots \\ h(\mathbf{w}_1 \cdot \mathbf{x}_n + e_1) & \cdots & h(\mathbf{w}_L \cdot \mathbf{x}_n + e_L) \end{bmatrix}. \tag{5}$$

$\boldsymbol{\beta}$ can be estimated by a least squares solution

$$\boldsymbol{\beta}' = \mathbf{H}^{\dagger}\mathbf{Y}, \tag{6}$$

where $\mathbf{H}^{\dagger}$ denotes the pseudoinverse matrix of $\mathbf{H}$. The output function of the ELM is

$$\mathbf{f}_L(\mathbf{x}_i) = \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} = \mathbf{h}(\mathbf{x}_i)\mathbf{H}^T \left(\frac{\mathbf{I}}{\rho} + \mathbf{H}\mathbf{H}^T\right)^{-1} \mathbf{Y}, \tag{7}$$

where ρ is a regularization parameter. A kernel matrix $\Omega_{\text{ELM}} = \mathbf{H}\mathbf{H}^T : \Omega_{\text{ELM}_{i,j}} = \mathbf{h}(\mathbf{x}_i) \cdot \mathbf{h}(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$ satisfying Mercer's conditions could be used if the feature mapping $\mathbf{h}(\mathbf{x}_i)$ is unknown. Therefore, the output function of KELM is given by

$$\mathbf{f}_L(\mathbf{x}_i) = \begin{bmatrix} K(\mathbf{x}_i, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}_i, \mathbf{x}_n) \end{bmatrix}^T \left(\frac{\mathbf{I}}{\rho} + \Omega_{\text{ELM}}\right)^{-1} \mathbf{Y}. \tag{8}$$

The label of a test sample is determined according to the index of the output nodes with the largest value.

4 Experiments

We now demonstrate the effectiveness of the proposed feature extraction method for remote sensing land-use scene classification using two standard public domain datasets. In our experiments, the radial basis function (RBF) kernel is employed in KELM. The classifier parameters (RBF kernel parameter and ELM regularization parameter) are obtained via cross-validation using the training data. The classification performance of the proposed method is compared with the state-of-the-art performance reported in the literature.

4.1 Experimental data and setup

The first dataset is the 21-class land-use dataset with ground truth labeling [3]. The dataset consists of images of 21 land-use classes, and each class contains 100 images with sizes of 256 × 256 pixels. This is a challenging dataset due to a variety of spatial patterns in those 21 classes. Sample images of each land-use class are shown in Fig. 6. To facilitate a fair comparison, the same experimental setting reported in [3] is followed. Fivefold cross-validation is performed in which the dataset is randomly partitioned into five equal subsets. There are 20 images from each land-use class in a subset. Four subsets are used for training, and the remaining subset is used for testing. The classification accuracy is the average over the five cross-validation evaluations.

The second dataset used in our experiments is the 19-class satellite scene dataset [23]. It consists of 19 classes of high-resolution satellite scenes. There are 50 images with sizes of 600 × 600 pixels for each class. An example of each class is shown in Fig. 7. The same experimental setup in [24] is used. We randomly select 30 images per class as training data and the remaining images as testing data. The experiment is repeated 10 times with different realizations of randomly selected training and testing images, and the classification accuracy is averaged over the 10 repeated trails.

Fig. 6 Example images from the 21-class land-use dataset

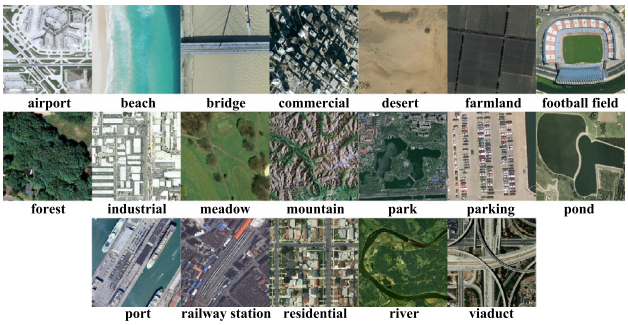

Fig. 7 Example images from the 19-class satellite scene dataset

4.2 Parameters selection

In the proposed feature extraction method, the number of scales and (m, r) of the CLBP operator are important parameters. First of all, we estimate the optimal parameter set (m, r) for the CLBP operator. For the 21-class land-use dataset, we randomly select four subsets for training and the remaining subset for testing. For the 19-class satellite scene dataset, 30 images per class are randomly selected for training and the remaining images for testing. Since the images in the two datasets are color images, we convert the images from the RGB color space to the YCbCr color space and use the Y component (luminance) to obtain the grayscale images as the original images. In this parameter tuning experiment, we apply the CLBP operator only to the original images to extract features. The classification results with various CLBP parameter sets are listed in Tables 1 and 2 for the two datasets, respectively.

Since the dimensionality of the CLBP histogram features is dependent on the number of neighbors (m), larger m will increase the feature dimensionality and computational complexity. Based on the results in Tables 1 and 2, we choose $(m, r) = (10, 3)$ for the 21-class land-use dataset and $(m, r) = (12, 4)$ for the 19-class satellite scene dataset in terms of classification accuracy and computational complexity, making the dimensionalities of the CLBP features (CLBP_S and CLBP_M histograms combined) for the 21-class land-use dataset and the 19-class satellite scene dataset 216 and 704, respectively.

Table 1 Classification accuracy (%) of CLBP with different parameters (m, r) on the 21-class land-use dataset

21-Class land-use dataset						
r	1	2	3	4	5	6
$m = 4$	69.05	70.71	72.38	69.52	68.33	65.24
$m = 6$	77.62	78.57	78.10	77.86	75.71	72.38
$m = 8$	80.24	82.62	83.57	83.10	80.95	78.33
$m = 10$	81.19	85.24	85.37	84.05	81.67	79.05
$m = 12$	81.43	85.24	84.52	85.48	83.10	82.62

Table 2 Classification accuracy (%) of CLBP with different parameters (m, r) on the 19-class satellite scene dataset

19-Class satellite scene dataset						
r	1	2	3	4	5	6
$m = 4$	75.79	75.79	75.79	75.53	73.68	70.79
$m = 6$	78.68	82.11	82.63	83.95	81.84	81.58
$m = 8$	81.84	87.11	87.11	88.42	88.68	86.05
$m = 10$	81.58	84.74	87.89	89.21	87.37	85.26
$m = 12$	82.37	88.68	89.47	92.37	90.00	87.63

Then, we study the number of scales for the MS-CLBP feature extraction method. In the first implementation of MS-CLBP, the number of neighbors m is fixed based on previously obtained tuning results, and different choices of multiple radii are examined. Specifically, 10 choices of multiple radii for MS-CLBP are {[1], [1:2], [1:3], . . . , [1:10]}, where [a:b] indicates $\{a, a+1, a+2, \ldots, b\}$. In the second implementation, parameter set (m, r) is fixed, 10 choices of multiple scales including {[1], 1/[1:2], 1/[1:3], . . . , 1/[1:10]} are considered. For example, 1/[1:2] indicates that scale = 1 (original image) and scale = 1/2 (down-sampled image with half of the size of the original image) are used in MS-CLBP. Figure 8 presents the classification results of two implementations of MS-CLBP using different numbers of scales (or radii) for the two datasets. It can be seen that the accuracy tends to be stable when 8 or more radii are used in the first implementation of MS-CLBP and 6 or more scales are used in the second implementation for the 21-class land-use dataset. More radii or scales will improve the accuracy but will also increase the dimensionality of MS-CLBP features and computational complexity. Therefore, for the 21-class land-use dataset, we choose 8 radii (i.e., r = [1:8]) for the first implementation of MS-CLBP and 6 scales (i.e., scale = 1/[1:6]) for the second implementation. Similarly, 8 radii and 4 scales, respectively, corresponding to the first and second implementations of MS-CLBP are chosen for the 19-class satellite scene dataset.

The dimensionality of the MS-CLBP features can be fairly high, e.g., it is 1728 for the 21-class land-use dataset using the second implementation with 8 scales, if the number of scales used for MS-CLBP is large. To gain computational efficiency,

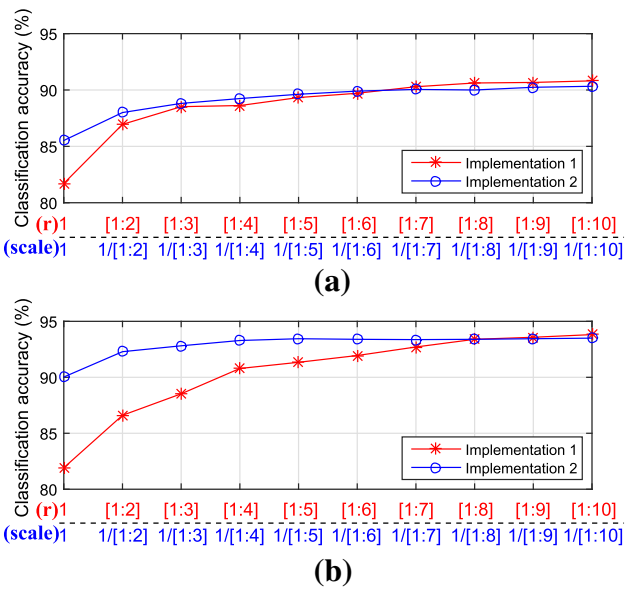

Fig. 8 Classification performance of two implementations of MS-CLBP using different numbers of scales (or radii) for **a** the 21-class land-use dataset and **b** the 19-class satellite scene dataset

principal component analysis (PCA) [25] is employed to reduce the dimensionality of MS-CLBP features. The PCA projection matrix was calculated using the features of the training data, and the principal components that accounted for 90 % of the total variation of the training features are considered in our experiments.

4.3 Comparison with the state-of-the-art methods

To evaluate the effectiveness of the proposed MS-CLBP feature extraction method, a comparison of its performance with previously reported performance in the literature is carried out on the 21-class land-use dataset under the same experimental setup (i.e., 80 % of the images from each class are used for training and the remaining images are used for testing in a fivefold cross-validation test). The first and second implementations of MS-CLBP are denoted as MS-CLBP1 and MS-CLBP2. The method applying CLBP operator to the original images (single scale) is denoted as CLBP. In addition, to demonstrate the enhanced discriminative power of fusion of the sign and magnitude components of CLBP, four methods involve only the CLBP_S operator or the CLBP_M operator (i.e., MS-CLBP_S1, MS-CLBP_S2, MS-CLBP_M1 and MS-CLBP_M2) are employed for comparison.

The comparison results reported in Table 3 show that our methods (MS-CLBP1 and MS-CLBP2) outperform the other methods. Especially, our methods achieve better performance than the popular BOVW classification framework, which demonstrates the effectiveness of the MS-CLBP approach for remote sensing land-use scene classification. The proposed methods have more than 4 % improvement over the CLBP method since the multi-scale analysis is able to capture texture and structure features of images at various resolutions. Furthermore, fusion of the sign and magnitude components of CLBP leads to a considerable classification improvement over the situation when the sign component (MS-CLBP_S) or the magnitude component (MS-CLBP_M) is used alone. It is also found that MS-CLBP1 performs a little better than MS-CLBP2 in terms of classification accuracy. This is probably because the down-sampling procedure in the second MS-CLBP implementation may smooth the images and lose some detailed texture and structure information. We also present the per-class classification accuracies for MS-CLBP1 and MS-CLBP2 in Fig. 9. We can see that the classification accuracies of the two MS-CLBP implementations are similar for most of the classes.

The comparison results for the 19-class satellite scene dataset are listed in Table 4. Although the multiple features fusion method described in [24] achieved slightly higher classification accuracy than our methods, three different sets of features including SIFT features, local ternary pattern histogram Fourier (LTP-HF) features and color histogram features were used in their approach. This multiple features

Table 3 Comparison of classification accuracy (mean ± SD) on the 21-class land-use dataset

Method	Accuracy (%)
BOVW [3]	76.8
SPM [3]	75.3
BOVW + spatial co-occurrence kernel [3]	77.7
Color Gabor [3]	80.5
Color histogram [3]	81.2
Structural texture similarity [10]	86.0
Wavelet BOVW [4]	87.4 ± 1.3
Unsupervised feature learning [12]	81.7 ± 1.2
Saliency-guided feature learning [26]	82.7 ± 1.2
Concentric circle-structured BOVW [7]	86.6 ± 0.8
Multifeature concatenation [27]	89.5 ± 0.8
Pyramid-of-spatial-relatons [8]	89.1
CLBP	85.5 ± 1.9
MS-CLBP_S1	87.4 ± 1.7
MS-CLBP_S2	83.3 ± 2.2
MS-CLBP_M1	85.7 ± 2.0
MS-CLBP_M2	83.0 ± 2.4
MS-CLBP1	**90.6 ± 1.4**
MS-CLBP2	89.9 ± 2.1

Bold values indicate the highest classification accuracy

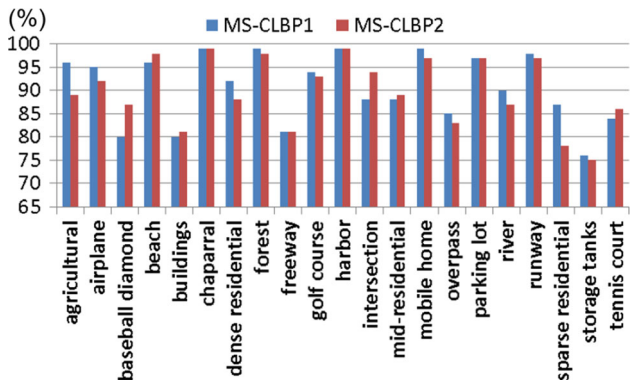

Fig. 9 Per-class accuracies for the 21-class dataset

Table 4 Comparison of classification accuracy (mean ± SD) on the 19-class satellite scene dataset

Method	Accuracy (%)
Bag of colors [27]	70.6 ± 1.5
Tree of c-shapes [27]	80.4 ± 1.8
Bag of SIFT [27]	85.5 ± 1.2
Multifeature concatenation [27]	90.8 ± 0.7
LTP-HF [24]	77.6
SIFT + LTP-HF + color histogram [24]	**93.6**
CLBP	90.0 ± 1.3
MS-CLBP_S1	88.3 ± 2.0
MS-CLBP_S2	84.8 ± 1.9
MS-CLBP_M1	89.4 ± 1.6
MS-CLBP_M2	89.1 ± 2.0
MS-CLBP1	93.4 ± 1.1
MS-CLBP2	93.3 ± 0.8

Bold values indicate the highest classification accuracy

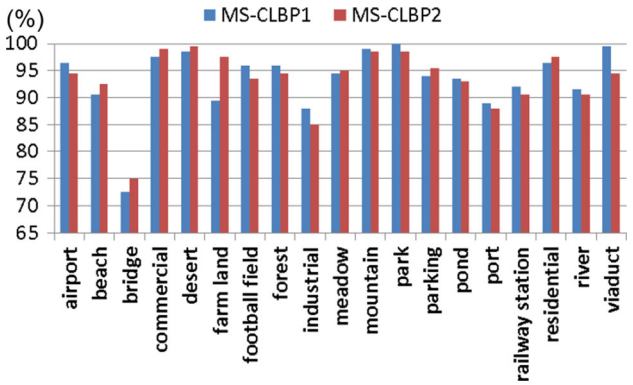

Fig. 10 Per-class accuracies for the 19-class dataset

Table 5 Processing time (ms) of the two implementations of MS-CLBP

Method	Time (mean ± SD)
MS-CLBP1	96.1 ± 6.3
MS-CLBP2	32.7 ± 5.2

fusion strategy not only requires parameter tuning for each feature but also increases the computational complexity of feature extraction. On the other hand, our proposed feature extraction method relies on a single feature descriptor (CLBP descriptor) to achieve discriminative feature representation of remote sensing land-use scene images. The per-class classification accuracies for MS-CLBP1 and MS-CLBP2 are shown in Fig. 10.

4.4 Computational complexity

Finally, we compare the computational complexity of two implementations of the proposed MS-CLBP feature extraction method. We use the 21-class land-use dataset to obtain the processing time for both implementations. To be a fair comparison, the same number of scales (6 scales) is used for MS-CLBP1 and MS-CLBP2. We run the two implementations on 2100 images (size of 256 × 256 pixels), and the processing time is the average over all the images. Our code is written in MATLAB, and the time reported in Table 5 is for an Intel i7 Quadcore 3.4 GHz desktop computer with 8 GB of RAM. As we can see that the second implementation of MS-CLBP is almost three times faster than the first implementation. It is worth mentioning that the proposed

MS-CLBP descriptor can be done in parallel to achieve high computational efficiency.

5 Conclusion

In this paper, we proposed an effective feature extraction method based on multi-scale completed local binary patterns (MS-CLBP) for land-use scene classification. Two implementations of the proposed MS-CLBP were presented and compared in terms of classification accuracy and computational complexity. We validated the proposed feature extraction method on two standard land-use scene datasets and quantitatively showed the excellence of the proposed method over some state-of-the-art land-use scene classification methods.

References

1. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Proceedings of ECCV Workshop on Statistical Learning in Computer Vision, Prague, pp. 1–2 (2004)
2. Yang, Y., Newsam, S.: Spatial pyramid co-occurrence for image classification. In: Proceedings of International Conference on Computer Vision, Barcelona, Spain, pp. 1465–1472 (2011)
3. Yang, Y., Newsam, S.: Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of the 18th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, pp. 270–279 (2010)
4. Zhao, L., Tang, P., Huo, L.: A 2-D wavelet decomposition-based bag-of-visual-words model for land-use scene classification. Int. J. Remote Sens. **35**(6), 2296–2310 (2014)
5. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the IEEE Computer Vision and Pattern Recognition, New York, NY, pp. 2169–2178 (2006)
6. Zhou, L., Zhou, Z., Hu, D.: Scene classification using a multi-resolution bag-of-features model. Pattern Recognit. **46**(1), 424–433 (2013)
7. Zhao, L., Tang, P., Huo, L.: Land-use scene classification using a concentric circle-structured multiscale bag-of-visual-words model. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **7**(12), 4620–4631 (2014)
8. Chen, S., Tian, Y.: Pyramid of spatial relatons for scene-level land use classification. IEEE Trans. Geosci. Remote Sens. **53**(4), 1947–1957 (2015)
9. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
10. Risojevic, V., Babic, Z.: Aerial image classification using structural texture similarity. In: Proceedings of IEEE International Symposium on Signal Processing and Information Technology, Bilbao, Spain, pp. 190–195 (2011)
11. Risojevic, V., Babic, Z.: Fusion of global and local descriptors for remote sensing image classification. IEEE Geosci. Remote Sens. Lett. **10**(4), 836–840 (2013)
12. Cheriyadat, A.M.: Unsupervised feature learning for aerial scene classification. IEEE Trans. Geosci. Remote Sens. **52**(1), 439–451 (2014)
13. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comput. Vis. **42**(3), 145–175 (2001)
14. Tola, E., Lepetit, V., Fua, P.: Daisy: an efficient dense descriptor applied to wide-baseline stereo. IEEE Trans. Pattern Anal. Mach. Intell. **32**(5), 815–830 (2010)
15. Zheng, X., Sun, X., Fu, K., Wang, H.: Automatic annotation of satellite images via multifeature joint sparse coding with spatial relation constraint. IEEE Geosci. Remote Sens. Lett. **10**(4), 652–656 (2013)
16. Zhang, Y., Zheng, X., Liu, G., Sun, X., Wang, H., Fu, K.: Semi-supervised manifold learning based multigraph fusion for high-resolution remote sensing image classification. IEEE Geosci. Remote Sens. Lett. **11**(2), 464–468 (2014)
17. Ojala, T., Pietikainen, M., Maenpaa, T.T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
18. Li, W., Chen, C., Su, H., Du, Q.: Local binary patterns and extreme learning machine for hyperspectral imagery classification. IEEE Trans. Geosci. Remote Sens. **53**(7), 3681–3693 (2015)
19. Chen, C., Jafari, R., Kehtarnavaz, N.: Action Recognition from depth sequences using depth motion maps-based local binary patterns. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Waikoloa Beach, HI, pp. 1092–1099 (2015)
20. Guo, Z., Zhang, L., Zhang, D.: A completed modeling of local binary pattern operator for texture classification. IEEE Trans. Image Process. **19**(6), 1657–1663 (2010)
21. Huang, G.B., Zhou, H., Ding, X., Zhang, R.: Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cybern. Part B Cybern. **42**(2), 513–529 (2012)
22. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: theory and applications. Neurocomputing **70**(1–3), 489–501 (2006)
23. Dai, D., Yang, W.: Satellite image classification via two-layer sparse coding with biased image representation. IEEE Geosci. Remote Sens. Lett. **8**(1), 173176 (2011)
24. Sheng, G., Yang, W., Xu, T., Sun, H.: High-resolution satellite scene classification using a sparse coding based multiple feature combination. Int. J. Remote Sens. **33**(8), 2395–2412 (2011)
25. Jolliffe, I.: Principal Component Analysis, 2nd edn. Springer, New York (2002)
26. Zhang, F., Du, B., Zhang, L.: Saliency-guided unsupervised feature learning for scene classification. IEEE Trans. Geosci. Remote Sens. **53**(4), 2175–2184 (2015)
27. Shao, W., Yang, W., Xia, G.S., Liu, G.: A hierarchical scheme of multiple feature fusion for high-resolution satellite scene categorization. In: Proceedings of the 9th International Conference on Computer Vision Systems, St. Petersburg, Russia, pp. 324–333 (2013)