Gabor-Filtering-Based Completed Local Binary Patterns for Land-Use Scene Classification

Chen Chen[1], Libing Zhou[2,*], Jianzhong Guo[1,2], Wei Li[3], Hongjun Su[4], Fangda Guo[5]

[1]Department of Electrical Engineering, University of Texas at Dallas, TX, USA (E-mail: chenchen870713@gmail.com)
[2]School of Electronic and Electrical Engineering, Wuhan Textile University, Wuhan, China
[3]College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China
[4]School of Earth Sciences and Engineering, Hohai University, Nanjing, China
[5]Department of Electrical, Computer, and Biomedical Engineering, University of Pavia, Pavia, Italy

Abstract—**Remote sensing land-use scene classification has a wide range of applications including forestry, urban-growth analysis, and weather forecasting. This paper presents an effective image representation method, Gabor-filtering-based completed local binary patterns (GCLBP), for land-use scene classification. It employs the multi-orientation Gabor filters to capture the global texture information from an input image. Then, a local operator called completed local binary patterns (CLBP) is utilized to extract the local texture features, such as edges and corners, from the Gabor feature images and the input image. The resulting CLBP histogram features are concatenated to represent an input image. Experimental results on two datasets demonstrate that the proposed method is superior to several existing methods for land-use scene classification.**

Keywords-Gabor filtering; local binary patterns; land-use sence classification; extreme learning machine

I. INTRODUCTION

Land-use scene classification aims to assign semantic labels (e.g., building, river, forest, mountain, etc.) to aerial or satellite images. It has a wide range of applications including agricultural planning, forestry, urban-growth analysis, and land use management. With the rapid development in sensor technology, high-resolution remote sensing images can be obtained using the advanced space-borne sensors. High-resolution remote sensing images with rich spatial and texture information have made it possible to categorize different land-use scene classes automatically [1].

There has been a great deal of effort in employing computer vision techniques for classifying aerial or satellite images. The Bag-of-Words (BoW) model [2] is one of the most popular approaches in image classification and image retrieval applications. In the BoW model, local image features such as color and texture are first quantized into a set of visual words using some clustering methods. An image is then represented by frequencies of the set of visual words. Although the BoW model has demonstrated the effectiveness for the remotely sensed land-use scene classification [1, 3], it ignores spatial relationships of the local features. To incorporate spatial context to the BoW model, a spatial pyramid matching (SPM) framework was proposed in [4] by partitioning an image into subregions and computing a BoW histogram for each subregion. Histograms from all subregions were concatenated to form the SPM representation of an image. In [5], a multi-resolution representation was incorporated into the BoW model to improve the SPM framework by constructing multiple resolution images and extracting local features from all the resolution images with dense regions. Since the SPM method uses the absolute spatial information, it may not improve the classification

performance for images exhibit rotation and translation variations due to rotated camera views. To overcome this limitation, a pyramid-of-spatial-relatons (PSR) model was proposed in [6] to capture both absolute and relative spatial relationships of local features.

The above-mentioned methods focused on improving the BoW framework by incorporating spatial information for land-use scene classification; however, extracting effective local features that can capture the rich texture information of the high-resolution remote sensing images was not exploited. On the other hand, some works evaluated various image feature descriptors and combinations of feature descriptors for scene classification. In [7], local structural texture descriptors and structural texture similarity with nearest neighbor classifier were utilized for semantic classification of aerial images. In [8], Gabor descriptor and Gist descriptor were evaluated individually for the task of aerial image classification. In [9], a global feature descriptor named enhanced Gabor texture descriptor (EGTD) and a local scale-invariant feature transform (SIFT) [10] descriptor were combined in a hierarchical approach to improve the remote sensing image classification performance. In [11], four types of features consist of DAISY [12], geometric blur [13], SIFT [10], and self-similarity [14] were used within the framework of multifeature joint sparse coding with spatial relation constraint. Although fusing a set of different features may enhance the discriminative power, it requires parameter tuning for each feature and the feature dimensionality may be increased significantly.

Gabor filters [15] and local binary patterns (LBP) [16] have been successfully applied for a variety of image processing and machine vision applications (e.g., [17-19]). In this paper, we present an efficient image representation method using Gabor-filtering-based completed local binary patterns (GCLBP). More specifically, multi-orientation Gabor filters are first applied to a remotely sensed input image to obtain multiple Gabor feature images which capture different orientation information of the input image. Completed local binary patterns (CLBP) [20] operator, a complete modeling of the LBP operator, is then employed to extract the rotation invariant texture features (histograms) from the Gabor feature images as well as the input image. The overall framework of the proposed representation approach is illustrated in Fig. 1. For classification, kernel-based extreme learning machine (KELM) [21] is utilized due to its efficient computation and good classification performance.

The remainder of this paper is organized as follows. Section II provides relevant background and related work. Section III describes the details of the proposed image representation approach. Section IV presents the experimental data and setup as well as comparison of the classification performance between the proposed method and the existing methods. Finally, Section V makes several concluding remarks.
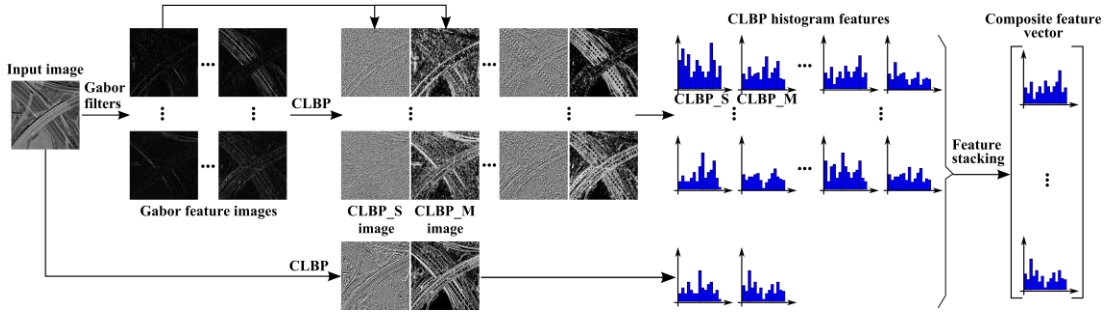
Fig. 1. The framework of the proposed GCLBP image representation approach.

II. RELATED WORK

A. Gabor Filtering

A Gabor wavelet is a filter whose impulse response is defined by a sinusoidal wave multiplied by a Gaussian function. In the 2-D spatial domain, a Gabor filter, including a real component and an imaginary term, can be represented as

$$G_{\delta,\theta,\psi,\sigma,\gamma}(a,b) = \exp\left(-\frac{a'^2 + \gamma^2 b'^2}{2\sigma^2}\right)\exp\left(j\left(2\pi\frac{a'}{\delta}+\psi\right)\right) \quad (1)$$

where

$$a' = a\cos\theta + b\sin\theta \quad (2)$$

$$b' = -a\sin\theta + b\cos\theta. \quad (3)$$

Here, a and b denote the pixel positions, δ represents the wavelength of the sinusoidal factor, θ represents the orientation of the Gabor wavelet (e.g., $\pi/8$, $\pi/4$, $\pi/2$, etc.). Note that we only need to consider $\theta \in [0^o, 180^o]$ since symmetry makes other directions redundant. ψ is the phase offset and γ is the spatial aspect ratio (the default value is 0.5 [17, 18]) specifying the ellipticity of the support of the Gabor function. $\psi = 0$ and $\psi = \pi/2$ return the real and imaginary parts of the Gabor filter, respectively. Parameter σ is the standard deviation of the Gaussian function and it is determined by δ and spatial frequency bandwidth bw as

$$\sigma = \frac{\delta}{\pi}\sqrt{\frac{\ln 2}{2}}\frac{2^{bw}+1}{2^{bw}-1}. \quad (4)$$

A visualization of Gabor filters for four orientations is presented in Fig. 2.
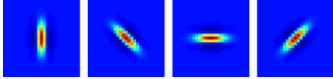

Fig. 2. Two-dimensional Gabor kernels with four orientations, from left to right: 0, $\pi/4$, $\pi/2$, and $3\pi/4$.

Typically, the Gabor texture feature image in a specific orientation is the magnitude part of convolving the input image with the Gabor function $G(a,b)$.

B. CLBP

LBP [16] is a simple yet efficient operator to summarize local gray-level structure of an image. 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:

$$LBP_{m,r}(t_c) = \sum_{i=0}^{m-1}s(t_i - t_c)2^i = \sum_{i=0}^{m-1}s(d_i)2^i, \quad (5)$$

where $d_i = (t_i - t_c)$ is the difference between the center pixel and each neighbor, $s(d_i) = 1$ 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 scheme, the image local differences are decomposed into two complementary components: the signs and magnitudes (absolute values of d_i, i.e., $|d_i|$). Fig. 3 shows an example of the sign and magnitude components of the CLBP extracted from a sample block. Note that "0" is coded as "-1" in CLBP [see Fig. 3 (c)]. Two operators, namely 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:

$$CLBP\_M_{m,r} = \sum_{i=0}^{m-1}p(|d_i|,c)2^i, \quad p(u,c)=\begin{cases}1, & u \geq c \\ 0, & u < c\end{cases} \quad (6)$$

where c is a threshold that is set to the mean value of $|d_i|$ from the whole image. The CLBP-Center part which codes the values of the center pixels is not used here.

93	61	73		41	9	21		1	1	1		41	9	21
83	**52**	56		31		4		1		1		31		4
31	46	43		-21	-6	-9		-1	-1	-1		21	6	9
	(a)				(b)				(c)				(d)	

Fig. 3. (a) 3×3 sample block; (b) the local differences; (c) the sign component of CLBP; and (d) the magnitude component of CLBP.

C. Extreme Learning Machine (ELM)

ELM [22] is an efficient learning algorithm for single-hidden-layer feed-forward neural networks (SLFNs). The hidden node parameters in ELM are randomly generated leading to a much faster learning rate.

Let $\mathbf{y} = [y_1,...,y_k,...,y_C]^T \in \mathfrak{R}^C$ be the class to which a sample belongs, where $y_k \in \{1,-1\}$ ($1 \leq k \leq C$) and C is the number of classes. Given n training samples $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^n$, where $\mathbf{x}_i \in \mathfrak{R}^M$ and $\mathbf{y}_i \in \mathfrak{R}^C$, the model of a single hidden layer neural network having L hidden nodes can be expressed as

$$\sum_{j=1}^{L} \boldsymbol{\beta}_j h\left(\mathbf{w}_j \cdot \mathbf{x}_i + e_j\right) = \mathbf{y}_i, \quad i = 1,...,n, \tag{7}$$

where $h(\cdot)$ is a nonlinear activation function (e.g., Sigmoid function), $\boldsymbol{\beta}_j \in \mathfrak{R}^C$ denotes the weight vector connecting the jth hidden node to the output nodes, $\mathbf{w}_j \in \mathfrak{R}^M$ 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 above n equations can be written compactly as:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{Y}, \tag{8}$$

where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T;...;\boldsymbol{\beta}_L^T] \in \mathfrak{R}^{L \times C}$, $\mathbf{Y} = [\mathbf{y}_1^T;...;\mathbf{y}_n^T] \in \mathfrak{R}^{n \times C}$, and $\mathbf{H}$ is the hidden layer output matrix of the neural network expressed as

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x}_1) \\ \vdots \\ \mathbf{h}(\mathbf{x}_n) \end{bmatrix} = \begin{bmatrix} h(\mathbf{w}_1 \cdot \mathbf{x}_1 + e_1) & \cdots & h(\mathbf{w}_L \cdot \mathbf{x}_1 + e_L) \\ \vdots & \ddots & \vdots \\ h(\mathbf{w}_1 \cdot \mathbf{x}_n + e_1) & \cdots & h(\mathbf{w}_L \cdot \mathbf{x}_n + e_L) \end{bmatrix}_{n \times L}. \tag{9}$$

$\mathbf{h}(\mathbf{x}_i) = [h(\mathbf{w}_1 \cdot \mathbf{x}_i + e_1),...,h(\mathbf{w}_L \cdot \mathbf{x}_i + e_L)]$ is the output of the hidden nodes in response to the input $\mathbf{x}_i$. A least-squares solution $\hat{\boldsymbol{\beta}}$ of the linear system $\mathbf{H}\boldsymbol{\beta} = \mathbf{Y}$ is found to be

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^\dagger \mathbf{Y}, \tag{10}$$

where $\mathbf{H}^\dagger$ is the *Moore-Penrose generalized inverse* of matrix $\mathbf{H}$. As a result, the output function of the ELM classifier can be expressed as

$$\mathbf{f}_L(\mathbf{x}_i) = \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} = \mathbf{h}(\mathbf{x}_i)\mathbf{H}^T \left(\frac{\mathbf{I}}{\rho} + \mathbf{H}\mathbf{H}^T\right)^{-1} \mathbf{Y}, \tag{11}$$

where $1/\rho$ is a regularization term. A kernel matrix $\Omega_{ELM} = \mathbf{h}(\mathbf{x}_i) \cdot \mathbf{h}(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$ is considered if the feature mapping $\mathbf{h}(\mathbf{x}_i)$ is unknown. Therefore, the output function of KELM is given by

$$\mathbf{f}_L(\mathbf{x}_i) = \begin{bmatrix} K(\mathbf{x}_i, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}_i, \mathbf{x}_n) \end{bmatrix}^T \left(\frac{\mathbf{I}}{\rho} + \Omega_{ELM}\right)^{-1} \mathbf{Y}. \tag{12}$$

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

III. PROPOSED IMAGE REPRESENTATION APPROACH

Inspired by the success of Gabor filters and LBP in computer vision applications, we propose an efficient image representation approach for land-use scene classification using Gabor-filtering-based CLBP. The Gabor filter belongs to a global operator while LBP is a local one. As a consequence, Gabor features and LBP features represent texture information from different perspectives.

An input land-use scene image is first convolved with the Gabor filters with different orientations to generate the Gabor-filtered images. The magnitudes of the Gabor-filtered images are used as the Gabor texture feature images. Fig. 4(b) - (e) are the Gabor feature images obtained by the Gabor filters with four orientations ($\theta=0$, $\theta=\pi/4$, $\theta=\pi/2$, and $\theta=3\pi/4$). As we can see the Gabor feature images reflect the global signal power in different orientations. In order to enhance the information in the Gabor feature images, we encode the Gabor feature images with the CLBP operator (i.e., CLBP_S and

CLBP_M). Each Gabor feature image results in one CLBP_S coded image (equivalent to an LBP coded image) and one CLBP_M coded image. Fig. 4 (a1) - (e1) show the CLBP_S coded images for the Gabor feature images and Fig. 4 (a2) - (e2) show the CLBP_M coded images for the Gabor feature images. It is obvious that the detailed local spatial texture features, such as edges, corners, and knots, are enhanced in the CLBP_S and CLBP_M coded images. Moreover, CLBP_S and CLBP_M coded images contain complementary texture information which motivates us to use CLBP in our image representation method to enhance the discriminative power. The CLBP operator is also applied to the input image. Histogram is computed from each CLBP_S and CLBP_M coded images. Finally, all the histograms are concatenated or stacked as a composite feature vector before it is fed into a KELM classifier. The overall framework of the proposed image representation approach (GCLBP) is illustrated in Fig. 1. Note that we use rotation invariant pattern in CLBP to achieve image rotation invariance.
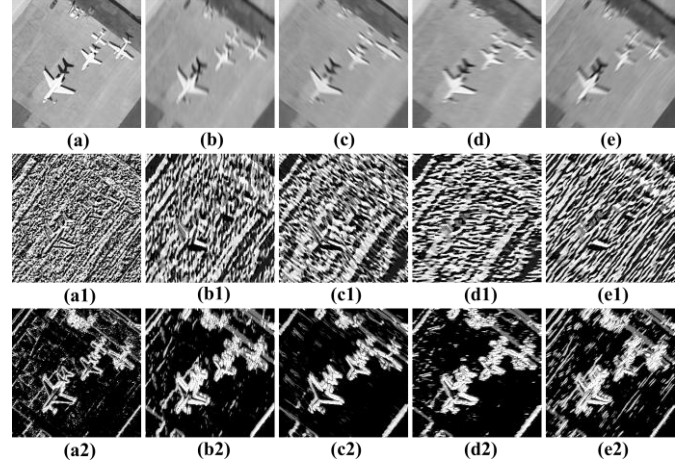

Fig. 4. Examples of Gabor feature images and the corresponding CLBP coded images. (a) Input image. (b) - (e) are the Gabor feature images obtained by the Gabor filters with $\theta=0$, $\theta=\pi/4$, $\theta=\pi/2$, and $\theta=3\pi/4$ (wavelength $\delta=8$ and bandwidth $bw=4$). (a1) - (e1) are CLBP_S coded images corresponding to (a) - (e). (a2) - (e2) are CLBP_M coded images corresponding to (a) - (e). The pixel values of the CLBP_S (CLBP_M) coded images are CLBP_S (CLBP_M) codes (binary strings) in decimal form.

IV. EXPERIMENT

To evaluate the efficacy of our proposed image representation method for remote sensing land-use scene classification, we conduct experiments using two publicly available datasets. The classification performance of the proposed method is compared with the state-of-the-art performance reported in the literatures. In our experiments, the radial basis function (RBF) kernel was employed in KELM.

A. 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 selected from aerial orthoimagery. 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. 5. To facilitate a fair comparison, the same experimental setting reported in [3] was followed. Five-fold 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 collected from Google Earth (Google Inc.). There are 50 images with sizes of 600×600 pixels for each class. The images are extracted from large satellite images. An example of each class is shown in Fig. 6. The same experimental setup in [24] was 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 classification accuracy is averaged over the 10 trails.
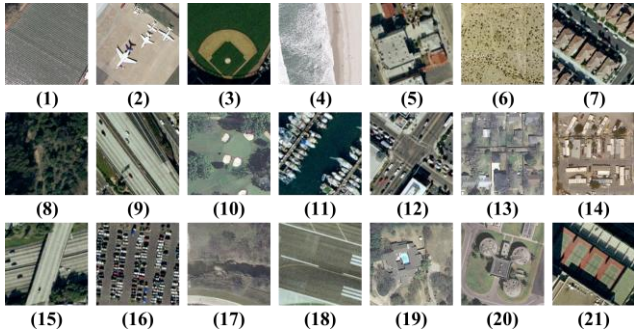

Fig. 5. Examples from the 21-class land-use dataset: (1) agricultural, (2) airplane, (3) baseball diamond, (4) beach, (5) buildings, (6) chaparral, (7) dense residential, (8) forest, (9) freeway, (10) golf course, (11) harbor, (12) intersection, (13) medium density residential, (14) mobile home park, (15) overpass, (16) parking lot, (17) river, (18) runway, (19) sparse residential, (20) storage tanks, (21) tennis courts.

Fig. 6. Examples from the 19-class satellite scene dataset: (1) airport, (2) beach, (3) bridge, (4) commercial, (5) desert, (6) farmland, (7) football field, (8) forest, (9) industrial, (10) meadow, (11) mountain, (12) park, (13) parking, (14) pond, (15) port, (16) railway station, (17) residential, (18) river, (19) viaduct.

B. Parameter Tuning

First of all, we study the Gabor filter parameters for land-use scene classification. According to (4), the parameters of Gabor filter with different δ and bw are investigated. Four Gabor orientations ($\theta=0$, $\theta=\pi/4$, $\theta=\pi/2$, and $\theta=3\pi/4$) are used. The parameters for the CLBP operator are set as: $m=10$ and $r=3$. 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. Fig. 7 and 8 show the classification results for the two datasets, respectively. From the results, the optimal δ for the 21-class land-use dataset is 8 and the optimal bw is 4. The optimal δ for the 19-class satellite scene dataset is 6 and the optimal bw is 2. Therefore, we fix these parameters in our subsequent experiments. We further examine different choices of orientations for the Gabor filter. Two orientations include $[0,\pi/2]$, four orientations include $[0,\pi/4,\pi/2,3\pi/4]$, six orientations include

$[0,\pi/6,\pi/3,\pi/2,2\pi/3,5\pi/6]$, and eight orientations include $[0,\pi/8,\pi/4,3\pi/8,\pi/2,5\pi/8,3\pi/4,7\pi/8]$. Fig. 9 illustrates the classification performance of GCLBP with different orientations. Thus, four orientations include $[0,\pi/4,\pi/2,3\pi/4]$ were chosen for the experiments.

Then, we assign appropriate values for the parameter set (m,r) of the CLBP operator. The classification results with various CLBP parameter sets are listed in Tables I and II for the two datasets, respectively. Note that the dimensionality of the CLBP histogram features is dependent on the number of neighbors (m) . Therefore, larger m will increase the feature dimensionality and computational complexity. In our experiments, we choose $(m,r)=(10,3)$ for the 21-class land-use dataset and $(m,r)=(8,3)$ for the 19-class satellite scene dataset in terms of classification accuracy and computational complexity, making the dimensionalities of the GCLBP features for the 21-class land-use dataset and the 19-class satellite scene dataset 1080 and 360, respectively. Furthermore, in all the experiments, the parameters for KELM (RBF kernel parameters) were chosen as the ones that maximized the training accuracy by means of a 5-fold cross-valiadation.
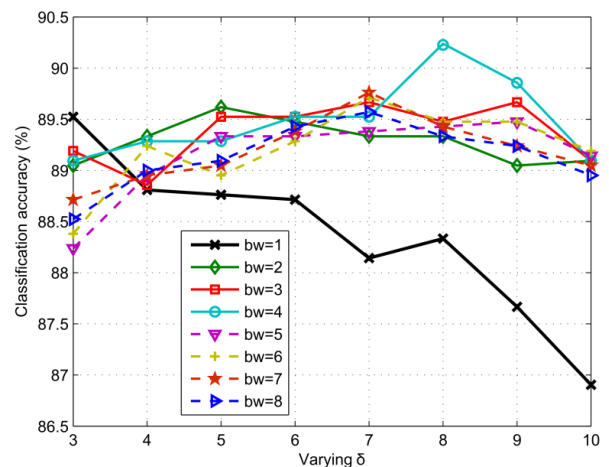

Fig. 7. Classification accuracy (%) versus varying δ and bw for the proposed GCLBP method for the 21-class land-use dataset.
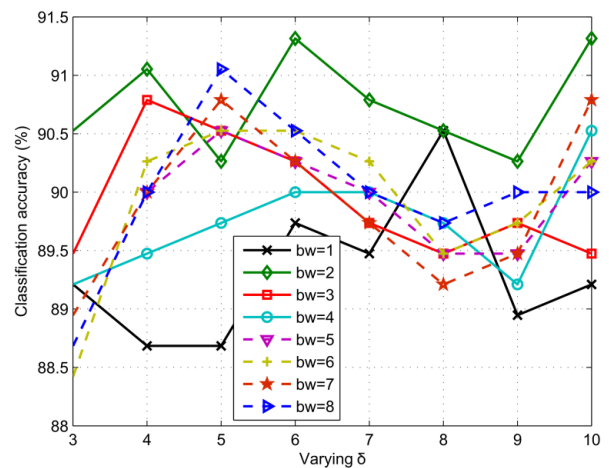

Fig. 8. Classification accuracy (%) versus varying δ and bw for the proposed GCLBP method for the 19-class satellite scene dataset.
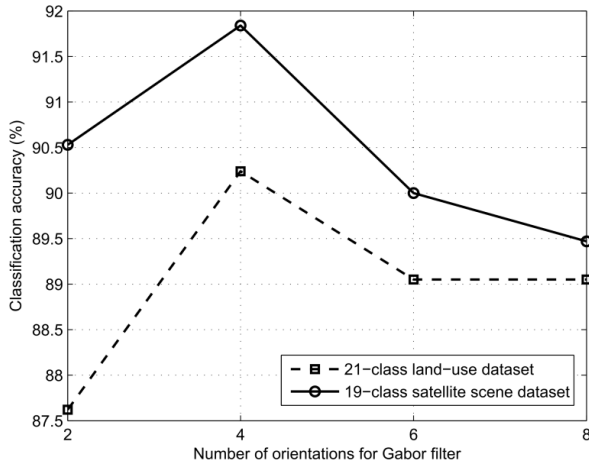
Fig. 9. Classification accuracy (%) versus different Gabor filter orientations for the proposed GCLBP.

TABLE I. CLASSIFICATION ACCURACY (%) OF GCLBP WITH DIFFERENT PARAMETERS (m, r) OF THE CLBP OPERATOR ON THE 21-CLASS LAND-USE DATASET

21-class land-use dataset					
r	1	2	3	4	5
$m=4$	85.48	85.00	83.57	82.86	80.71
$m=6$	88.81	88.10	87.14	86.67	85.71
$m=8$	89.52	89.05	89.76	88.33	87.38
$m=10$	89.05	89.29	90.24	88.10	86.67
$m=12$	89.52	89.52	90.48	90.00	88.10

TABLE II. CLASSIFICATION ACCURACY (%) OF GCLBP WITH DIFFERENT PARAMETERS (m, r) OF THE CLBP OPERATOR ON THE 19-CLASS SATELLITE SCENE DATASET

19-class satellite scene dataset					
r	1	2	3	4	5
$m=4$	85.79	89.21	87.89	87.11	85.00
$m=6$	86.84	89.47	88.68	89.21	88.68
$m=8$	89.74	90.53	91.84	90.79	90.53
$m=10$	89.21	90.26	91.32	91.58	91.05
$m=12$	89.74	91.32	91.32	91.32	91.05

C. Comparison With the State of the Art

To evalute the effectiveness of the proposed GCLBP representation method, a comparison of its performance with previsouly reported performance in the literatures was 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 as training, and the remaining images are used as testing). Since the images in the dataset 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 gray scale images. The GCLBP features are extracted from the gray scale images. We also implement the method which uses the CLBP operator on the input image only, denoted as CLBP. The comparison results are reported in Table III, which demonstrates that our method achieves superior classification performance over the other methods. Especially, our method achieved better performance than the popular BoW classification framework, which demonstrates the effectiveness of the proposed GCLBP approach for remote sensing land-use scene classification. Moreover, the proposed GCLBP has 4.5% improvement over the CLBP method since the multi-orientation Gabor filters captured the global texture information in different directions.

We also present the confusion matrix of our method for the 21-class land-use dataset in Fig. 10. For a compact representation, numbers along the x-axis and y-axis in this figure are used to indicate the land-use scene classes listed in Fig. 5. The diagonal elements of the matrix denote the mean class-specific classification accuracy (%).

TABLE III. COMPARISON OF CLASSIFICATION ACCURACY (MEAN±STD) ON THE 21-CLASS LAND-USE SCENE DATASET

Method	Accuracy (%)
BoW [3]	76.8
SPM [3]	75.3
BoW+Spatial Co-occurrence Kernel [3]	77.7
Color Gabor [3]	80.5
Color histogram (HLS) [3]	81.2
Structural texture similarity [7]	86.0
Wavelet BoW [25]	87.4
Concentric circle-structured multiscale BoW [27]	86.6
Multiple feature fusion [26]	89.5
Pyramid-of-Spatial-Relatons (PSR) [6]	89.1
CLBP	85.5
Ours (GCLBP)	**90.0±2.1**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	88				1		1	1	2	2		1	1						2		1
2		94	2	1								1						1		1	
3		3	80					6		1					1			6	3		
4				99					1												
5		4	1		80		2			3	2	2							5	1	
6						99	1														
7					1		88			2	5	1					1	1	1		
8								99										1			
9			1						94		2			2					1		
10		3								91					1		5				
11							1				98					1					
12	1	1			3		1					88	1		1	1			2	1	
13	1				1		4			1		90					3				
14							1					3	96								
15		1					2			6			88				3				
16									2	1			97								
17	1		1	4			1		1					88	4						
18	1						3							95	1						
19	2	5			1	1		5		1	4		1	79			1				
20	6	3		4		1			2	1	1		1	3	76	2					
21	1	1		3		1				4				3	4	83					

Fig. 10. Confusion matrix of our method for the 21-class land-use dataset.

The comparison results for the 19-class satellite scene dataset are listed in Table IV. Although the multiple features fusion method described in [24] achieved higher classification accuracy than our method, three different sets of features including SIFT features, Local Ternary Pattern Histogram Fourier (LTP-HF) features, and color histogram features were used, thus leading to increased computational complexity. The confusion matrix of our method for the 19-class satellite scene dataset is shown in Fig. 11. The numbers along the x-axis and y-axis in this figure are used to indicate the land-use scene classes listed in Fig. 6.

TABLE IV. COMPARISON OF CLASSIFICATION ACCURACY (MEAN±STD) ON THE 19-CLASS SATELLITE SCENE DATASET

Method	Accuracy
Bag of colors [26]	70.6
Tree of c-shapes [26]	80.4
Bag of SIFT [26]	85.5
Multifeature concatenation [26]	90.8
Local Ternary Pattern Histogram Fourier (LTP-HF) [24]	77.6
SIFT+LTP-HF+Color histogram [24]	**93.6**
CLBP	86.7
Ours (GCLBP)	91.0 ±1.5

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	89.5			1.5			1		4								0.5	1	2.5
2	1	90		1.5	2.5	0.5	1.5					2						1	
3	1	1	80	0.5			0.5			3.5			0.5	7.5	3.5			2	
4				97.5			0.5					1					1		
5					95	2.5						0.5						2	
6	1	2.5			6	84				2.5	0.5	0.5	0.5			1.5			
7		1					98					1							
8								97				1						2	
9				4		0.5	5.5		86.5							0.5	2		1
10			0.5		1	2.5				94.5								1.5	
11					1						97.5							1.5	
12				1								99							
13	3		0.5						1.5			1.5	86		6.5			1	
14		1	0.5				0.5						0.5	85	5			7.5	
15	1		4.5				3.5					1	1.5	2.5	86				
16	4								1.5							86			8.5
17				6.5			1.5		2								90		
18								1.5				5.5	1					92	
19	0.5		3						1							0.5		0.5	94.5

Fig. 11. Confusion matrix of our method for the 19-class satellite scene dataset.

The dimensionality of the GCLBP features can be fairly high, e.g., it is 1080 for the 21-class land-use dataset, if a large m is used for the CLBP operator. To gain computational efficiency, dimensionality reduction techniques such as principal component analysis (PCA) [28] can be applied to the GCLBP features to reduce the dimensionality.

V. CONCLUSION

In this paper, an effective image representation method for remote sensing land-use scene classification was introduced. This representation method was derived from the Gabor filters and the completed local binary patterns (CLBP) operator. Gabor filters were employed to capture the global texture information from different directions of an input image, whereas CLBP histogram features were extracted from the Gabor feature images to enhance the texture information (e.g., edges and corners). The combination of a global operator (Gabor filters) and a local operator (CLBP) greatly enhanced the representation power of the spatial histogram. The experimental results on two datasets demonstrated that our proposed Gabor-filtering-based CLBP (GCLBP) representation method achieved superior classification performance over the existing methods for land-use scene classification.

REFERENCES

[1] Y. Yang, and S. Newsam, "Spatial pyramid co-occurrence for image classification," in *ICCV*, Barcelona, Spain, pp. 1465-1472, November 2011.

[2] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorizatoin with bags of keypoints," in *Proceedings of ECCV Workshops on Statistical Learning in Computer Vision*, 2004.

[3] Y. Yang, and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings ofthe 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, San Jose, CA, pp. 270-279, November 2010.

[4] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *CVPR*, New York, NY, pp. 2169-2178, June 2006.

[5] L. Zhou, Z. Zhou, and D. Hu, "Scene classification using a multi-resolution bag-of-features model," *Pattern Recognition*, vol. 46, no. 1, pp. 424-433, January 2013.

[6] S. Chen, and Y. Tian, "Pyramid of spatial relatons for scene-level land use classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 1947-1957, April 2015.

[7] V. Risojevic, and Z. Babic, "Aerial image classification using structureal texture similarity," in *Proceedings of IEEE International Symposium on Signal Processing and Information Technology*, Bilbao, Spain, pp. 190-195, December 2011.

[8] V. Risojevic, S. Momic, and Z. Babic, "Gabor descriptors for aerial image classification," in *Proceedings of 10th International Conference on Adaptive and Natural Computing Algorithms*, Ljubljana, Slovenia, pp. 51-60, April 2011.

[9] V. Risojevic, and Z. Babic, "Fusion of global and local descriptors for remote sensing image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 836-840, July 2013.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, November 2004.

[11] X. Zheng, X. Sun, K. Fu, and H. Wang, "Automatic annotation of satellite images via multifeature joint sparse coding with spatial relation constraint," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 652-656, July 2013.

[12] E. Tola, V. Lepetit, and P. Fua, "A fast local descirptor for dense matching," in *CVPR*, Anchorage, AK, pp. 1-8, June 2008.

[13] A. C. Berg, and J. Malik, "Geometric blur for template matching," in *CVPR*, Kauai, HI, vol. 1, pp. I-607-I-614, December 2001.

[14] E. Shechtman, and M. Irani, "Matching local self-similarities across images and videos," in *CVPR*, Minneapolis, MN, pp. 1-8, June 2007.

[15] I. Fogel, and D. Sagi, "Gabor fitlers as texture discriminator," *Biological Cybernetics*, vol. 61, no. 2, pp. 103-113, June1989.

[16] T. Ojala, M. Pietikainen, and T. T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, July 2002.

[17] C. Chen, W. Li, H. Su, and K. Liu, "Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine," *Remote Sensing*, vol. 6, no. 6, pp. 5795-5814, June 2014.

[18] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns for spatial-spectral classification of hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3681-3693, July 2015.

[19] C. Chen, R. Jafari, and N. Kehtarnavaz, "Action recognition from depth sequences using depth motion maps-based local binary patterns," in *WACV*, Waikoloa Beach, HI, pp. 1092-1099, January, 2015.

[20] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657-1663, June 2010.

[21] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 2, pp. 513-529, April 2012.

[22] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew,"Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no.1-3, pp. 489-501, December 2006.

[23] D. Dai, and W. Yang, "Satellite image classification via two-layer sparse coding with biased image representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 1, pp. 173–176, January 2011.

[24] G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *International Journal of Remote Sensing*, vol. 33, no. 8, pp. 2395-2412, October 2011.

[25] L. Zhao, P. Tang, and L. Huo, "A 2-D wavelet decomposition-based bag-of-visual-words model for land-use scene classification," *International Journal of Remote Sensing*, vol. 35, no. 6, pp. 2296-2310, March 2014.

[26] W. Shao, W. Yang, G.-S. Xia, and G. Liu, "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, July 2013.

[27] L. Zhao, P. Tang, and L. Huo, "Land-Use scene classification using a concentric circle-structured multiscale bag-of-visual-words model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 12, pp. 4620-4631, December 2014.

[28] J. Ren, J. Zabalza, S. Marshall, and J. Zheng, "Effective feature extraction and data reduction in remote sensing using hyperspectral imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 149-154, July 2014.