Journal of
Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Residential land extraction from high spatial resolution optical images using multifeature hierarchical method

Zhongliang Fu
Xiaoli Liang

SPIE.

Zhongliang Fu, Xiaoli Liang, "Residential land extraction from high spatial resolution optical images using multifeature hierarchical method," *J. Appl. Remote Sens.* **13**(2), 026515 (2019), doi: 10.1117/1.JRS.13.026515.

Residential land extraction from high spatial resolution optical images using multifeature hierarchical method

Zhongliang Fu[a,b] and Xiaoli Liang[a,*]
[a]Wuhan University, School of Remote Sensing and Information Engineering, Wuhan, China
[b]Collaborative Innovation Center of Geospatial Technology, Wuhan, China

Abstract. Residential land (RL), as a typical kind of urban functional zone, plays an important role in urban planning and land census. Recent years have witnessed frequent changes in RL via the process of urbanization. The extraction of RL from high spatial resolution optical images can reflect the status quo of land use/land cover to a certain extent, which is of great significance to land census and urban planning. We adopt a scene classification strategy to extract RL and mainly focus on the extraction of four common types of RL in China: old-style village, low-density high-rise, medium-density low-rise, and low-density low-rise. We design a multifeature hierarchical (MFH) algorithm for RL extraction. First, RL is extracted based on the gray level concurrence matrix and a fuzzy classification algorithm. Then an improved bag-of-visual-words algorithm is introduced to further realize the extraction of RL. The effectiveness of our proposed method is analyzed with a sample dataset and large images. We also analyze the separability among different kinds of RL. We compare the experimental results with those of three other algorithms, and the results demonstrate that the MFH algorithm performs better in terms of the accuracy and efficiency of the RL extraction. The results can provide services for land surveying and urban planning, and the technological processes and experimental design in the algorithm can provide a reference for the research in related fields. [DOI: 10.1117/1.JRS.13.026515]

Keywords: residential land; high spatial resolution optical images; scene classification; multifeature; hierarchical.

Paper 190005 received Jan. 3, 2019; accepted for publication May 23, 2019; published online Jun. 19, 2019.

1 Introduction

With the rapid development of urbanization, the effective use and management of urban land have attracted extensive attention. The places where people partake in different socioeconomic activities are divided into different functional zones.[1] Residential land (RL), as a typical functional zone, is the most extensive land use type in cities, and it can reflect the situation of land use to a certain extent. Moreover, it is one of the most transformed land use/land cover (LULC) types in the process of urbanization in China. Timely information about the quantity and spatial distribution of this type of land is important for urban planning and investigations.

Currently, the methods employed for image information extraction mainly use remote sensing technologies. The rapid development of remote sensing technology enables people to quickly obtain a large amount of ground information through satellites and spacecrafts. High spatial resolution optical images (HSROIs) can represent the surface of the Earth in detail and are widely used in urban LULC extraction, which provides data resources for the extraction of RL. The application of HSROIs provides favorable conditions for the updating and application of GIS data and is of great significance for map updating, image matching, and target detection.[2,3] However, previous studies pay more attention to the information extraction of diverse land-cover objects (e.g., building,[4–8] road,[9–13] vegetation,[14–17] and cultivated land[18–21]), and fewer studies focus on the extraction of functional zones. Functional zones are spatially aggregated by different land-cover objects, and their categories are semantically abstracted from land use functions,[22]

*Address all correspondence to Xiaoli Liang, E-mail: 2017102130006@whu.edu.cn

e.g., commercial zones, industrial zones, residential districts, shanty towns, campuses, and parks.[23]

Functional zone studies are usually implemented by scene-based classification with very high resolution satellite images, where functional zones are represented by image scenes.[24,25] Recently, numerous efforts have been made to extract proper features from scene images and build effective classification models. In terms of feature extraction, the visual features, including spectral, textural, and geometrical features, are usually used to characterize the scene images.[26–34] However, only the visual features were used in these studies, and the semantic features that represent the special geographic information were ignored. Therefore, these methods are only effective in classifying simple scenes, rather than heterogeneous scenes with diverse kinds of objects.[35–37] To solve this issue, scale invariant feature transform (SIFT) features were introduced.[38] Unlike some visual features that are variant to affined transformations, SIFT features overcome the variability of scale and affinity issues and are widely used in image classification, scene recognition, and target detection.[39–41] In terms of classification methods, they mainly use techniques for measuring the feature similarity between scene images and labeling scenes using various classifiers, such as the K-nearest neighbor, maximum likelihood, support vector machine (SVM), artificial neural network, and random forests.[42] However, these classifiers are only capable of dealing with the visual features and are easily affected by feature changes. Therefore, more effective models, such as the latent Dirichlet allocation[43] and bag of words,[44] are gradually being introduced to improve the classification accuracy. Nevertheless, the previous methods are not capable of tackling the urban RL recognition and classification task by solely using simple features and classification models, as the residential scenes are often heterogeneous with complex components and various semantic categories.

In addition, other issues, such as the classification of RL, the universality of the dataset, and the generalization of the method, also impede RL extraction in Chinese urban areas. There are great differences in the morphological structures and geographical distributions of different types of RL, so it is more practically significant for further classification of RL. However, in the study by Yang and Newsam,[45] RL was not subdivided, and Xia et al.[46] divided RL into dense residential and rural residential. These classification systems are too general to distinguish different types of RL. In other relevant studies,[47,48] the housing types in RL are quite different from those in China. In Chinese towns, residential areas are dominated by high-rise housings, whereas in the United States, residential areas mainly include single-family housing, multifamily residential, and mobile homes. These three residential types also appear in the frequently used scene classification dataset, the UC merced land use dataset, which can be downloaded from the United States Geological Survey National Map.[49] Other commonly used datasets such as the SIRI-WHU[50] and WHU-RS19 datasets[51] are selected from Chinese areas, and the RL is taken as a class in the scene classification without being further subdivided. Moreover, the images of these datasets and the classification schemes were built according to land use scenes, not functional zones. Therefore, these datasets are not suitable for the extraction of RL in China. In addition, these studies only involve the classification of scene images in the dataset without considering the applicability of large-scale remote sensing images.

In summary, this study aims to address its four key issues: the classification scheme, dataset, features and models, and applicability of large area images. This study built a classification scheme in line with the current status of LULC in China and subdivided RL into four types [old-style village (OSV), low-density high-rise (LDHR), medium-density low-rise (MDLR), and low-density low-rise (LDLR)] according to the characteristics of Chinese residential buildings, such as the morphological structure, distribution location, floor height, and floor spacing. In our study, the HSROIs provided by Google Earth were used to collect the samples and build the dataset. Though the Google Earth images have been preprocessed using RGB renderings from the original optical aerial images, there is no significant difference between the Google Earth images and the real optical aerial images, even in the pixel-level LULC mapping.[52] Thus Google Earth images can also be used as aerial images for scene classification. Many datasets, such as the AID,[45] SIRI-WHU[50], WHU-RS19,[51] and RSSCN7[53] datasets, are collected from Google Earth images. For the features and models, we proposed a multifeature hierarchical (MFH) method to extract RL, and the validity of our algorithm was verified by large area images. At present, many works on information extraction adopt the idea of deep learning, but it requires

a large number of samples. Since our algorithm does not require a large number of samples to achieve the extraction of RL, the deep learning method is not considered here.

The main contributions of our work are listed below:

- In this study, we subdivided RL and constructed a reasonable category system and sample dataset that is in line with China's geographical conditions.
- We designed the MFH algorithm for the RL extraction and further analyzed the separability of single-class RL. In our work, the traditional fuzzy classification and bag-of-visual-words (BOVW) model were improved to realize the rapid and automatic extraction of RL.
- The MFH algorithm was applied to the extraction of RL in large-scale images, which provided strategies and approaches for the realization of automatic and fast national LULC investigations, and further provided reference basis for urban planning.

The remainder of this paper is organized as follows: In Sec. 2, the details of our proposed methods are described. Section 3 demonstrates the effectiveness of the proposed model using a sample dataset and large images. The discussion and conclusion are given in Secs. 4 and 5, respectively.

2 Methodology

In our study, a classification scheme was first built, and the samples of each class were collected through Google Earth. Our classification scheme was built according to the study areas, Beijing and Tianjin, where Google Earth provides high precision images, and the LULC types are diverse, as well as a certain number of scattered residential areas. The experimental data collected from the images of such a research area are more typical and representative, which can better reflect the characteristics of the ground objects in northern China. We collected samples with a size of 300 m $\times$ 300 m extracted from 1-m spatial resolution optical images through Google Earth. Then the MFH algorithm was designed for RL extraction. The algorithm consists of two steps. First, the gray level concurrence matrix (GLCM) texture features and fuzzy classification (GLCM-FC) algorithm were used to realize the extraction of RL. Then an improved BOVW (IBOVW) model was constructed for the classification of the categories that cannot be distinguished from RL. We also further analyzed the separability of each type of RL based on the proposed methods. The framework of the MFH algorithm for RL extraction is shown in Fig. 1.

Our algorithm considers the following aspects:

- *Multiscale and multifeature*. Multitype and multiscale rotation-invariant GLCM texture global features and SIFT local features were used to characterize the different kinds of LULCs.
- *Integrated*. We improved the traditional fuzzy classification and the BOVW model and combined the advantages of the two algorithms to realize the rapid and automatic extraction of RL.
- *Universality*. The proposed algorithms not only apply to the retrieval of RL images in the sample dataset but also work well in images covering large areas. Moreover, the algorithm can be used for the extraction of other kinds of categories.

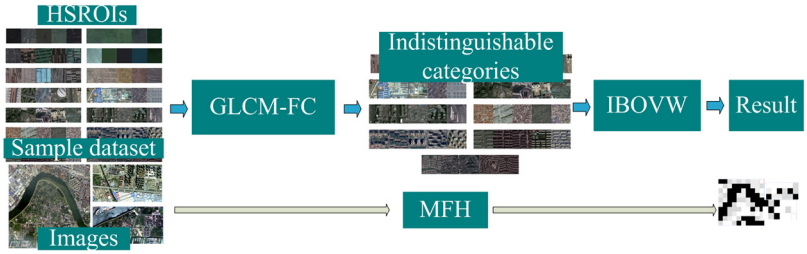

Fig. 1 The framework of the MFH algorithm.

2.1 *Classification Scheme Building and Sample Collection*

A classification scheme was built based on current Chinese land use classification[54] and the characteristics of the land-cover objects in the study area. In this classification scheme, 14 categories commonly seen in urban and rural areas were selected, as listed in Table 1, including woodland, grassland, cultivated land, water, farming facilities, bare land, transportation, industrial, public management and public service (PMPS), commercial and service district (CSD), OSV, LDHR, MDLR, and LDLR. This detailed classification strategy can help us to emphasize RL and the inner feature differences in the LULC types in the image, at the same time, it can help us to understand which classes are easily confused with RL.

RL is mainly a polygonal ground object with irregular shapes, different sizes, and complicated boundaries. Therefore, it is difficult to define and characterize it on a uniform scale. In our work, the samples were selected with a size of 300 m × 300 m extracted from a 1-m spatial resolution optical image through Google Earth based on field survey and expert interpretation. This scale can better describe RL by investigating a large number of RL images in the research area. Selected samples were randomly divided into training samples and validation samples.

2.2 *GLCM-FC Algorithm*

In the GLCM-FC algorithm, we use the idea of underdeveloped village extraction from our previous work to extract RL.[55] This work could be divided into two steps, including the GLCM texture feature extraction and fuzzy classification. The novelties of the algorithm are twofold: (1) multitype and multiscale rotation-invariant GLCM texture features were used to characterize the local spatial arrangements of different kinds of LULC and (2) a fuzzy deduction process was constructed to fuse the image features with different ranges and properties. Figure 2 illustrates the detailed steps.

2.2.1 *GLCM texture feature extraction*

The GLCM is a method to describe texture by studying the grayscale spatial correlation of an image, and it is one of the commonly used texture statistical analysis methods. Its essence is the frequency of co-occurrence of pixel pairs with fixed relative positions, which is the basis for analyzing the local pattern structure of an image and its arrangement rules. We assume that

Table 1 Classification scheme.

No.	01	02	03	04	05	06	07
Class	Woodland	Grassland	Cultivated land	Water	Farming facilities	Bare land	Transportation
No.	08	09	10	11	12	13	14
Class	Industrial	PMPS	CSD	OSV	LDHR	MDLR	LDLR

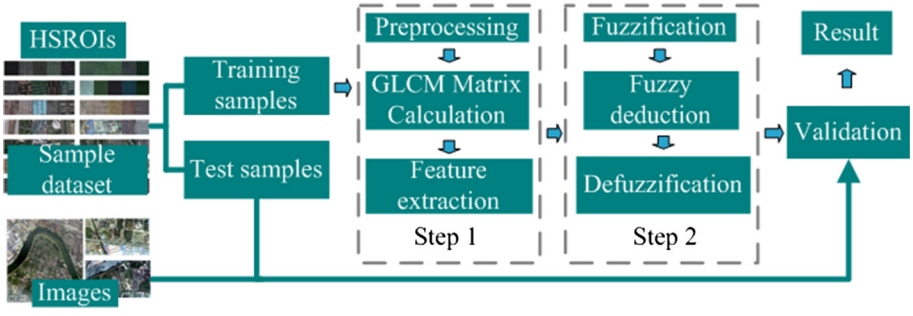

Fig. 2 Framework of the GLCM-FC algorithm.

$f(x, y)$ is a digital image, with a size of $M \times N$ and gray level of N_g. Then the GLCM satisfying a certain spatial relationship is shown in

$$p(i, j) = \#[(x_1, y_1), (x_2, y_2) \in M \cdot N | f(x_1, y_1) = i, f(x_2, y_2) = j]. \quad (1)$$

where # denotes the number of elements in the set, p is a matrix with a size of $N_g \times N_g$, (x_1, y_1), and (x_2, y_2) are two pixels in the image with gray tones i and j, respectively. If the distance between (x_1, y_1) and (x_2, y_2) is d, and the angle between them and the horizontal axis of the coordinates is θ; then, the GLCM $p(i, j, d, \theta)$ of various spacings and angles can be obtained.

In our approach, feature extraction was conducted based on the GLCM textures and training samples. Before texture extraction, the input image needed to be preprocessed, including grayscale image extraction and grayscale quantification. The grayscale image was obtained by a band calculation. By comparing with other methods (including weighting, maximum, mean value, and band selection), we finally chose to extract the green channel of the image to turn the RGB image into a grayscale one. To realize the grayscale quantification, the gray value range of the grayscale image was linearly compressed into 64 levels. Haralick et al.[56] adopted statistical methods to extract 14 second-order statistics from the GLCM as texture feature measures. In our method, four commonly used textures were selected, including contrast (CON), entropy (ENT), homogeneity, (HOM), and angular second moment (ASM). Each texture was made rotation-invariant by averaging over its four directions (0 deg/45 deg/90 deg/135 deg). The multiscale texture features were obtained using step sizes of 1/2/3; thus a total of 12 texture features were obtained. Four commonly used texture features[57] are shown in Table 2. The factors in texture feature extraction are listed in Table 3. Each texture feature was represented by texture feature abbreviations and the step size, e.g., ASM3 means angular second moment with a step size of three. The ability of each texture to discriminate RL from other classes was visually interpreted by plotting the values of training samples in a box plot, as shown in Fig. 3. The rotation-invariant textures with good discriminative performance were selected as input features.

Table 2 Four commonly used texture features.

Texture	Equation	Characteristic
CON (C_1)	$C_1 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i-j)^2 p(i, j; d, \theta)$	CON indicates the difference in the grayscale in the image neighborhood, reflecting the local change in image sharpness and the degree of texture value. The larger the local variation of the image, the higher the value is
ENT (C_2)	$C_2 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p(i, j; d, \theta) \lg p(i, j; d, \theta)$	ENT represents the degree of nonuniformity or complexity of texture in the image
HOM (C_3)	$C_3 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{p(i,j;d,\theta)}{1+(i-j)^2}$	HOM is a measure of image texture similarity. The higher the value is, the less change in the local area and the smaller the gray level difference
ASM (C4)	$C_4 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [p(i, j; d, \theta)]^2$	ASM reflects the uniformity of grayscale distribution and the texture thickness of the image. The more similar the pixel values in the region are, the higher the HOM, and the larger the ASM value

Table 3 Factors in texture feature extraction.

Texture	Step size	Window size	Direction
CON	1/2/3	300 m × 300 m	The average of four directions (0/45/90/135 deg)
ENT			
HOM			
ASM			

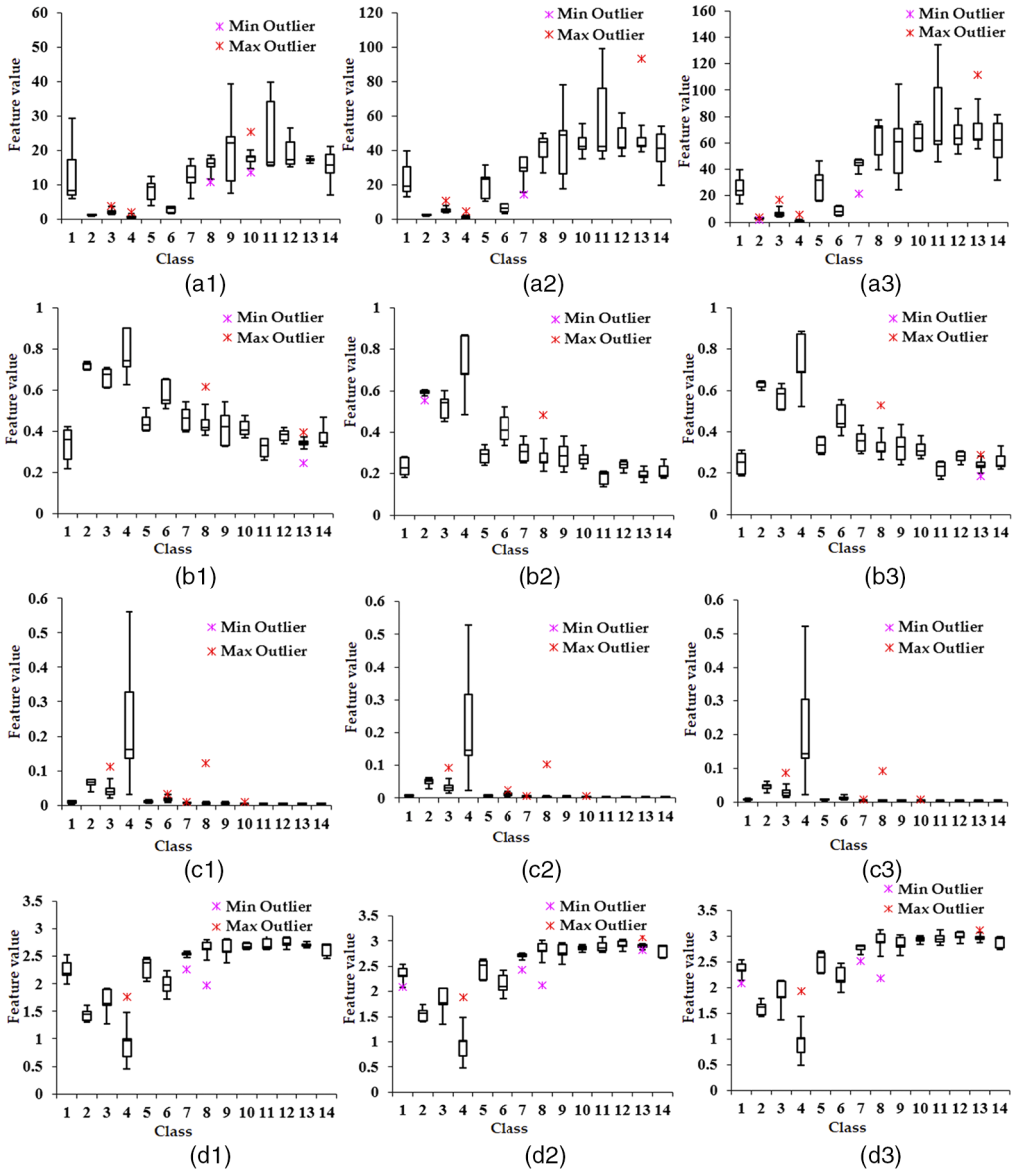

Fig. 3 Comparison chart of the GLCM texture features: (a1) CON1, (a2) CON2, (a3) CON3, (b1) HOM1, (b2) HOM2, (b3) HOM3, (c1) ASM1, (c2) ASM2, (c3) ASM3, (d1) ENT1, (d2) ENT2, and (d3) ENT3. In the above chart, the ends of the box plot are set at 1.5* interquartile range (IQR) above the third quartile (Q3) and 1.5*IQR below the first quartile (Q1). Values above the range and below the range are treated as abnormal and are indicated by purple and red, respectively.

Figure 3 indicates that RL (classes 11/12/13/14) is easily confused with classes 5/7/8/9/10 by comparing 12 texture features. Among the four texture measures, ENT, ASM, and HOM have better discriminative performances. The best features that can discriminate RL and other classes are ENT with steps 2/3, namely, ENT2, ENT3, respectively, which can distinguish RL from classes 1/2/3/4/6.

2.2.2 *Fuzzy classification*

Fuzzy classification is a powerful and flexible soft classifier. It contains the membership degree of multidimensional types, which describes the assignment degree of the object under consideration to n different categories. The specific equation is described as follows:

$$f_{\text{class,obj}} = [u_{\text{class1}}(\text{obj}), u_{\text{class2}}(\text{obj}), \ldots, u_{\text{classn}}(\text{obj})]. \quad (2)$$

Typical fuzzy classification includes three main steps: fuzzification, fuzzy deduction, and defuzzification.[58] In this algorithm, fuzzification was realized based on the histogram distribution of selected features in the training samples. In the fuzzification process, only training samples belonging to RL were used. For each feature, the algorithm retrieved its training samples in each class, then the histogram of the samples was calculated and normalized to (0, 1). Finally, the convolution of the normalized histogram with a predefined one-dimensional kernel leads to the membership function of the feature for each class. Each class was assigned several membership functions, which worked like lookup tables for fuzzy deduction. In the fuzzy deduction process, for each input object, the algorithm calculated its fuzzy membership in each class. This was achieved by calculating the maximum membership of all selected features in each class. The membership of each feature in each class could be easily extracted from the lookup table. Finally, defuzzification was achieved by comparing the memberships of all classes for each input object. The input object was assigned to the class with the maximum membership. If the obtained membership is equal to the maximum or very low, the class of the input object remains undetermined. It should be noted that the limitation of the training samples confined the representational ability of the lookup tables and the convolution in the fuzzification process is to generalize the ability of lookup tables and plays an important role in our approach.

2.3 *IBOVW Algorithm*

This model includes four steps: SIFT feature descriptor creation, vocabulary generation, visual word histogram construction, and classifier selection. In this algorithm, we proposed a dictionary generation method and adopted an improved minimum distance (MD) classifier instead of the traditional SVM. The framework of our proposed method is shown in Fig. 4.

2.3.1 *SIFT feature descriptor*

We choose the Lowe's SIFT feature descriptor to achieve interest point detection and description. An image analysis using SIFT features typically has two steps: a detection step is first needed to identify the interest points in the image, and then the descriptors are extracted from each image patch centered at the detected locations. The SIFT detection step is achieved by searching for local extreme points in the difference in the Gaussian images in scale space. Then the position and scale of the salient points in the image are determined. Finally, the SIFT local descriptor is obtained by extracting the normalized region gradient histograms at the locations of the found salient points. The feature descriptor consists of the histograms of gradient directions computed over 4×4 spatial grids. The interest point orientation estimate is used to align the gradient directions to make the descriptor rotation invariant. The gradient directions are quantified into eight bins, and the final feature can be represented by a vector of 128 $(4 \times 4 \times 8)$ dimensions.

In our algorithm, SIFT feature extraction was implemented using Lowe's demo code for detecting and matching SIFT features. The generated feature file includes the position, scale, direction, and 128-dimensional feature vector information of each feature point. The length of the SIFT feature vector was normalized to further remove the influence of illumination variation.

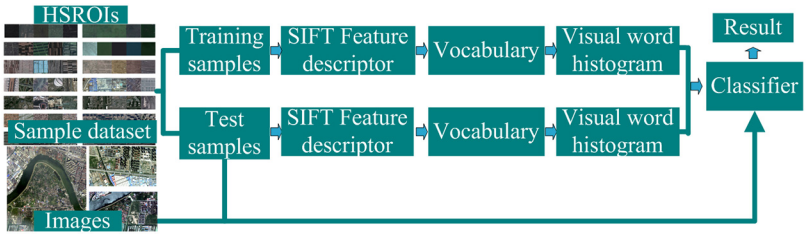

Fig. 4 Framework of the IBOVW.

2.3.2 *Vocabulary generation*

The vocabulary in traditional BOVW (TBOVW) model is created by applying *k*-means clustering to the SIFT descriptors of all categories. However, the number of interest points extracted in each category is different, and samples of some categories have fewer interest points (e.g., water and grassland). The TBOVW would cause the feature suppression of these categories, and the generated vocabulary could not express well, which would inevitably affect the retrieval performance.

We proposed a method to generate the vocabulary. In our method, *k*-means clustering was first conducted on the SIFT features of the training sample of each class to generate the vocabulary of each category, and then the vocabulary of each class was combined to form the overall vocabulary. In this way, each category generates its own vocabulary, which avoids feature suppression and loss. The overall vocabulary is a collection of visual words from all categories that can comprehensively express various characteristics. The flowchart of the vocabulary generation of our proposed method is given in Fig. 5.

2.3.3 *Visual word histogram construction*

The visual word histogram was achieved through three steps in our method. First, we calculated the Euclidean distance between each interest point of each sample and each word in the vocabulary. The point was then assigned to the closest visual word. Finally, we counted the frequencies of each word in the image and used the visual word frequency histogram as the feature vector of the image. The final representation of an image was the frequency count or histogram of the labeled SIFT features as shown in

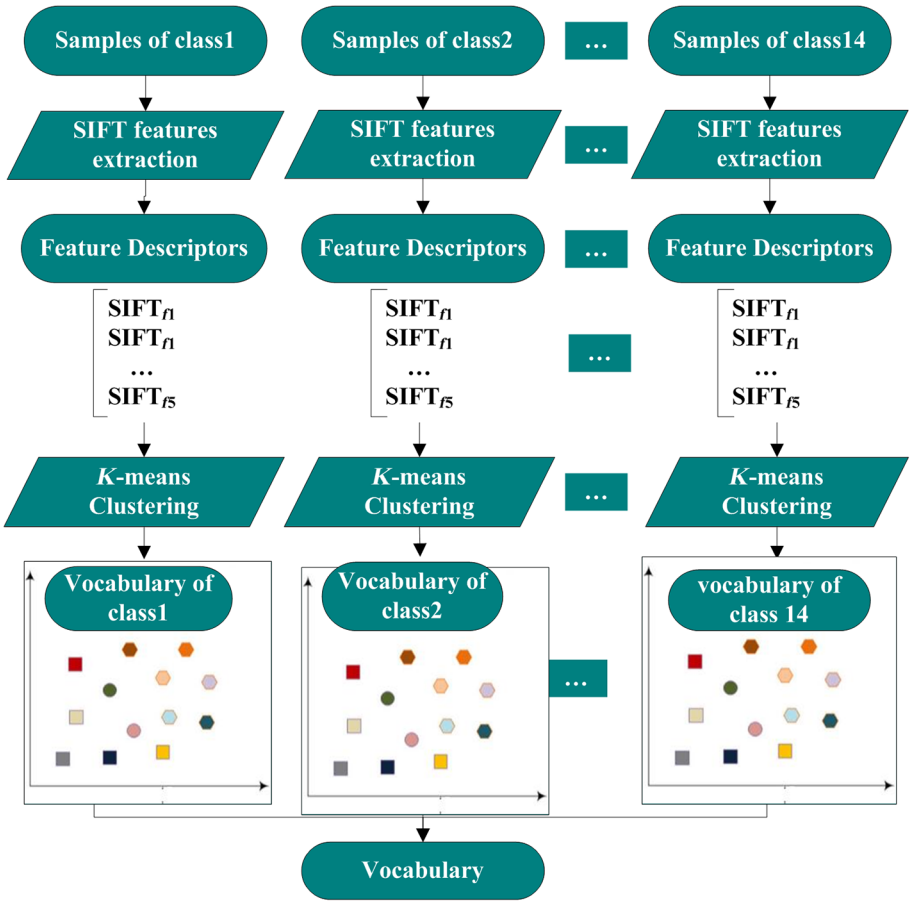

Fig. 5 Flowchart of the generation of the vocabulary.

$$\text{image} = (v_1, v_2, \ldots, v_n), \tag{3}$$

where v_i is the number of times a visual word appears in the image and n is the size of the vocabulary. Each image was represented by a multidimensional feature vector. To account for the difference in the number of interest points between the images, the feature histogram vector should be normalized to have unit L1 norm.

2.3.4 *Classifier*

Appropriate classifiers were selected to classify the feature histogram constructed using the training samples and test samples. Here we choose the MD classifier instead of the traditional SVM classifier for RL extraction.

The MD method is realized by distance similarity retrieval. In this paper, the histogram distance measures were used to compare the SIFT histogram features, and the phase anisotropy of histogram was measured by Euclidean distance. Since the variances in the different categories are different, it is not possible to divide the attribution of pixels by simply using the distance from the pixel to the center of the category. We improved the Euclidean distance to increase the classification accuracy as shown in

$$d_{12} = \sqrt{\sum_{i=1}^{n} (x_{i1} - x_{i2})^2 / \sigma_{12}^2}, \tag{4}$$

where x_{i1} is the first-dimensional coordinate of the first point, x_{i2} represents the first-dimensional coordinate of the second point, and σ_{12} is the standard deviation of the two points. In the MD classification method, the MD between each test sample and each category was first calculated. This was realized by calculating the minimum Euclidean distance of the feature histogram between each test sample and all training samples of each category. In our paper, there are a total of 14 categories; thus 14 MD values were obtained between each test sample and all classes. Then the category of each test sample was determined, which was obtained by calculating the minimum value of these 14 distance values.

3 Experiments and Analysis

In this section, we introduce the experimental data, design the experiment setup, and analyze the experimental results. The sample dataset and large area images are used to verify the effectiveness of the MFH algorithm in our experiments. In the first experiment, we analyzed the extraction results of RL for the GLCM-FC algorithm using a sample dataset, in which the effectiveness of convolution processing in fuzzy classification was also analyzed, and we also further verified the algorithm in three large area images. In the second experiment, the IBOVW algorithm was applied to the undifferentiated categories, and three other comparison experiments were used to prove the availability of the MFH algorithm using the sample dataset and three images. Finally, we further analyzed the separability of single-class RL.

3.1 *Experimental Data*

The experimental data include a sample dataset and three large area images of the Tianjin area. The dataset contains 10 samples from each of the 14 LULC classes, and half of them are training samples. Five samples of each class are displayed in Fig. 6. The other three large area images (Fig. 7) are selected to test and verify the applicability of our algorithm, including two Google Earth images with a size of 1800 m × 900 m [image TA (upper right) and image TB (lower right)] and an aerial RGB image with a size of 3000 m × 3000 m [image TC (left)], which was resampled from the 0.3-m resolution aerial RGB image. The selection of images from different sources also further verifies the adaptability of our proposed algorithm to different data sources.

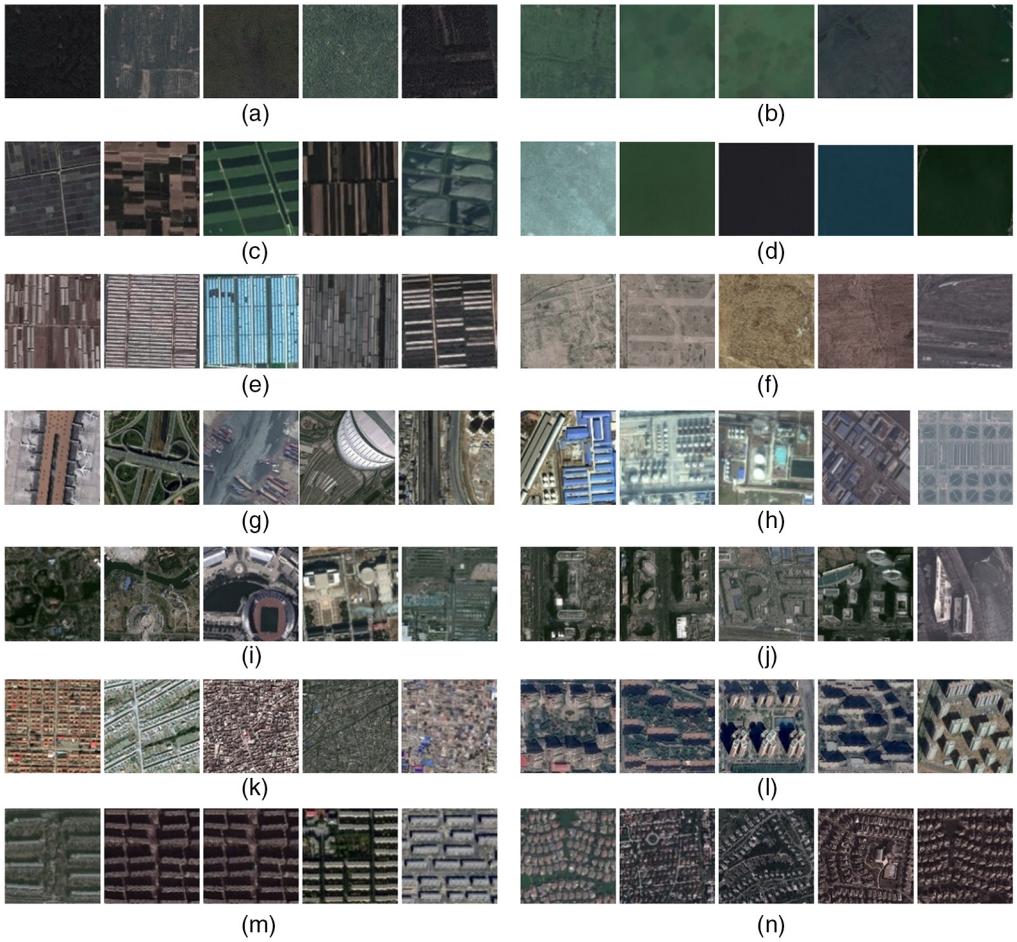

Fig. 6 Five samples of each class in dataset: (a) woodland, (b) grassland, (c) cultivated land, (d) water, (e) farming facilities, (f) bare land, (g) transportation, (h) industrial, (i) PMPS, (j) CSD, (k) OSV, (l) LDHR, (m) MDLR, and (n) LDLR.

3.2 *Experimental Setup*

Two groups of experiments were designed to show the effectiveness of the MFH algorithm, and the sample dataset and large area images were used for experimental verification. Since our purpose is to extract RL, categories other than RL are taken as a signal category, nonresidential land (non-RL), and the extraction results of any class of object within non-RL are not further studied in our experiments. In the accuracy assessment, we only analyzed the extraction accuracy of RL. We chose commonly used accuracy assessment metrics[59–61] in information retrieval, including precision (P), recall (R), and F-measure (i.e., the F-score. $F1$-measure are commonly used), and the overall accuracy (OA) of RL extraction was also added to our experiments as a reference index.

In the first experiment, the optimum texture features were chosen for the extraction of RL, in which the sample dataset and images were used in the GLCM-FC algorithm. In this method, only the training samples of RL were trained with fuzzification. The bin size in the histogram calculation was set to 32, and the convolution kernel for the look-up tables generalization was set to (0.05, 0.1, 0.2, 0.3, 1, 0.3, 0.2, 0.1, 0.05).

In the second experiment, the categories that were not distinguished from RL in the GLCM-FC algorithm were further classified by means of the IBOVW. In this experiment, five visual words were generated in each category, with a total of 70 words in the vocabulary, and this method was iterated 10 times to obtain the overall vocabulary. Three other RL extraction methods, i.e., the GLCM + fuzzy,[55] TBOVW, and IBOVW, were designed to compare with the MFH algorithm. In the comparison experiments, the sample dataset and three large area images were

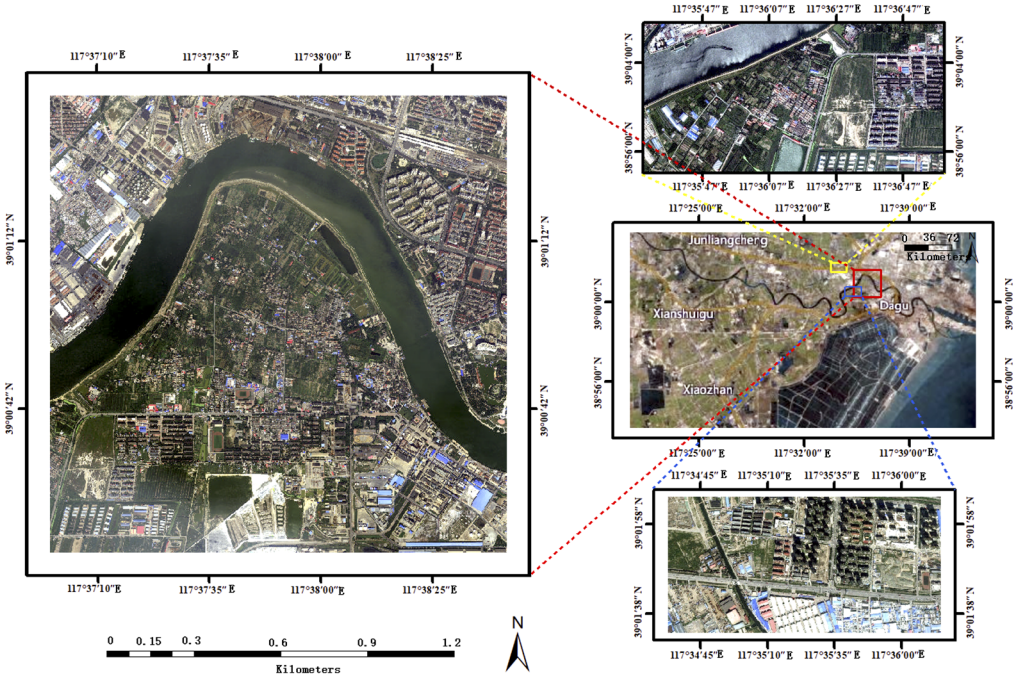

Fig. 7 Images of Tianjin area. The upper right image (image TA) and the lower right image (image TB) are taken from Google Earth with a size of 1800 m × 900 m. The left image (image TC) is an aerial RGB image with a size of 3000 m × 3000 m.

used to extract RL, and the samples in all categories participated in training and classification in the TBOVW and IBOVW experiments. In the first two experiments, four types of RL were analyzed as a whole, and the separability of the four types of RL and the distinction between non-RL and RL were carried out in the final separability experiment.

3.3 *Results and Analysis*

3.3.1 *First experiment*

Training samples of RL were used and the best texture features, ENT 2 and ENT 3, were selected as input features. Then a fuzzy classification process was constructed to fuse the best sample image features. The experiment results in the GLCM-FC algorithm are shown in Table 4 and Table 5.

From Table 4, we can see that almost all samples of RL are correctly extracted (see bold values), and no samples of classes 1/2/3/4/6 are divided into RL. The wrongly classified categories mainly come from classes 5/7/8/9/10, which is consistent with the interpretation results of the box-plot using the training samples. Table 5 shows the results obtained without using kernel convolution in the building of the lookup tables in the fuzzification process. Although classes 1/2/3/4/5/6 can be distinguished from RL, many samples of RL have not been extracted;

Table 4 Experimental result of the GLCM-FC algorithm under the condition of convolution in sample dataset.

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	∑
Sample number	10	10	10	10	10	10	10	10	10	10	10	10	10	10	140
Unclassified	10	10	10	10	4	10	4	3	1	0	1	0	0	0	63
Classified to RL	0	0	0	0	6	0	6	7	9	10	**9**	**10**	**10**	**10**	77

Table 5 Experimental result of the GLCM-FC algorithm under the condition of no convolution in sample dataset.

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	$\sum$
Sample number	10	10	10	10	10	10	10	10	10	10	10	10	10	10	140
Unclassified	10	10	10	10	10	10	7	8	5	7	5	3	4	3	102
Classified to RL	0	0	0	0	0	0	3	2	5	3	**5**	**7**	**6**	**7**	38

Table 6 Accuracy assessment and analysis of RL extraction in sample dataset based on the GLCM-FC algorithm.

Experiment	*P* (%)	*R* (%)	*F*1 (%)	OA (%)	Distinguishable categories
GLCM-FC (with convolution)	50.649	97.500	66.666	72.143	1/2/3/4/6
GLCM-FC (without convolution)	65.789	62.500	64.102	80.000	1/2/3/4/5/6

we can see the results from the bold values in Table 5. The accuracy assessment and analysis of RL extraction are given in Table 6, which was obtained from the above two tables.

In the convolution experiment, a total of 77 samples were classified as RL, among which, 38 samples of non-RL were incorrectly classified into RL. The precision of RL extraction is 50.649%, the recall is 97.500%, and the OA is 72.143%. In the experiment without convolution, the distinguishable categories, precision, and OA are improved, whereas the recall is 35% lower, and only 25 samples of RL were correctly extracted. This indicates that some of the test samples in RL remained unclassified. Because some of the test samples are not in the same bin with the training samples in the normalized histogram, the value of $F1$ (the comprehensive evaluation index of P and R) is higher in the former group experiment, which means that it performs better in RL extraction. This proves the usefulness of the convolution in the lookup table construction.

Next, we apply this algorithm to the RL extraction in large area images. The extraction results of three images were obtained under the HOM3 feature, as shown in Fig. 8. The accuracy assessment of the experiments was evaluated by the extraction result images and the corresponding membership images. Each square in the results represents an image block with a size of 300 m × 300 m. In the extraction result images, the white and gray squares represent the identified RL, and the black ones are non-RL. The membership images indicate the probability of each image block being classified as RL. The brighter the square is, the higher the probability of the image block being classified into RL, and the darker the square is, the lower the probability is.

The results and accuracy analysis of the three images are shown in Table 7. In image TA, a total of 14 image blocks are classified into RL, of which only 8 are correctly classified. The extraction precision of RL is 57.143%, and the recall is 80.000%. The precision of the retrieval accuracy of RL in image TB is 50.000%, only half of the image blocks classified as RL are correct, and 87.500% of RL in this image is extracted. Although in image TC, the precision of RL extraction is 1.667% higher than that of image TB, the OA is 5.444% higher than that of the first two images, and the recall ratio and $F1$ are lower than that of the first two images. In image TC, the industrial zone is confused with RL, and woodland, water, the transportation district, and PMPS are also wrongly classified. This is caused by different data sources, as image TC was an aerial RGB image, with more detailed feature information, whereas the training samples were taken from the satellite images of Google Earth, with smoother texture features.

The results show that this algorithm can distinguish RL from some types of categories, but the wrongly classified rate is high. In addition, the algorithm is affected by the number of categories in the training stage. We chose two metrics ($F1$ and OA) to show the relationship between the extraction accuracy of RL and the number of categories involved in the training stage (Fig. 9). With the increase in categories added in training, the extraction accuracy of RL shows a declining trend as a whole, especially for the OA, which decreases faster. Moreover, the applicability of the algorithm to different data sources still needs to be improved.

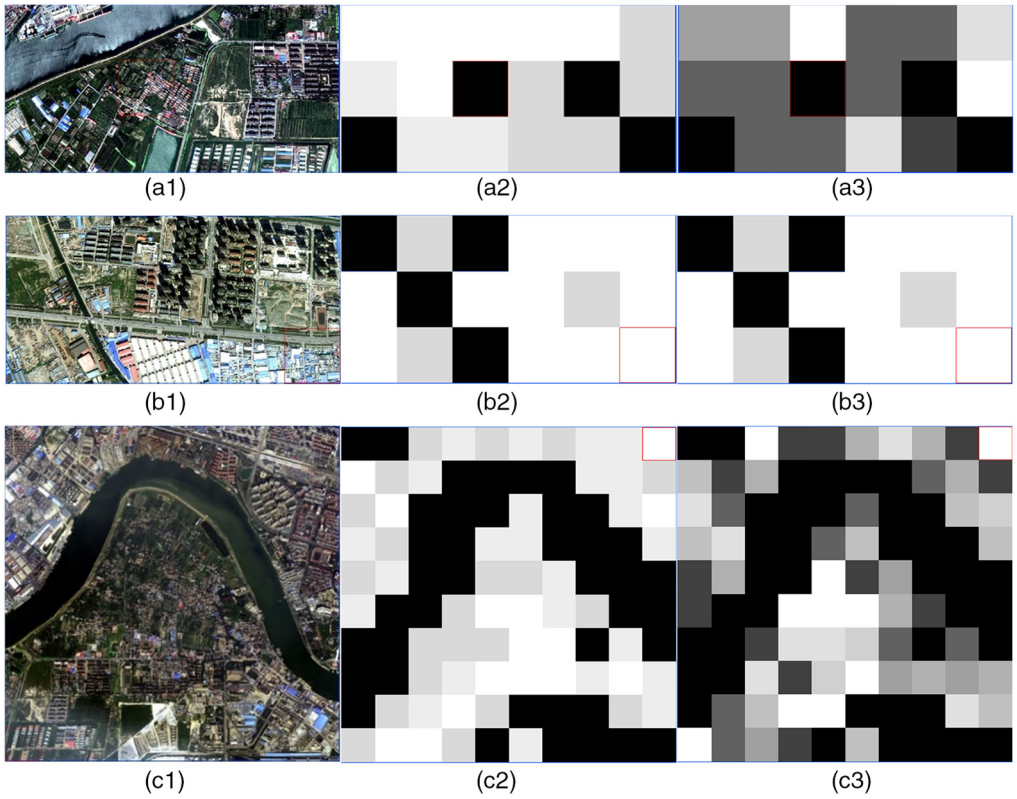

Fig. 8 RL extraction results for three images of the Tianjin area for the GLCM-FC algorithm: (a1) Image TA, (a2) extraction result in image TA, (a3) membership of extraction result in image TA, (b1) image TB, (b2) extraction result in image TB, (b3) membership of extraction result in image TB, (c1) image TC, (c2) extraction result in image TC, and (c3) membership of extraction result in image TC.

Table 7 Accuracy assessment and analysis of RL extraction in large area images based on the GLCM-FC algorithm.

Image	Image block	RL	Non-RL	$\sum$	*P* (%)	*R* (%)	*F*1 (%)	OA (%)
TA	Number of each class	10	8	18	57.143	80.000	66.667	55.556
	Number of classified into RL	8	6	14				
TB	Number of each class	8	10	18	50.000	87.500	63.636	55.556
	Number of classified into RL	7	7	14				
TC	Number of each class	41	59	100	51.667	75.610	61.386	61.000
	Number of classified into RL	31	29	60				

However, our method is very flexible, in which only samples of the target objects (in our algorithm, only training samples of RL) were trained, avoiding the interference of other categories.

3.3.2 *Second experiment*

Although some classes can be distinguished from RL in the GLCM-FC algorithm, it is difficult to achieve an effective extraction of RL in complex urban scenes by solely using texture features. The indistinguishable categories (classes 5/7/8/9/10) and RL (11/12/13/14) are then classified by means of IBOVW, and the classification results and accuracy assessment of RL extraction based on MFH extraction are displayed in Table 8.

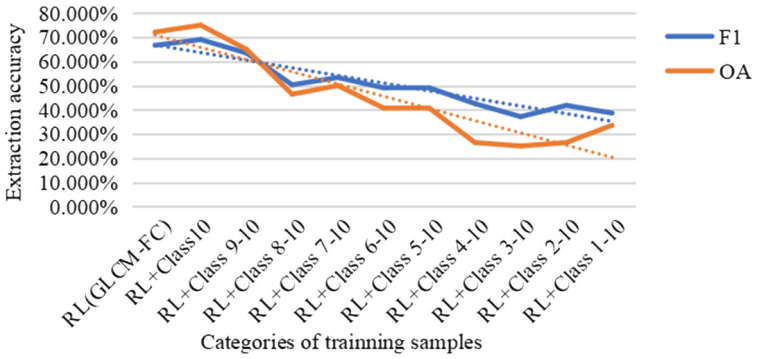

Fig. 9 Relationship between the extraction accuracy of RL and the number of categories involved in the training stage. The red line indicates the OA, and the blue line indicates the *F*-score.

Table 8 Classification results and accuracy assessment in the MFH algorithm.

		Prediction class						
Confusion matrix		5	7	8	9	10	RL	$\sum$
True class	5	6	0	0	1	0	3	10
	7	0	9	0	0	1	0	10
	8	0	1	7	1	0	1	10
	9	1	0	1	6	1	1	10
	10	0	0	0	0	5	5	10
	RL	0	1	1	3	1	34	40
	$\sum$	7	11	9	11	8	44	90
Accuracy assessment of RL extraction		P:34/44 = 77.273% R:34/40 = 85.000% $F1$:$2P*R/(P+R)$ = 80.953% OA:(34 + 40)/90 = 82.222%						

According to the classification results, we can obtain the extraction accuracy of RL. A total of 44 samples are classified as RL, in which there are 34 residential areas, 10 nonresidential areas, and the wrongly classified samples come from classes 5/8/9/10. The precision, recall, and *F*-score of RL extraction are 77.273%, 85.000%, and 80.953%, respectively. In this algorithm, the overall extraction accuracy of RL extraction is 82.222%.

The MFH algorithm is a fusion algorithm that is realized by combining on two methods. We further analyzed the RL extraction results using these two methods and the TBOVW algorithm. The detailed comparison analysis of four experiments is shown in Table 9. Compared with the experiments of GLCM-FC (Table 10), IBOVW (Table 11), and TBOVW (Table 12), the MFH algorithm reduced the false detection rate, and the extraction precision of RL was raised by

Table 9 Comparative experiments of RL extraction in the sample dataset.

Experiment	*P* (%)	*R* (%)	*F*1 (%)	OA (%)	Misclassification number	Time (s)
GLCM-FC	50.649	97.500	66.666	72.143	38	9.549
IBOVW	52.174	60.000	55.814	72.857	22	366.727
TBOVW	41.558	80.000	54.700	62.143	45	705.475
MFH	**77.273**	85.000	**80.953**	**82.222**	**10**	252.608

Table 10 Classification results of the GLCM-FC algorithm.

Confusion matrix		Prediction class											
		1	2	3	4	5	6	7	8	9	10	RL	$\sum$
True class	1	10	0	0	0	0	0	0	0	0	0	0	10
	2	0	10	0	0	0	0	0	0	0	0	0	10
	3	0	0	10	0	0	0	0	0	0	0	0	10
	4	0	0	0	10	0	0	0	0	0	0	0	10
	5	0	0	0	0	5	0	0	0	0	0	5	10
	6	0	0	0	0	0	10	0	0	0	0	0	10
	7	0	0	0	0	0	0	3	0	0	0	7	10
	8	0	0	0	0	0	0	0	2	0	0	8	10
	9	0	0	0	0	0	0	0	0	2	0	8	10
	10	0	0	0	0	0	0	0	0	0	0	10	10
	RL	0	0	0	0	1	0	0	0	0	0	39	40
	$\sum$	10	10	10	10	6	10	3	2	2	0	77	140

26.624%, 25.099%, and 35.715%, respectively. Ten nonresidential samples (see the bold values in Table 9) were misclassified as RL. The $F1$ and OA of RL extraction in the MFH algorithm are higher than those in the comparative experiments, the extraction results are shown in bold values in Table 9. Compared with the experimental results of IBOVW and TBOVW, the recall ratio of RL in the MFH algorithm is higher, and due to the reduced number of classes involved in the classification, the MFH algorithm saved time and increased the efficiency.

Table 11 Classification results of the IBOVW algorithm.

Confusion matrix		Prediction class											
		1	2	3	4	5	6	7	8	9	10	RL	$\sum$
True class	1	7	2	0	0	0	0	0	0	0	0	1	10
	2	2	5	0	0	1	0	0	0	0	0	2	10
	3	0	0	5	0	2	0	1	0	0	1	1	10
	4	2	0	0	5	0	0	0	2	0	0	1	10
	5	0	0	0	0	5	0	0	0	0	1	4	10
	6	0	0	0	0	0	8	0	0	0	0	2	10
	7	0	0	1	0	0	0	5	1	0	1	2	10
	8	0	0	0	0	1	0	0	5	1	2	1	10
	9	0	0	2	0	0	0	0	0	3	1	4	10
	10	0	0	0	0	0	0	0	0	1	5	4	10
	RL	1	0	1	1	1	1	2	5	1	3	24	40
	$\sum$	12	7	9	6	10	9	8	13	6	14	46	140

Table 12 Classification results of the TBOVW algorithm.

Confusion matrix		Prediction class											
		1	2	3	4	5	6	7	8	9	10	RL	$\sum$
True class	1	0	0	1	0	1	0	0	0	1	0	7	10
	2	1	2	1	2	0	0	1	0	0	0	3	10
	3	1	0	0	0	3	1	2	0	2	1	0	10
	4	1	3	0	1	0	1	0	1	0	0	3	10
	5	0	2	0	0	0	2	0	0	0	0	6	10
	6	0	1	0	0	2	0	1	1	0	1	4	10
	7	0	0	0	0	1	0	0	1	2	2	4	10
	8	0	0	0	0	0	1	1	0	1	1	6	10
	9	0	0	0	0	0	0	1	2	0	1	6	10
	10	0	0	0	0	0	1	1	1	1	0	6	10
	RL	1	0	0	0	0	1	0	3	0	3	32	40
	$\sum$	4	8	2	3	7	7	7	9	7	9	77	140

We use the above algorithms to extract RL in three images of the Tianjin area. The specific comparison analysis is displayed in Tables 13–15. In images TA and TB, the MFH algorithm has the best comprehensive performance, in which the recall, $F1$-score, and the overall extraction accuracy of RL are all better than those of the other three algorithms. In image TA, 80% of the residential areas is extracted, the precision is 88.889%, the $F1$ is 84.211%, and the OA is 83.337%, as shown in bold values in Table 13. In image TB, all residential areas are retrieved, the precision is 72.727%, the comprehensive evaluation index $F1$ reaches 84.210%, and the overall extraction accuracy is 83.333% (see bold values in Table 14). The extraction accuracy of the MFH algorithm in image TC is not as high as that of the first two images. The recall of RL extraction is only 58.537%, and the precision is only 61.538%. $F1$ and OA are lower by ~24.211% and 15.335%, respectively. Although different data sources lead to a lower extraction accuracy, the overall experimental results indicate that the MFH algorithm improved the extraction accuracy of RL, reduced the misclassification rate, and effectively realized the extraction of RL.

The separability within the four types of RL is listed in Fig. 10. The values on the diagonal line of the confusion matrix represent the accuracy of each class. We mapped the value of classification to the color band to visualize the results. The green cells have a higher accuracy, whereas the yellow ones have lower values. The red cells indicate that the probability of being classified into this class is zero, that is, they can be distinguished. In the GLCM-FC algorithm (a), there is a case of misclassification among four types of RL, and the classification accuracy is

Table 13 Comparison analysis of three proposed algorithms in image TA.

Image	Algorithm	P (%)	R (%)	$F1$ (%)	OA (%)
TA	GLCM-FC	57.143	80.000	66.667	55.556
	IBOVW	100.000	30.000	46.154	61.111
	TBOVW	50.000	20.000	28.571	44.444
	MFH	88.889	**80.000**	**84.211**	**83.337**

Table 14 Comparison analysis of three proposed algorithms in image TB.

Image	Algorithm	*P* (%)	*R* (%)	*F*1 (%)	OA (%)
TB	GLCM-FC	50.000	87.500	63.636	55.556
	IBOVW	100.000	12.500	22.222	61.111
	TBOVW	100.000	12.500	22.222	61.111
	MFH	72.727	**100.00**	**84.210**	**83.333**

Table 15 Comparison analysis of three proposed algorithms in image TC.

Image	Algorithm	*P* (%)	*R* (%)	*F*1 (%)	OA (%)
TC	GLCM-FC	51.667	75.610	61.386	61.000
	IBOVW	78.571	53.658	63.768	75.000
	TBOVW	48.837	51.220	50.000	58.000
	MFH	61.538	58.537	60.000	68.000

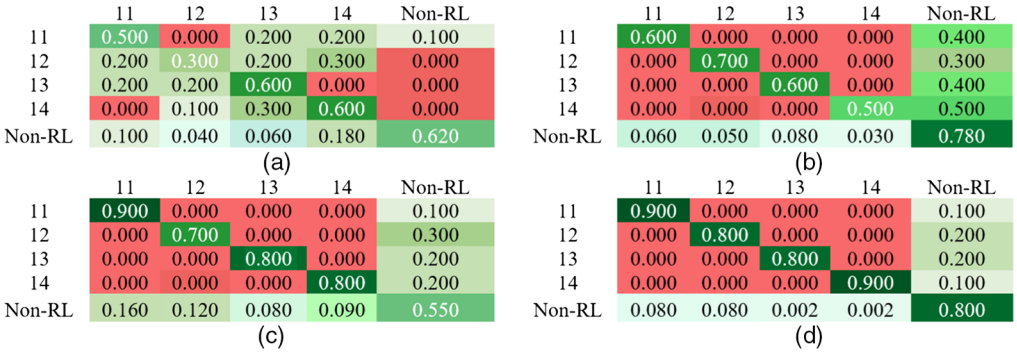

Fig. 10 The classification results of each type of RL: (a) the classification results of GLCM-FC, (b) the classification results of IBOVW, (c) the classification results of TBOVW, and (d) the classification results of MFH.

not ideal, which indicates that relying only the texture features cannot realize the distinction among RL types. Although there is no confusion within the four types of RL in the IBOVW (b) and TBOVW (c) experiments, more samples of non-RL are wrongly divided into RL. Compared with the other algorithms, the MFH algorithm has a higher classification accuracy, as class 11 and class 14 are both 90.000%, and classes 12/13 and non-RL are 80.000%. In the MFH algorithm, four types of RL can be distinguished from each other, and the number of misclassified samples is reduced, which provides some reference for the investigation and analysis of different types of RL.

4 Discussion

The experimental results indicate that the MFH algorithm shows a better performance in RL extraction in the sample dataset and large area images, and the classification of different types of RL. The main challenge of the algorithms is the distinction between RL and easily mixed categories, including farming facilities, industrial, PMPSs, and CSDs. There are many reasons for the confusion in these categories. First, the complexity of the ground objects in urban functional zones and the similar morphological structures are the main reasons. Additionally, the

geographical proximity of these functional zones also makes it difficult to extract RL. The image scale of 300 m × 300 m cannot achieve a good description for all categories, in which most of the image blocks are mixed ground objects, and the boundary areas connecting different types of LULC are also one of main reasons for misclassification. The algorithm can achieve better extraction results for the sample dataset and large area images from Google Earth while the extraction accuracy in aerial images is lower, which is related to the representativeness of the training samples, complex ground objects, image scales, and different data sources. The different experimental results in the two data sources indicate that the applicability of this algorithm needs to be improved to achieve good generalization.

The next step is to realize the distinction between the confused categories and RL. In addition, the classification accuracy of different types of RL also needs to be improved. Multiple features extracted from multisource data should be considered to achieve the differentiation among the four types of RL and non-RL, such as the elevation information provided by LiDAR data, geographic information provided by the Google street view map, and so on. As we want to achieve the extraction of LULC in large range areas, the algorithm still needs to be enhanced in the universality of images covering large areas. Meanwhile, the applicability of the algorithm to different data sources remains to be studied. Furthermore, the algorithm needs to be studied in terms of the selection and limitations of the parameters. The algorithm retrieves image blocks with a size of 300 m × 300 m in a 1-m resolution image; that is, only the spatial resolution and image block size parameters are constrained. The factors that affect the information extraction are extremely complicated, and different parameters will result in different extraction accuracies. Different solar elevation angles and atmospheric conditions will also affect the feature extraction of the image. In the next step, the algorithm will consider distinguishing different features at different scales and restricting more parameters to achieve fast and efficient automatic retrieval of RL. Our algorithm will be applied to a large area of Google Earth imagery to achieve information extraction of target objects quickly and automatically, providing reference information for land use and urban planning.

5 Conclusion

In this study, we proposed an MFH algorithm for RL extraction. The GLCM texture features and fuzzy classification strategy were first designed to extract RL. The results indicated that this method can differentiate classes 1/2/3/4/6 from RL. As this approach is very flexible, in which each class can be well extracted by only using feature characterization, features with more scales and texture measures can be included to improve the results in the future. Furthermore, only samples of the target classes were trained in this algorithm, avoiding the interference of other categories. Both the sample dataset and large area images were used to verify the algorithm. The algorithm has a higher recall rate, whereas the precision is not ideal, and the misclassification is serious. Therefore, an improved BOVW model based on SIFT features was further proposed to classify the undifferentiated categories. In this algorithm, we boosted the traditional BOVW model by proposing a method to generate the vocabulary. The extraction accuracy of RL was improved, and the rate of misclassification was significantly deduced. The MFH algorithm was used to further analyze the internal separability of the four types of RL. The results show that the four types of RL can be distinguished from each other, and the number of samples wrongly divided into RL is reduced, which is helpful for a dynamic investigation and the analyses of different types of RL.

Acknowledgments

This paper was substantially supported by the National Key R&D Program of China (Grant No. 2017YFB0503004) and the National Natural Science Foundation of China (Project Nos. 41301525 and 41571440). We owe great appreciation to the anonymous reviewers for their critical, helpful, and constructive comments and suggestions. Disclosures: The authors declare no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose.

References

1. X. Y. Zhang et al., "Multiscale geoscene segmentation for extracting urban functional zones from VHR satellite images," *Remote Sens.* **10**(2), 281 (2018).
2. Z. R. Liu et al., "The study on the extraction of habitation based on the high resolution satellite images," *Image Technol.* **24**(1), 25–28 (2012).
3. Z. Fu et al., "Multiscale and multifeature segmentation of high-spatial resolution remote sensing images using superpixels with mutual optimal strategy," *Remote Sens.* **10**(8), 1289 (2018).
4. M. Janalipour and A. Mohammadzadeh, "A fuzzy-GA based decision making system for detecting damaged buildings from high-spatial resolution optical images," *Remote Sens.* **9**(4), 349–372 (2017).
5. W. Z. Shi, Z. Y. Mao, and J. Q. Liu, "Building area extraction from the high spatial resolution remote sensing imagery," *Earth Sci. Inf.* **12**(1), 19–29 (2019).
6. T. Mahmood et al., "A survey on block-based copy move image forgery detection techniques," in *11th IEEE Int. Conf. Emerg. Technol.*, Peshawar, Pakistan (2015).
7. B. Sirmacek and C. Unsalan, "Urban-area and building detection using SIFT key points and graph theory," *IEEE Trans. Geosci. Remote Sens.* **47**(4), 1156–1167 (2009).
8. X. Huang and L. Zhang, "Morphological building/shadow index for building extraction from high-resolution imagery over urban areas," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **5**(1), 161–172 (2012).
9. S. Das, T. T. Mirnalinee, and K. Varghese, "Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images," *IEEE Trans. Geosci. Remote Sens.* **49**(10), 3906–3931 (2011).
10. Z. Miao et al., "A semi-automatic method for road centerline extraction from VHR images," *IEEE Geosci. Remote Sens. Lett.* **11**(11), 1856–1860 (2014).
11. Z. Lv et al., "An adaptive multifeature sparsity-based model for semiautomatic road extraction from high-resolution satellite images in urban areas," *IEEE Geosci. Remote Sens. Lett.* **14**(8), 1238–1242 (2017).
12. H. Pan et al., "An adaptive multifeature method for semiautomatic road extraction from high-resolution stereo mapping satellite images," *IEEE Geosci. Remote Sens. Lett.* **16**(2), 201–205 (2018).
13. S. Valero et al., "Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images," *Pattern Recognit. Lett.* **31**(10), 1120–1127 (2010).
14. M. Wulder, K. O. Niemann, and D. G. Goodenough, "Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery," *Remote Sens. Environ.* **73**(1), 103–114 (2000).
15. D. A. Pouliot et al., "Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration," *Remote Sens. Environ.* **82**(2), 322–334 (2002).
16. D. S. Culvenor, "TIDA: an algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery," *Comput. Geosci.* **28**(1), 33–44 (2002).
17. L. Wang, P. Gong, and G. S. Biging, "Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery," *Photogramm. Eng. Remote Sens.* **70**(3), 351–357 (2004).
18. Y. X. Liu et al., "High quality prime farmland extraction pattern based on object-oriented image analysis," in *Proc. Geoinf. 2008 and Joint Conf. GIS and Built Environ.: Classif. Remote Sens. Images*, SPIE, Guangzhou, China, pp. 71470–71471 (2008).
19. J. Shen et al., "Cropland extraction from very high spatial resolution satellite imagery by object-based classification using improved mean shift and one-class support vector machines," *Sens. Lett.* **9**(3), 997–1005 (2011).
20. X. D. Sun and H. Q. Xu, "Comer extraction algorithm for high-resolution imagery of agricultural land," *Trans. Chin. Soc. Agric. Eng.* **25**(10), 135–141 (2009).
21. L. Ma et al., "Cultivated land information extraction from high-resolution unmanned aerial vehicle imagery data," *J. Appl. Remote Sens.* **8**(1), 083673 (2014).

22. X. Y. Zhang et al., "Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data," *ISPRS. J. Photogramm. Remote Sens.* **132**, 170–184 (2017).
23. X. Y. Zhang et al., "Integrating bottom-up classification and top-down feedback for improving urban land-cover and functional-zone mapping," *Remote Sens. Environ.* **212**, 231–248 (2018).
24. M. R. Boutell et al., "Learning multi-label scene classification," *Pattern Recognit.* **37**(9), 1757–1771 (2004).
25. E. Farahzadeh, T. J. Cham, and W. Q. Li, "Semantic and spatial content fusion for scene recognition," Chapter 3 in *New Development in Robot Vision*, Y. Sun, A. Behal, and C. K. R. Chung, Eds., Springer, Berlin, Heidelberg, pp. 33–53 (2015).
26. J. D. Sun et al., "Image retrieval based on color distribution entropy," *Pattern Recognit. Lett.* **27**(10), 1122–1126 (2006).
27. L. C. Sim, H. Schroder, and G. Leedham, "Fast line detection using major line removal morphological Hough transform," in *Proc. of the 9th Int. Conf. Neural Inf. Process.*, 4, 2127–2131 (2002).
28. D. S. Zhang and G. J. Lu, "Shape-based image retrieval using generic Fourier descriptor," *Signal Proc. Image Commun.* **17**(10), 825–848 (2002).
29. A. A. Popescu, I. Gavat, and M. Datcu, "Contextual descriptors for scene classes in very high resolution SAR images," *IEEE Geosci. Remote Sens. Lett.* **9**(1), 80–84 (2012).
30. A. C. Berg and J. Malik, "Geometric blur for template matching," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, pp. 607–614 (2001).
31. L. F. Posada et al., "Semantic classification of scenes and places with omnidirectional vision," in *Proc. IEEE Eur. Conf. Mob. Rob.*, pp. 113–118 (2013).
32. C. Couprie et al., "Indoor semantic segmentation using depth information," in *Proc. 1st Int. Conf. Learn. Represent.*, pp. 1–8 (2013).
33. A. Payne and S. Singh, "Indoor vs. outdoor scene classification in digital photographs," *Pattern Recognit.* **38**(10), 1533–1545 (2005).
34. N. Serrano, A. Savakis, and J. Luo, "A computationally efficient approach to indoor/outdoor scene classification," *Pattern Recognit.* **37**(9), 1773–1784 (2004).
35. T. S. Pan et al., "Rapid semantic categorization of scenes: an advantage for emotional images," *J. Vis.* **13**(9), 1312–1312 (2013).
36. N. Ikizler-Cinbis and S. Sclaroff, "Object, scene and actions: combining multiple features for human action recognition," in *Proc. Eur. Conf. Comput. Vis.*, pp. 494–507 (2010).
37. L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2(2), 524–531 (2005).
38. D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Kerkyra, Greece (1999).
39. B. Sirmacek and C. Unsalan, "Urban area detection using local feature points and spatial voting," *IEEE Geosci. Remote Sens. Lett.* **7**(1), 146–150 (2010).
40. B. Sirmacek and C. Unsalan, "A probabilistic framework to detect buildings in aerial and satellite images," *IEEE Trans. Geosci. Remote Sens.* **49**(1), 211–221 (2011).
41. Y. Yang and S. Newsam, "Geographic image retrieval using local invariant features," *IEEE Trans. Geosci. Remote Sens.* **51**(2), 818–832 (2013).
42. X. Y. Zhang, S. Du, and Y. C. Wang, "Semantic classification of heterogeneous urban scenes using intrascene feature similarity and interscene semantic dependency," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **8**(5), 2005–2014 (2015).
43. D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
44. D. M. Blei, "Probabilistic models of text and images," PhD Dissertation, University of California, Berkeley, CA, USA (2004).
45. Y. Yang and S. Newsam, "Comparing SIFT descriptors and Gabor texture features for classification of remote sensed imagery," in *IEEE Int. Conf. Image Proc.* (2008).
46. G. S. Xia et al., "AID: a benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.* **55**(7), 3965–3981 (2017).

47. L. J. Zhao, P. Tang, and L. Z. Huo, "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).
48. W. Zhang et al., "Parcel-based urban land use classification in megacity using airborne LiDAR, high resolution orthoimagery, and Google Street View," *Comput. Environ. Urban Syst.* **64**, 215–228 (2017).
49. Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. ACM SIGSPATIAL GIS.*, pp. 270–279 (2010).
50. Q. Zhu et al., "Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.* **13**(6), 747–751 (2017).
51. G. Sheng et al., "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *Int. J. Remote Sens.* **33**(8), 2395–2412 (2012).
52. Q. Hu et al., "Exploring the use of Google earth imagery and object-based methods in land use/cover mapping," *Remote Sens*. **5**(11), 6026–6042 (2013).
53. Q. Zou et al., "Deep learning based feature selection for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.* **12**(11), 2321–2325 (2015).
54. H. Z. Leng et al., "Current land use classification (GB/T21010-2017)," online (2017), http://www.tdzyw.com/2017/1113/45597.html.
55. X. L. Liang et al., "Underdeveloped village extraction from high spatial resolution optical image based on GLCM textures and fuzzy classification," in *IEEE Int. Workshop Earth Obs. Remote Sens. Appl.*, Changsha, China (2014).
56. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **SMC-3**(6), 610–621 (1973).
57. X. Huang, "Multiscale texture and shape feature extraction and object-oriented classification for very high resolution remotely sensed imagery," PhD dissertation, Wuhan Univeisity, Wuhan (2009).
58. A. K. Shackelford and C. H. Davis, "A fuzzy classification approach for high-resolution multispectral data over urban areas," *IEEE Int. Geosci. Remote Sens. Symp.* **41**(9), 1621–1623 (2002).
59. A. Turpin and F. Scholer, "User performance versus precision measures for simple search tasks," in *ACM Sigir Process.*, New York, pp. 11–18 (2016).
60. D. M. W. Powers, "Evaluation: from precision, recall and F-factor to ROC, informedness, markedness & correlation," *J. Mach. Learn. Technol.* **2**(1), 37–63 (2011).
61. N. Chinchor, "MUC-4 evaluation metrics," in *MUC4 '92 Proc. 4th Conf. Message Understanding*, McLean, Virginia (1992).

Zhongliang Fu received his BS, MS, and PhD degrees in photogrammetry and remote sensing from Wuhan University in 1985, 1988, and 1996, respectively. He is a professor at the School of Remote Sensing and Information Engineering Wuhan University, Wuhan, China. His research interests include spatial data management and updates, remote sensing image processing and analysis, 3-D GIS, video GIS, spatiotemporal big data analysis, multisource data processing and analysis, and smart cities.

Xiaoli Liang received her PhD from the School of Remote Sensing and Information Engineering Wuhan University, Wuhan, China. Her research interests include remote sensing image processing and analysis, point cloud 3-D scene target semantic segmentation based on deep learning, and indoor 3-D scene reconstruction based on point cloud.