Arab J Geosci (2019) 12: 446
https://doi.org/10.1007/s12517-019-4600-0

ORIGINAL PAPER

Change detection in high-resolution images based on feature importance and ensemble method

Xin Wang[1,2,3] · Peijun Du[1,2,3] · Sicong Liu[4] · Gang Lu[2] · Xiaoming Gao[2,5]

Received: 6 December 2018 / Accepted: 2 July 2019 / Published online: 17 July 2019

Abstract

A robust framework for change detection in high-resolution (HR) images which takes into account feature importance of morphological attribute profiles (APs) is proposed. Although APs allow the discriminative extraction of geometrical features, the selection of optimal filtering thresholds is always a challenging task. In order to address this problem, the importances of multi-attribute APs, which are generated by ranges of thresholds, are calculated in advance. The APs with higher importances are chosen as input features for change detectors. Furthermore, owing to the possible incomplete feature importance model caused by randomness and small size of selected training samples through random forest, an ensemble method is conducted by integrating different results to enhance the stability of the final output. Experimental results obtained by two pairs of bi-temporal HR images demonstrate the effectiveness of the proposed approach.

Keywords Change detection · High resolution (HR) · Morphological attribute profiles (APs) · Feature importance · Ensemble method

Introduction

Change detection is a process of finding changes or phenomenon by observing images at different times (Lu et al. 2004; Hussain et al. 2013). Given that the new generation of satellite sensors has increased the availability of images with higher geometrical resolution, remote sensing images allow us to recognize different kinds of complex structures within a scene. Hence, the spatial information of high-resolution (HR) images becomes very important, in addition to spectral information which may be limited compared with multi-spectral medium-resolution images (Bruzzone and Carlin 2006).

With the increasing spatial resolution of remote sensing imagery, traditional features, such as spectral indexes and statistics (Frick and Tervooren 2019), texture features such as GLCM (Haralick et al. 1973; Hall-Beyer 2017), Gabor features (Angelo and Haertel 2003), and wavelet transform (Weeks and Bayoumi 2002; Qian et al. 2013), are not sufficient to describe the details of HR images. In recent years, morphological attribute profiles (APs), which are composed by performing the filtering with different thresholds, have been used as a powerful tool for extracting spatial information (Dalla Mura et al. 2010). APs can model the spatial information of the connectivity regions. Numerous studies have used APs for remote sensing classification, e.g., hyperspectral images (Aptoula et al. 2016), very high–resolution images (Demir and Bruzzone 2016), SAR images (Marpu et al. 2011), and LiDAR data (Pedergnana et al. 2012). For change detection, APs of area attribute derived from panchromatic image were used, and good results were achieved (Falco et al. 2013). However, many thresholds should be usually used for APs in order

Editorial handling: Hongchao Fan

✉ Peijun Du
dupjrs@gmail.com

1 School of Geography and Ocean Science, Nanjing University, Nanjing, China

2 Key Laboratory for Satellite Mapping Technology and Applications of National Administration of Surveying, Mapping and Geoinformation of China, Nanjing University, Nanjing, China

3 Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China

4 College of Surveying and Geoinformatics, Tongji University, Shanghai, China

5 Land Satellite Remote Sensing Application Center, Ministry of Natural Resource of the People's Republic of China, Beijing, China

to include all the possible structures in the image, and high-dimensional features were inevitably produced, which might be highly related with each other.

Filtering thresholds are critical parameters in constructing APs, and the optimal thresholds are difficult to define. Three methods are often used to construct APs: (1) Original APs created by tuned series of thresholds: As mentioned before, a wide range of thresholds are tuned to cover as many scales of the attributes as possible to structure the spatial features in the image. As a consequence, the high-dimensional APs are generated and difficult to deal with (Dalla Mura et al. 2010, 2011; Falco et al. 2013); (2) Feature extraction (FE) of APs (namely FEAPs): FE can be explained as finding a set of features which represents an observation while reducing the dimensionality (Ghamisi et al. 2014). In this method, APs acquired from multiple thresholds are conducted by FE method to cut down the dimensionality; (3) Ensemble based on random thresholds (namely ERTAPs) method: APs, which are generated by randomly selected thresholds, are used as the input of detectors, and multiple results are then integrated via a decision-level fusion method to generate the final result (Bao et al. 2016). This method can address the problem of inappropriate threshold selection through the fusing process. However, some problems still exist: (1) The thresholds of each single change detection process are not reasonable enough due to randomness in threshold selection process; (2) The importances of APs generated by one threshold through thickening and thinning operations are not always equal in different change detection task, which means that they may be not necessary for a change detection task in the same time.

Although an effective method may be used in some studies, the trained model is not optimal and representative, and

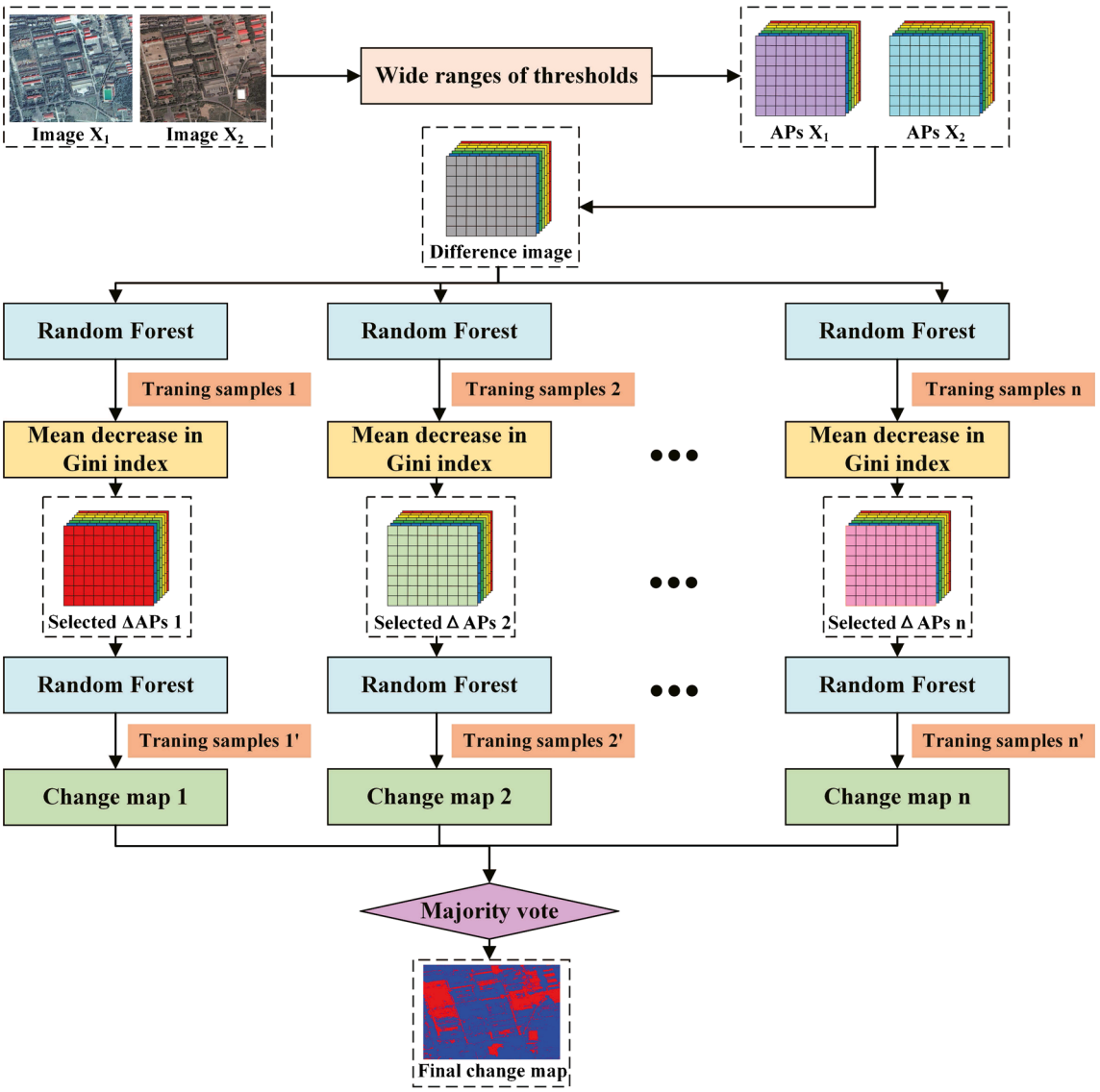

Fig. 1 Framework of the proposed change detection approach based on EITAPs

the result is unstable due to the available different training samples and the complex scenes. Their performances are limited by the advantages and disadvantages of the generated training models. Hence, it is necessary to randomly select samples to train the models for many times and then integrate the multiple results obtained by different training models to optimize their advantages and finally achieve the effective and stable results (Du et al. 2013).

A novel change detection approach using APs based on feature importance and ensemble method (namely EITAPs) is proposed for HR remote sensing images. Wide ranges of thresholds are defined for each attribute to construct a high-dimensional difference AP image in advance as in previous work (Falco et al. 2013). The APs with higher importance in each attribute are selected and combined to rebuild a new feature set for change detection. Finally, an ensemble strategy combining multiple results is used to avoid an incomplete feature importance model of the whole scene constructed by selecting a small-size training sample in each single change detection process. Compared with the aforementioned methods, the proposed approach not only reduces the dimensionality of APs but also solves the problem of randomness and using both APs generated by one threshold in each detector in the ERTAPs method.

Change detection approach using APs based on feature importance and ensemble method

A mean decrease in the Gini index is an effective indicator for describing the feature importance and can be used for selecting the useful variables from high-dimensional features such as APs for supervised classification and change detection. Ensemble methods can use a set of results generated from different training samples to learn partial solutions for a problem and then integrate them in some manner to construct a final solution for the original task. Taking advantages of these methods, we proposed a novel change detection approach, whose framework is shown in Fig. 1. This scheme is based on four steps: (1) APs with four kinds of attributes of bi-temporal images are calculated by wide ranges of thresholds, respectively. (2) Image differencing is conducted to generate the difference APs which can model the change information of bi-temporal images. (3) Importances of difference APs are figured out via random forest (RF), and difference APs with higher importances are selected to reestablish a new feature set, which is used as the input of RF to obtain the change detection result. (4) The multiple results, which are generated by repeating step 3, are integrated using *majority voting* rule to generate the final result.

Morphological attribute profiles

APs are generalization of conventional morphological profiles (MPs) and model the features rather than the sizes of objects to overcome the limitations of MPs (Dalla Mura et al. 2010). APs perform a contextual analysis of image considering measures computed on the regions. This permits obtaining a richer description of the regions in the scene since the filtering is performed according to measurement of their spectral, spatial, textural, and other characteristics. The definition of APs is based on the concept of granulometry, in case of thickening operation, and antigranulometry, in case of thinning operation,

$$\mathrm{AP}(f) = \{\phi^{\lambda_\varepsilon}(f), \phi^{\lambda_\varepsilon - 1}(f), ..., \phi^{\lambda_1}(f), f, \gamma^{\lambda_1}(f), ..., \gamma^{\lambda_{\varepsilon-1}}(f), \gamma^{\lambda_\varepsilon}(f)\} \quad (1)$$

where ϕ^i and γ^i represent the thickening and thinning operations with a given sequence of thresholds $\{\lambda_1, \lambda_2, ..., \lambda_\varepsilon\}$ (Dalla Mura et al. 2010).

At present, the attributes commonly used in AP filtering are as follows: (1) s, standard deviation of the gray-level values of the pixels in the regions; (2) a, the area of connectivity area; (3) d, length of the diagonal of the box bounding the region; (4) i, moment of inertia. The area of the connectivity area and the length of the diagonal of the box bounding the region measure the size and shape of the connected areas. Standard deviation of the gray-level values describes the homogeneity of the pixel gray value in the connected regions, and the moment of inertia describes the non-compactness of the connected regions (Dalla Mura et al. 2010).

When the information related to spatial context is included, profiles of unchanged regions in bi-temporal images will show similar behaviors, whereas profiles related to the changed areas will be characterized by different behaviors. Therefore, the difference AP vectors, which are described as $\delta\mathrm{AP} = \mathrm{AP}(f_2) - \mathrm{AP}(f_1)$, are generated to describe the differences of bi-temporal images from time 1 to time 2, where $\mathrm{AP}(f_1)$ and $\mathrm{AP}(f_2)$ are bi-temporal images with APs. As all unchanged pixels result in similar differences (with $\delta\mathrm{AP} \approx 0$) , the class of such pixels cannot be modeled. On the contrary, those showing difference AP vectors far from 0 in at least one feature have high probability of transitions in ground cover. In addition, normalization is applied to APs, which aims at reducing the effects of data expression in different attributes.

Importance-based feature selection through random forest

RF is a machine learning classification method which can deal with a large quantity of input variables for remote sensing image classification and change detection (Belgiu and Drăguţ 2016; Duro et al. 2012). It combines a series of tree-structured classifiers where each classifier contributes with a single vote for the assignation of the most frequent class to the input data (Rodriguez-Galiano et al. 2012). RF can also be used as an embedded method for feature selection (Tuia et al. 2009). It provides an assessment of the relative importance of the different features or variables during the classification process. This aspect is useful for high-dimensional data to know how each predictive variable influences the classification model and selects the better variables. In the process of estimating feature importance, the RF changes one of the input random variables while keeping the others constant to measure the decrease in accuracy through the out-of-bag error estimation or the Gini index decrease (Breiman 2001). The Gini index is a standard impurity measure which can avoid variable selection bias. It is described as:

$$\mathrm{Gini}(m) = 1 - \sum_{i=1}^{L} {p_{wi}}^2 \tag{2}$$

where L is the number of classes and p_{wi} is the probability or relative frequency of class wi at node m (Archer and Kimes 2008). The mean decrease in the Gini index is a measurement of how each variable contributes to the homogeneity of the nodes and leaves in the resulting RF. Each time a particular variable is used to split a node, the Gini index for the child nodes is calculated and compared with that of the original node. The changes in the Gini index are summed for each variable and normalized at the end of the calculation. Variables that result in nodes with higher purity have a higher decrease in the Gini index. Therefore, it becomes easy to evaluate the performance of input features in change detection process by using the mean decrease in the Gini index.

The difference APs are used to describe the information of changes. In order to cover the comprehensive features of changes from small scale to large scale, a series of thresholds are used to calculate the APs. In this case, high-dimensional difference APs are produced. Hundreds of generated input features may be highly relevant and redundant for a task. Under this circumstances, the feature selection method through RF, which can sort the input variables by their contributions to the result, is beneficial to enhance accuracy. Therefore, the bi-temporal difference APs with higher importance in each attribute are selected and combined as a new input feature set.

Ensemble method

In the context of pattern recognition field, there is no guarantee that one specific classification process will always achieve the optimal performance in different condition. However, considering multiple operational processes can achieve a better predictive performance than any single result. In other words, an ensemble method can integrate multiple results to achieve a better result and increase generalization capability (Oza and Tumer 2008; Healey et al. 2018). In the proposed method, after multiple results are generated from repeating step 3 with different training samples for several times, an ensemble method called majority vote is applied, in which the most frequent produced label in each pixel is determined as the final result of change detection (Du et al. 2012; Wang et al. 2018).

Results and analysis

Dataset description

For the sake of comparing the proposed approach with other existing strategies, the experimental analysis was conducted with two real HR remote sensing images, and

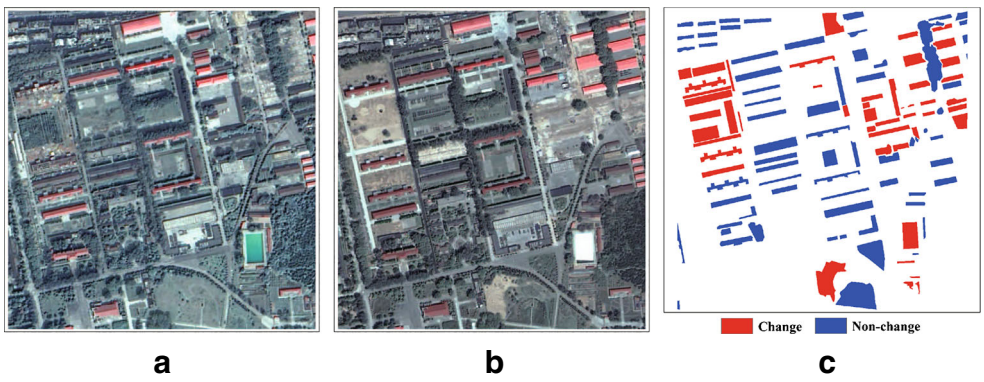

Fig. 2 True color composite images of Xuzhou dataset acquired on **a** 15 September 2004 and **b** 2 May 2005, and **c** standard reference change map of Xuzhou dataset

the performances are evaluated in terms of change detection accuracies.

Xuzhou: The first dataset was acquired by QuickBird satellite on 15 September 2004 and 2 May 2005 covering the same area of Xuzhou City, China. After data preprocessing, it is composed of 1070 × 1035 pixels with spatial resolution of 0.61 m. In order to perform a quantitative analysis of the results, the ground reference maps were defined according to accurate manual photointerpretation of the images with prior knowledge. Figure 2 depicts the true color composite images and standard reference change map, which contains 84,736 changed pixels and 121,101 unchanged pixels.

Jiangyin: The second dataset were acquired by Ziyuan-3 and Gaofen-1 at the same location of Jiangyin County, China, on 10 February 2015 and 27 March 2016. Although they were acquired by different satellites, they share the same bands with the same spectral range in each channel of four multi-spectral bands and thus can be used for the image to image change detection process. After data preprocessing, they were composed of 800 × 880 pixels with spatial resolution of 2 m. In order to perform a quantitative analysis of the results, the ground reference map was defined according to accurate manual photointerpretation of the images with prior knowledge. Figure 3 depicts the true color composite images and standard reference change map, which contains 24,586 changed pixels and 42,704 unchanged pixels.

Experimental setup

Four different attributes with defined ranges of thresholds were considered to generate the original APs according to the automatic scheme in Ghamisi et al. (2014) and prior knowledge of the datasets: (1) s ($\lambda_s = u/100 \times \{\sigma_{\min}, \sigma_{\min} + \delta_s, \sigma_{\min} + 2\delta_s, ..., \sigma_{\max}\}$); (2) a ($\lambda_a = 1000/v \times \{a_{\min}, a_{\min} + \delta_a, a_{\min} + 2\delta_a, ..., a_{\max}\}$); (3) d ($\lambda_d = [5, 10, 15, ..., 100]$); (4) i ($\lambda_i = [0.24, 0.28, 0.32, ..., 1.00]$), where u was the mean of the bands of bi-temporal images and $\sigma_{\min}$, $\sigma_{\max}$, and δ_s were 0.15, 3, and 0.15 and 0.25, 5, and 0.25 for Xuzhou and Jiangyin, respectively, where v was the spatial resolution of the input data and $a_{\min}$, $a_{\max}$, and δ_a were 0.075, 1.5, and 0.075 and 0.25, 5, and 0.25, for Xuzhou and Jiangyin, respectively. In the proposed approach, randomly selected samples with the previously generated original APs were put into RF to figure out their importances at first. Then, the most important APs, whose quantity was equal to the number generated by two thresholds, were chosen in each attribute and combined to construct a new feature set for the global detection. This process were repeated 10 times with different randomly selected training samples, and the results were integrated to generate the final change map.

In order to validate the effectiveness of the proposed method, (1) four original spectral bands, (2) original APs, (3) FEAPs, and (4) ERTAPs were conducted with RF as comparisons. In the FEAPs, kernel principle component analysis (KPCA) (Cao et al. 2003), and discriminant analysis feature extraction (DAFE) (Belhumeur et al. 1997) were used as unsupervised and supervised FE methods to reduce the dimensionality of APs. The Gaussian kernel function was used in the KPCA. In order to be consistent with the proposed method, the dimension of the output features of KPCA was set to 64. In DAFE, The maximum dimension of output features was equal to $L - 1$ where L was the number of classes. Therefore, only one feature was generated after DAFE in binary change detection. In ERTAPs, two thresholds are randomly selected from the same defined range in each attribute to construct 64-dimensional APs for each RF classifier, and multiple results were then integrated. In order to compare the performance of RF in dealing with APs, SVM, which has intrinsic robustness to high-dimensional datasets and ill-posed problems (Volpi et al. 2013), was conducted for comparison based on original APs. In the experiments, 1000 training samples were randomly selected from the standard reference change map each time for the first feature importance calculation step and the second change detection step. The number of the trees and the features in a subset in RF were set to be 10. The type of kernel function used in SVM was RBF. The experimental results of all the listed methods were achieved by the mean of 10 Monte Carlo runs.

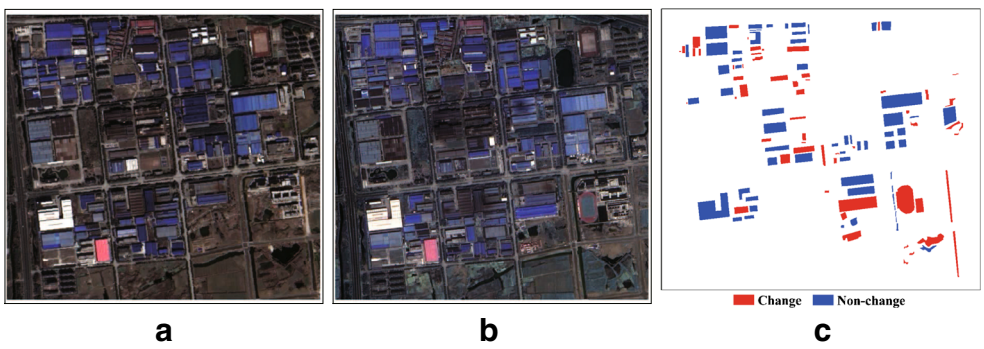

Fig. 3 The true color composite images of Jiangyin dataset acquired on **a** 10 February 2015 and **b** 27 March 2016, and **c** standard reference change map of Jiangyin dataset

Table 1 Change detection accuracies obtained by the proposed approach and the reference methods in Xuzhou dataset

	Spec	APs	APs_{svm}	KPCA	DAFE	ERTAPs	EITAPs
UC (%)	90.37	96.58	98.15	91.20	96.25	97.89	**98.91**
CH (%)	92.54	96.98	95.86	92.28	95.95	97.20	**97.66**
OA (%)	91.27	96.74	97.21	91.64	96.13	97.60	**98.39**
AA (%)	91.46	96.78	97.00	91.74	96.10	97.54	**98.28**
Kappa	0.8213	0.9329	0.9422	0.8287	0.9203	0.9505	**0.9668**
CE (%)	12.96	4.81	2.69	12.02	5.29	3.01	**1.57**
OE (%)	7.46	3.02	4.14	7.72	4.05	2.80	**2.34**

Note that UC and CH indicate the class accuracies of unchanged and changed areas

The bold characters represent the best result among the different methods in each evaluation index

Results

Results of Xuzhou

The change detection results in Xuzhou are reported in Table 1. With the comparison of the reference schemes, the results of the proposed approach achieve the highest overall accuracy (OA). Meanwhile, they keep the least commission error (CE) and omission error (OE) in the same time.

The importances of APs in different attributes are displayed in Fig. 4. They vary greatly with different thresholds and some of them are very small, which means that they are redundant and make little contribution to the result. Besides that, the importance of APs created through opening and thinning operations by one threshold is indeed not similar all the time. In this condition, the proposed method can take its advantage of selecting the more important AP separately to avoid having to use both features. The mean and standard deviations of importances are calculated (i.e., $s = 0.91 \pm 1.64$, $a = 0.86 \pm 1.57$, $d = 1.09 \pm 1.86$, $i = 0.26 \pm 0.40$.). It can be observed that the values of attributes s, a, and d are obviously higher than those of i. That is to say, the attributes s, a, and d are more important and fluctuant internally than i. Change detection maps of the different schemes are listed in Fig. 5, in which some improvements obtained by the proposed method are marked with yellow rectangles.

Results of Jiangyin

Table 2 shows the change detection results of Jiangyin. Compared with the reference methods, the OA is the highest and the CE and OE are the lowest in the proposed approach. These results indicate that the proposed approach makes better performances than traditional methods of bi-temporal HR image change detection with APs.

The importances of APs are displayed in Fig. 6. The following observations can be seen in the figure: (1) APs generated by different thresholds have significantly different magnitudes of importance. (2) The importances of APs generated through one threshold are not always equal for a task. The feature with small importance should be split out and removed separately. (3) The mean and standard deviation of importances in each attribute of APs are calculated (i.e., $s = 0.71 \pm 0.86$, $a = 0.97 \pm 1.25$, $d = 1.21 \pm 1.52$, $i = 0.23 \pm 0.28$.). The values of attributes s, a, and d are more higher than i, which indicates that the attributes s, a, and d are more important and fluctuant

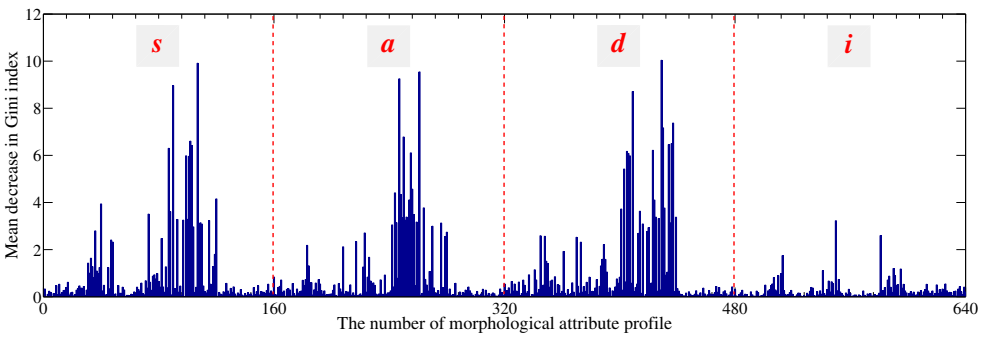

Fig. 4 Feature importance of APs in Xuzhou dataset by measuring the mean decrease in the Gini index. No.1-160, No.161-320, No.321-480, and No.481-640 of X axes represent the attributes s, a, d, and i of APs, respectively

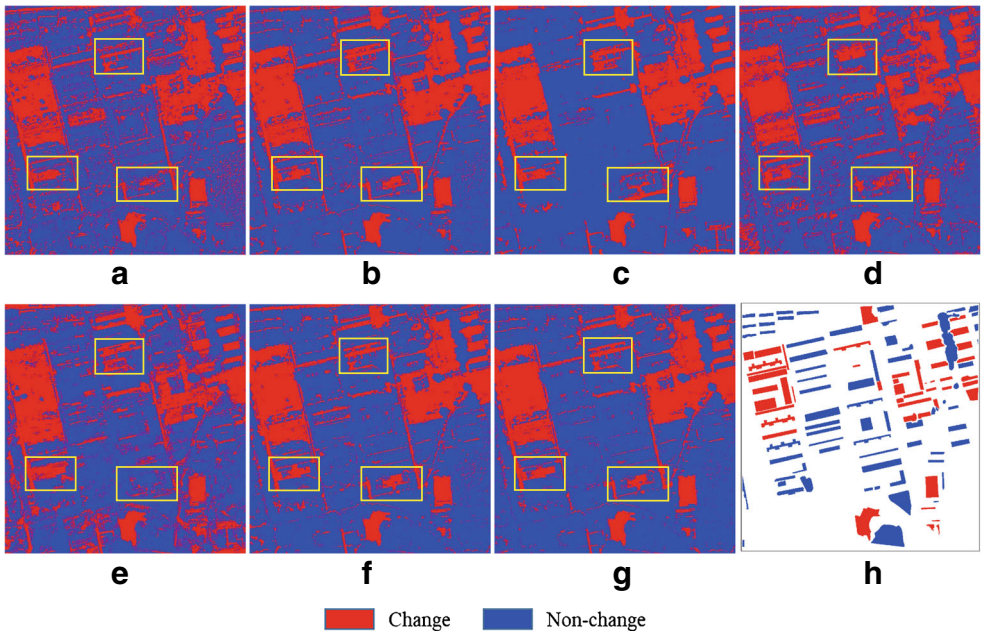

Fig. 5 Change detection maps of Xuzhou dataset. The results obtained by using RF with **a** spectral bands, **b** original APs, feature reduction of original APs with **d** KPCA and **e** DAFE, **f** ERTAPs, and **g** EITAPs, respectively. **c** is the result generated by SVM with original APs. **h** is the standard reference change map

Table 2 Change detection accuracies obtained by the proposed approach and the reference methods in Jiangyin dataset

	Spec	APs	APs_{svm}	KPCA	DAFE	ERTAPs	EITAPs
UC (%)	94.35	97.34	97.29	95.91	94.75	98.02	**98.11**
CH (%)	93.53	96.29	92.80	96.04	95.04	96.91	**97.04**
OA (%)	94.05	96.96	95.66	95.95	94.86	97.62	**97.72**
AA (%)	93.94	96.81	95.05	95.97	94.90	97.47	**97.57**
Kappa	0.8723	0.9344	0.9057	0.9132	0.8899	0.9486	**0.9508**
CE (%)	9.57	4.61	4.86	6.95	8.82	3.42	**3.27**
OE (%)	6.47	3.71	7.20	3.96	4.96	3.09	**2.96**

The bold characters represent the best result among the different methods in each evaluation index

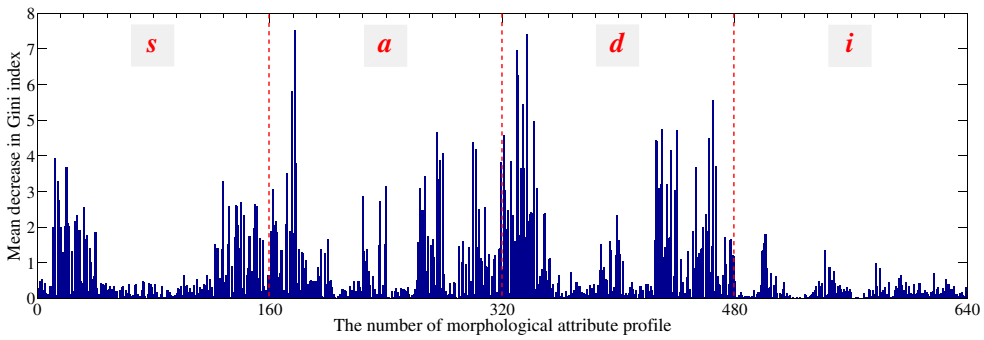

Fig. 6 Feature importance of APs in Jiangyin dataset by measuring the mean decrease in the Gini index. No.1-160, No.161-320, No.321-480, and No.481-640 of *X* axes represent the attributes *s*, *a*, *d*, and *i* of APs, respectively

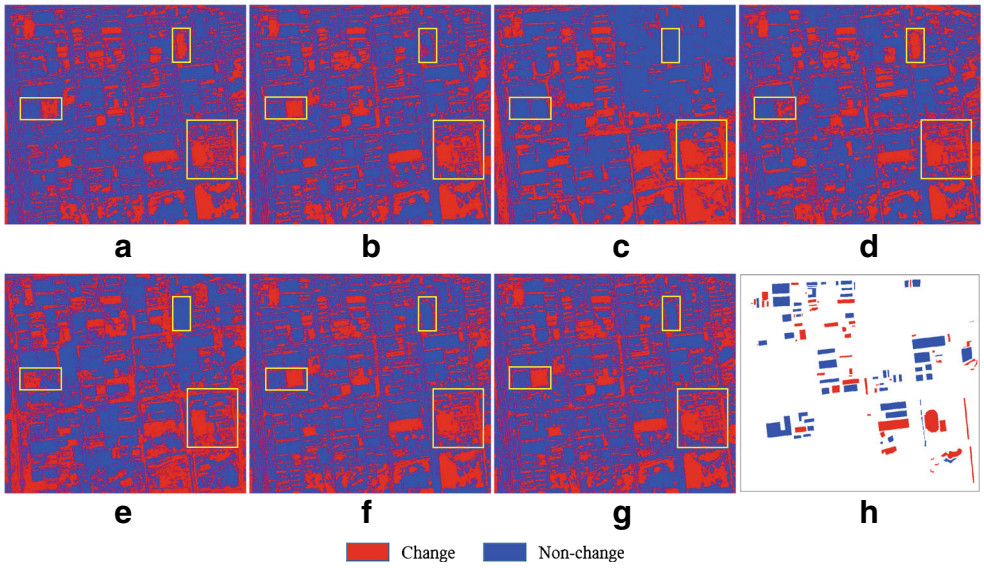

Fig. 7 Change detection maps of Jiangyin dataset. The results obtained by using RF with **a** spectral bands, **b** original APs, feature reduction of original APs with **d** KPCA and **e** DAFE, **f** ERTAPs, and **g** EITAPs, respectively. **c** is the result generated by SVM with original APs. **h** is the standard reference change map

internally than *i*. Figure 7 shows the change detection maps of the different schemes. The yellow rectangles in the figures show the results that have been improved.

Discussion

Table 3 shows the time consumption of the methods used in the experiments. In the feature reduction methods, the ERTAPs and EITAPs consumed more time than the feature reduction of APs with KPCA and DAFE. Actually, the ERTAPs and EITAPs were both ensemble methods, in which the detecting process was repeated in different conditions. Furthermore, RF was needed twice during each independent detecting process in the EITAPs method, and that was why it took double time as long as the ERTAPs method. From the aspects of detectors, the time consumed by SVM was huge in such the amount of feature dimension of APs and the accuracies of the results were no better than using RF.

Two parameters, which are the number of decision trees (*Ntree*) and the number of features in a subset for the best split when growing the trees (*Mtry*), are user defined in RF. For the sake of exploring their impacts on the EITAPs method, a series of *Mtry* and *Ntree* were tested for RF in the experiments. Tables 4 and 5 show the results of the EITAPs method with different *Ntree* and *Mtry* used in RF. Actually, it can been seen from the tables that the accuracies of the change detection results of the EITAPs are not sensitive to *Ntree* and *Mtry*. Moreover, in terms of the time consumption, the time consumed increases significantly as the number of *Ntree* increases, while the number of *Mtry* has no such notable effect on the time-consuming. The increase of decision trees will directly increase the operation time since each tree needs to be calculated and integrated in RF. The computation time will also increase as the *Mtry* increases because it has to compute the information gain contributed by all of the variables used to split the nodes of each decision tree. However, compared with *Ntree*, the computation time increase caused by *Mtry* is slight. Therefore, the EITAPs method can save the process of tuning the parameters of RF. That is to say, as long as a relatively small number of *Ntree* in the method is set, a high accuracy result with a less time-consuming process can be obtained.

In order to test the validity of the proposed method and figure out the influence of the threshold number in ensemble

Table 3 Time consumed by the proposed and reference methods in two datasets

	Spec	APs	APs_{svm}	KPCA	DAFE	ERTAPs	EITAPs
Xuzhou (s)	0.67	6.27	335.32	6.16	1.10	55.23	111.24
Jiangyin (s)	0.46	4.11	335.58	2.80	0.78	34.87	76.33

Table 4 Change detection results obtained with different numbers of *Ntree* with *Mtry*=10 in RF

Number of *Ntree*		10	25	50	75	100	200	500
Xuzhou	OA (%)	98.39	98.41	98.49	98.33	98.24	98.23	98.19
	time (s)	111.24	152.19	212.96	266.65	325.86	552.19	1301.78
Jiangyin	OA (%)	97.72	97.68	97.54	97.46	97.63	97.55	97.53
	time (s)	76.33	103.76	149.01	195.37	247.96	425.37	954.55

Table 5 Change detection results obtained with different numbers of *Mtry* with *Ntree* = 10 in RF

Number of *Mtry*		10	20	40	80	160	320	640
Xuzhou	OA (%)	98.39	98.39	98.39	98.37	98.43	98.46	98.52
	time (s)	111.24	119.68	117.90	119.18	119.87	115.17	123.12
Jiangyin	OA (%)	97.72	97.70	97.68	97.69	97.69	97.69	97.74
	time (s)	76.33	77.34	76.81	78.64	82.15	81.39	85.36

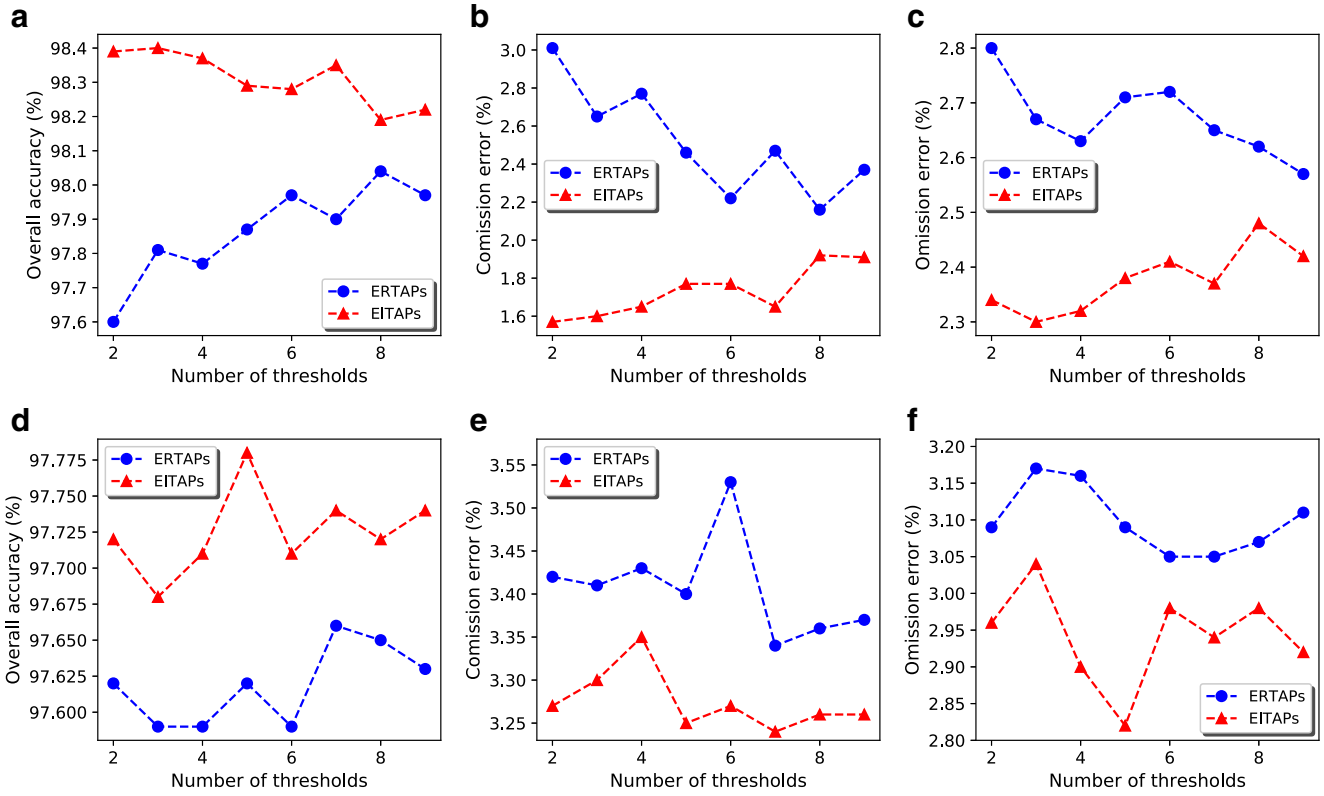

Fig. 8 Change detection OAs, CEs, and OEs with different numbers of thresholds selected in each attribute. **a–c** Xuzhou dataset. **d–f** Jiangyin dataset

methods, APs generated by two to nine thresholds in each attribute in the ERTAPs method and the corresponding numbers of APs in each attribute in the proposed method were used for two datasets (Fig. 8). The proposed approach always achieved the higher OA and the lower CE and OE in all conditions.

In addition, the proposed method is provided with the potential for multiple change detection in which the information of change type or direction in change region is calculated as well. With the training samples containing the change type information, the EITAPs can be used to calculate and select the important features for separating all the change types from the original APs. Liu et al. (2017) has used features obtained from morphological operations to detect the directions of change. The results confirmed that using morphological feature is feasible to detect different change classes. As mentioned before, APs have a better performance due to modeling the features of objects from different attributes rather than the size only compared with MPs. However, such research using APs has not been explored yet. Therefore, EITAPs may have potential for multiple change detection and we will explore its performance in the future work.

Conclusion

In this letter, a novel HR change detection approach using APs based on feature importance and ensemble method has been presented. The importances of APs, which were generated by measuring the mean decrease in the Gini index via RF, were various and asymmetric. Therefore, effective APs were selected based on high magnitude of feature importance to reduce dimensionality and achieve better result. Finally, multiple results obtained by using optimized APs were integrated to achieve the final result. Compared with using all APs or the other feature reduction methods, the proposed approach selected useful spatial features directly for the specific task to improve the discriminability of changed and unchanged areas, and the experimental results illustrated that it outperformed the other methods in terms of accuracy and stability.

References

Angelo NP, Haertel V (2003) On the application of gabor filtering in supervised image classification. Int J Remote Sens 24(10):2167–2189

Aptoula E, Ozdemir MC, Yanikoglu B (2016) Deep learning with attribute profiles for hyperspectral image classification. IEEE Geosci Remote Sens Lett 13(12):1970–1974

Archer KJ, Kimes RV (2008) Empirical characterization of random forest variable importance measures. Comput Statist Data Anal 52(4):2249–2260

Bao R, Xia J, Dalla Mura M, Du P, Chanussot J, Ren J (2016) Combining morphological attribute profiles via an ensemble method for hyperspectral image classification. IEEE Geosci Remote Sens Lett 13(3):359–363

Belgiu M, Drăguţ L (2016) Random forest in remote sensing: a review of applications and future directions. ISPRS J Photogramm Remote Sens 114:24–31

Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal& Mach Intell 7:711–720

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Bruzzone L, Carlin L (2006) A multilevel context-based system for classification of very high spatial resolution images. IEEE Trans Geosci Remote Sens 44(9):2587–2600

Cao L, Chua KS, Chong W, Lee H, Gu Q (2003) A comparison of pca, kpca and ica for dimensionality reduction in support vector machine. Neurocomputing 55(1–2):321–336

Dalla Mura M, Benediktsson JA, Waske B, Bruzzone L (2010) Morphological attribute profiles for the analysis of very high resolution images. IEEE Trans Geosci Remote Sens 48(10):3747–3762

Dalla Mura M, Villa A, Benediktsson JA, Chanussot J, Bruzzone L (2011) Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. IEEE Geosci Remote Sens Lett 8(3):542–546

Demir B, Bruzzone L (2016) Histogram-based attribute profiles for classification of very high resolution remote sensing images. IEEE Trans Geosci Remote Sens 54(4):2096–2107

Du P, Liu S, Gamba P, Tan K, Xia J (2012) Fusion of difference images for change detection over urban areas. IEEE J Selected Topics Appl Earth Observ Remote Sens 5(4):1076–1086

Du P, Liu S, Xia J, Zhao Y (2013) Information fusion techniques for change detection from multi-temporal remote sensing images. Inform Fus 14(1):19–27

Duro DC, Franklin SE, Dubé MG (2012) Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. Int J Remote Sens 33(14):4502–4526

Falco N, Mura MD, Bovolo F, Benediktsson JA, Bruzzone L (2013) Change detection in vhr images based on morphological attribute profiles. IEEE Geosci Remote Sens Lett 10(3):636–640

Frick A, Tervooren S (2019) A framework for the long-term monitoring of urban green volume based on multi-temporal and multi-sensoral remote sensing data. J Geovisual Spatial Anal 3(1):6

Ghamisi P, Benediktsson JA, Sveinsson JR (2014) Automatic spectral-spatial classification framework based on attribute profiles and supervised feature extraction. IEEE Trans Geosci Remote Sens 52(9):5771–5782

Hall-Beyer M (2017) Practical guidelines for choosing glcm textures to use in landscape classification tasks over a range of moderate spatial scales. Int J Remote Sens 38(5):1312–1338

Haralick RM, Shanmugam K et al (1973) Textural features for image classification. IEEE Trans Syst Man Cybern 6:610–621

Healey SP, Cohen WB, Yang Z, Brewer CK, Brooks EB, Gorelick N, Hernandez AJ, Huang C, Hughes MJ, Kennedy RE et al (2018) Mapping forest change using stacked generalization: an ensemble approach. Remote Sens Environ 204:717–728

Hussain M, Chen D, Cheng A, Wei H, Stanley D (2013) Change detection from remotely sensed images: from pixel-based to object-based approaches. ISPRS J Photogramm Remote Sens 80:91–106

Liu S, Du Q, Tong X, Samat A, Bruzzone L, Bovolo F (2017) Multiscale morphological compressed change vector analysis for unsupervised multiple change detection. IEEE J Selected Topics Appl Earth Observ Remote Sens 10(9):4124–4137

Lu D, Mausel P, Moran EF (2004) Change detection techniques. Int J Remote Sens 25(12):2365–2407

Marpu PR, Chen KS, Chu CY, Benediktsson JA (2011) Spectral-spatial classification of polarimetric sar data using morphological profiles. Proceedings of SPIE - The International Society for Optical Engineering 8180(1):547–556

Oza NC, Tumer K (2008) Classifier ensembles: select real-world applications. Inform Fus 9(1):4–20

Pedergnana M, Marpu PR, Mura MD, Benediktsson JA, Bruzzone L (2012) Classification of remote sensing optical and lidar data using extended attribute profiles. IEEE J Selected Topics Signal Process 6(7):856–865

Qian Y, Ye M, Zhou J (2013) Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. IEEE Trans Geosci Remote Sens 51(4):2276–2291

Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J Photogramm Remote Sens 67:93–104

Tuia D, Pacifici F, Kanevski M, Emery WJ (2009) Classification of very high spatial resolution imagery using mathematical morphology and support vector machines. IEEE Trans Geosci Remote Sens 47(11):3866–3879

Volpi M, Tuia D, Bovolo F, Kanevski M, Bruzzone L (2013) Supervised change detection in vhr images using contextual information and support vector machines. Int J Appl Earth Observ Geoinform 20:77–85

Wang X, Liu S, Du P, Liang H, Xia J, Li Y (2018) Object-based change detection in urban areas from high spatial resolution images based on multiple features and ensemble learning. Remote Sens 10(2):276

Weeks M, Bayoumi MA (2002) Three-dimensional discrete wavelet transform architectures. IEEE Trans Signal Process 50(8):2050–2063