A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion

Chen Wu [a,d], Bo Du [b], Xiaohui Cui [a], Liangpei Zhang [c,*]

[a] International School of Software, Wuhan University, Wuhan, PR China
[b] School of Computer, Wuhan University, Wuhan, PR China
[c] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, PR China
[d] Research Center of Spatial Information & Digital Engineering, International School of Software, Wuhan University, PR China

ARTICLE INFO

ABSTRACT

Post-classification with multi-temporal remote sensing images is one of the most popular change detection methods, providing the detailed "from-to" change information in real applications. However, due to the fact that it neglects the temporal correlation between corresponding pixels in multi-temporal images, the post-classification approach usually suffers from an accumulation of misclassification errors. In order to solve this problem, previous studies have separated the change and non-change candidates with change vector analysis, and they have only updated the classes of the changed pixels with the post-classification; however, this approach with thresholding loses the continuous change intensity information, where larger values indicate higher probability to be changed. Therefore, in this paper, a new post-classification method with iterative slow feature analysis (ISFA) and Bayesian soft fusion is proposed to obtain reliable and accurate change detection maps. The proposed method consists of three main steps: 1) independent classification is implemented to obtain the class probability for each image; 2) the ISFA algorithm is used to obtain the continuous change probability map of multi-temporal images, where the value of each pixel indicates the probability to be changed; and 3) based on Bayesian theory, the a posteriori probabilities for the class combinations of coupled pixels are calculated to integrate the class probability with the change probability, which is named as Bayesian soft fusion. The class combination with the maximum a posteriori probability is then determined as the change detection result. In addition, a class probability filter is proposed to avoid the false alarms caused by the spectral variation within the same class. Two experiments with multi-temporal Landsat Thematic Mapper (TM) images indicated that the proposed method achieves a clearly higher change detection accuracy than the current state-of-the-art methods. The proposed method based on Bayesian theory and ISFA was also verified to have the ability to improve the change detection rate and reduce the false alarms at the same time. Given its effectiveness and flexibility, the proposed method could be widely applied in land-use/land-cover change detection and monitoring at a large scale.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Land-use/land-cover (LULC) change detection is an important factor in the dynamic study of ecosystems and city development (Coppin et al., 2004). The locations and the "from-to" types of LULC changes can provide valuable information for environmental change and urban expansion monitoring (Bindschadler et al., 2010; Huang et al., 2016; Nielsen et al., 2008; Tang and Zhang, 2017; Xian and Homer, 2010; Yang et al., 2012). Timely and accurate LULC change detection is also important for a better understanding of the interactions between human activity and natural phenomena (Fu and Weng, 2016; Lu et al., 2004; Zhao et al., 2011; Zillmann et al., 2014).

Due to its large-scale view and long-period observation, remote sensing has been the primary data source for LULC change detection (Kennedy et al., 2009; Song et al., 2014; Yao et al., 2016; Zhang and Zhang, 2007; Zillmann et al., 2014). Numerous methods have addressed change detection using multi-temporal remote sensing data, including image algebra (Bovolo and Bruzzone, 2007; Chen et al., 2003), image transformation (Canty and Nielsen, 2008; Celik, 2009; Nielsen, 2007), and classification-based methods (Hussain et al., 2013; Singh, 1989). Among them, classification-based methods can provide the detailed "from-to" change type information in the study area, and have thus been widely applied in LULC change detection (Chen et al., 2012; Coppin et al., 2004; Hussain et al., 2013; Lu et al., 2004; Xian et al., 2009; Yuan et al., 2005).

There are two main types of methods in classification-based change detection: multi-date classification methods and post-classification

* Corresponding author.
E-mail address: zlp62@whu.edu.cn (L. Zhang).

methods (Hu and Zhang, 2013; Huang et al., 2010; Singh, 1989). Multi-date classification regards each "from-to" change type as one separate class, and performs a direct classification for the stacked data (Nemmour and Chibani, 2006). However, it may be difficult to obtain enough high-quality training samples for all the possible change types (Huang et al., 2010). Therefore, the post-classification methods, which compare the class maps from independent classification, are more popular in practical applications, (Ahlqvist, 2008; Serra et al., 2003; Yuan et al., 2005). The disadvantage of the post-classification methods is that they neglect the correlation information of the corresponding landscapes from multi-temporal images, and thus suffer from an accumulation of misclassification errors (Singh, 1989). From the perspective of change type identification, the misclassification in one image will lead to errors in the "from-to" change map, no matter whether or not the corresponding pixels in the other image are correctly classified (Singh, 1989). From the perspective of change/non-change detection, pixels with the same spectral features may be detected as changes due to the classification hyperplane variations in the multi-temporal images. Therefore, for the purpose of obtaining a better performance, integrating temporal correlation with independent classifications is a feasible and effective way to improve a post-classification method (Cardille and Fortin, 2016; Xian et al., 2009).

In a previous study, Xian et al. (2009) proposed an updated post-classification method based on a "hard fusion" approach, with the independent classification and change intensity obtained by change vector analysis (CVA). CVA is first implemented to separate the change and non-change candidate areas. For the unchanged areas, the landscape is assigned as the same class as that in the previous map, and the changed areas follow the independent classification (Xian and Homer, 2010). This straightforward approach can provide the opportunity to maintain the consistency of landscapes, and has been demonstrated to be effective in reducing false alarms (Yu et al., 2016). However, there are two problems with this method: 1) it is not easy to automatically and accurately determine the threshold to identify non-changes in previous works (Xian and Homer, 2010; Xian et al., 2009; Yu et al., 2016); 2) it is a hard fusion method, because it neglects the continuous probability information from the change intensity map. The hard fusion method generates the binary map for change and non-change candidates by thresholding, while the continuous change intensity information, where larger values indicate higher probabilities to be changed, is lost in the process of binarization.

In this paper, in order to solve the aforementioned problems, we present a new post-classification change detection method with iterative slow feature analysis (ISFA) and Bayesian soft fusion for multi-temporal multispectral remote sensing data. After the class probabilities of corresponding pixels in multi-temporal images are obtained by independent classification, the continuous probabilities of pixel pairs to be changed are calculated by ISFA algorithm. And then, instead of determining the class labels of pixel pairs independently, we proposed a novel Bayesian soft fusion method to find their optimal class combinations fusing the class probabilities and the change probabilities. The class combination with the same label will be more preferred with low change probability, and *vice versa*. In this way, the continuous change probability without thresholding is utilized to avoid error accumulation. The coupled class labels in multi-temporal images will be determined accurately by integrating the class probabilities from independent classification and the change probabilities from change detection.

Therefore, the proposed approach consists of three main parts: 1) independent classification is performed to obtain the class probability; 2) the ISFA algorithm is used to obtain the change probability for each pixel; and 3) based on Bayesian theory, the *a posteriori* probabilities for the class combinations of coupled pixels from the multi-temporal images are calculated with the class probability and the change probability. The class combination with the maximum *a posteriori* probability is then determined as the final result. In addition, a class probability

filter is proposed to avoid the false alarms caused by the spectral variation within the same class. We tested the proposed method using two multi-temporal Landsat TM datasets covering two cities in China, where obvious urban expansion had occurred over the time period.

The rest of this paper is organized as follows. Section 2 gives a brief description of the study areas and the data. Section 3 details the proposed method. The experimental results and discussion are presented in Sections 4 and 5. Finally, we draw our conclusion in Section 6.

2. Study areas and data description

The experiments involved two datasets of multi-temporal TM images acquired by Landsat 5, as shown in Fig. 1. The images contain six spectral bands, with a spatial resolution of 30 m. The first two multi-temporal images were collected on 2000/05/03 and 2002/07/12, covering the developed and newly developing regions of the city of Nanjing, Jiangsu Province, China. The second dataset was collected on 2007/07/26 and 2010/08/19, located in the region of the city of Maanshan, Anhui Province, China. The image sizes are both 800×800. During the time periods, both Nanjing and Maanshan experienced a dramatic change from agricultural to urban land use in their suburban areas. Therefore, these two study areas are suitable to evaluate the performance of the proposed method.
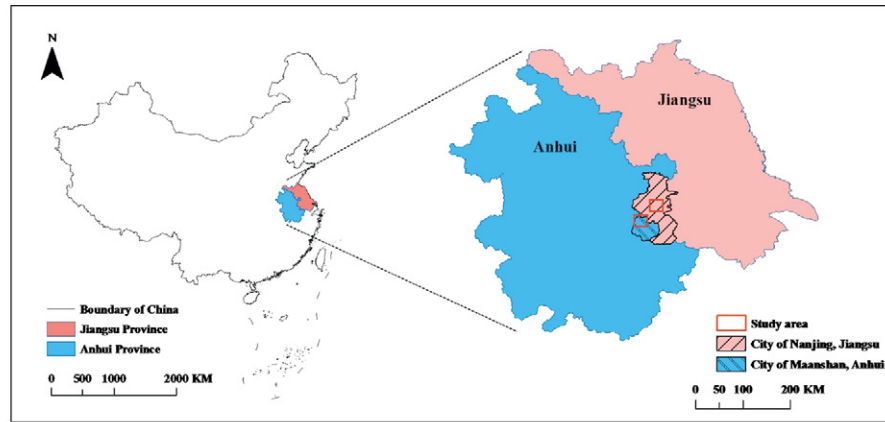
The multi-temporal datasets were provided by the International Scientific & Technical Data Mirror Site of the Computer Network Information Center directed by the Chinese Academy of Sciences (http://www.gscloud.cn). The datasets are the L1 T standard data product, which is processed by systematic radiometric correction and geometric correction with ground control points (GCPs) and a digital elevation model (DEM). We categorized the landscapes in the study areas into four land-cover classes: city, soil, vegetation, and water. For the classification training, we selected 37 samples for city, 17 samples for soil, 34 samples for vegetation, and 30 samples for water in the multi-temporal images of Nanjing; and 66 samples for city, 36 samples for soil, 61 samples for vegetation, and 65 samples for water in the multi-temporal images of Maanshan. In each dataset, the training samples, which differ from the test samples, are selected from the unchanged pixels with the same class label in multi-temporal images, thus used in the independent classification of each image. Actually, the training samples for the proposed method can be selected independently, since the continuous change probabilities are obtained by unsupervised method. The reason to select unchanged samples is to facilitate the comparison between the proposed method and the state-of-the-art supervised method in the following experiments. For the quantitative assessment, we chose 14,756 samples from the Nanjing TM dataset and 9641 samples from the Maanshan TM dataset for evaluation by careful visual interpretation. The reference sample distributions for these two datasets are listed in Appendix A.
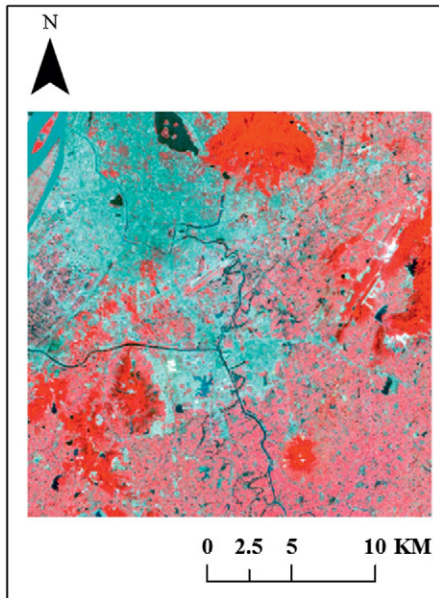
3. Methodology

The flowchart of the proposed method is shown in Fig. 2. The method consists of three steps: 1) independent classification is performed for each multispectral image to obtain the class probability; 2) the ISFA change detection algorithm is employed to obtain the change probability of each pixel; 3) the *a posteriori* probabilities of the coupled class combinations are calculated with the class probability and the change probability according to Bayesian soft fusion; 4) with the class probability filter, the definitely classified pixel pairs will be labeled only by their class probabilities, and the class combinations of the others will be determined by the Bayesian soft fusion. The details are as follows.
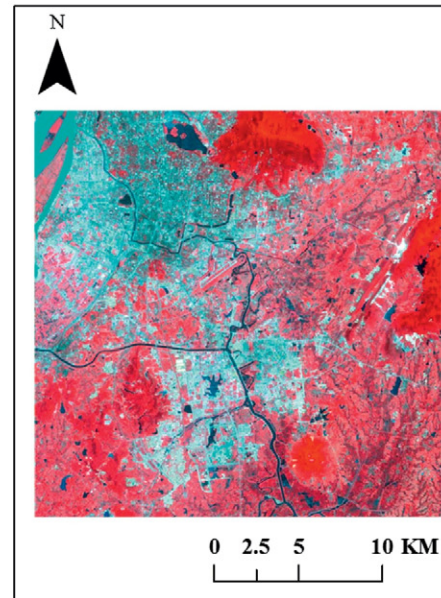
3.1. Classification

The multi-temporal multispectral images are first classified to obtain the class probability of each pixel. In this study, support vector machine
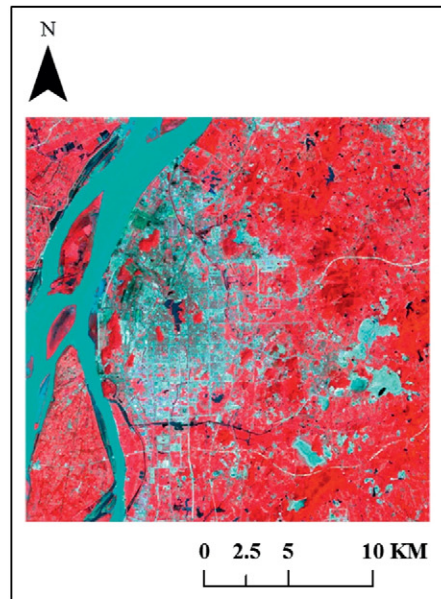
Fig. 1. Study areas and datasets used in this paper: (a) the study areas and their locations; (b)–(c) the pseudo-color images acquired in 2000 and 2002 in the study area of Nanjing; and (d)–(e) the pseudo-color images acquired in 2007 and 2010 in the study area of Maanshan.
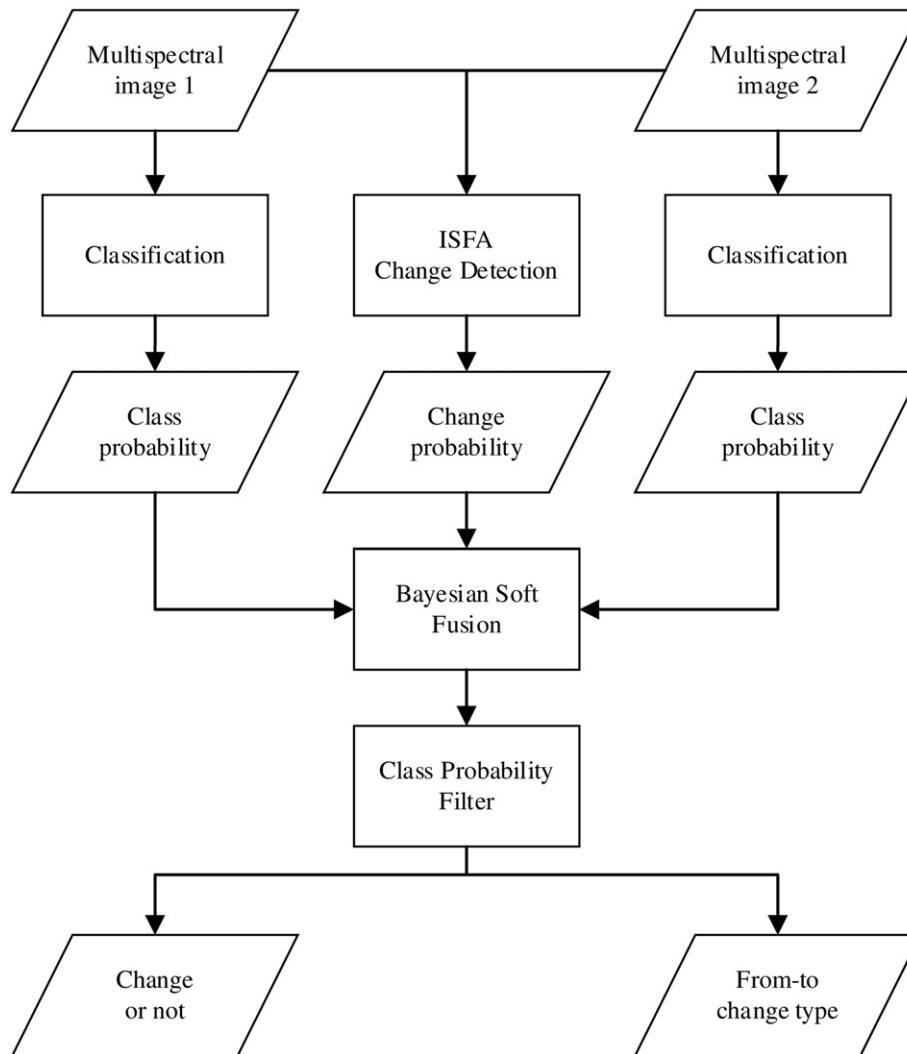
Fig. 2. Flowchart of the post-classification change detection method with ISFA and Bayesian soft fusion.

(SVM) is used as the classification method, since it is robust to small numbers of training samples and can obtain a very good performance. LIBSVM is used for the implementation of the classification (Chang and Lin, 2011). After the classification, the probabilities of all the pixels for each landscape class in the multi-temporal images are obtained for use in the following process.

3.2. Iterative slow feature analysis

In order to fusing the temporal information in the proposed Bayesian soft fusion, the change probability should be obtained to measure the probability of a pixel pair to be changed during the interval. In this paper, an effective unsupervised change detection, which is iterative slow feature analysis (ISFA), is utilized for the accurate change probability map of multi-temporal images (Wu et al., 2014).

In practical applications, the observed signals from sensors are sensitive to small environmental changes, and thus vary greatly over time. However, for the observed targets, they usually have common structures during the observation period, and show statistical regularities in the original signals (Lin et al., 2014). Therefore, their high-level representation extracted from the temporal remote sensing data mostly changes very slowly (Berkes, 2005; Berkes and Wiskott, 2005; Wilbert et al., 2011). Slow feature analysis (SFA) was thus proposed to discover

the underlying essential pattern from the original input data (Wiskott and Sejnowski, 2002; Zhang and Tao, 2012).

In 2014, Wu et al. (2014) improved the SFA algorithm to solve the problem of change detection in multi-temporal remote sensing images. The environmental changes caused by the different acquisition times result in spectral variance in multi-temporal remote sensing images, which leads to pseudo-changes and reduces the accuracy of the change detection. Therefore, SFA was reformulated for the discrete case to extract the temporally invariant features, where the spectral variance is suppressed and the real changes are highlighted (Wu et al., 2014; Wu et al., 2015; Zhang et al., 2014).

Mathematically, SFA change detection methods can be defined as an optimization problem. For multi-temporal images, x_i and y_i indicate the spectral feature vectors of the corresponding pixels in the same location, where $x_i = [x_{i,1}, x_{i,2}, \cdots, x_{i,b}]^T$, $x_{i,b}$ is the band value of sample i in band b. We want to find a set of transformation vectors w_j, such that the optimization objective $\min_{w_j} \frac{1}{n} \sum_{i=1}^{n} (w_j^T \hat{x}_i - w_j^T \hat{y}_i)^2$ will be fulfilled (Wu et al., 2014), where $\hat{x}$ and $\hat{y}$ are the normalized spectral vectors with zero mean and unit variance, and n is the number of samples. The optimization problem needs to be solved under three constraints: zero mean, unit variance, and uncorrelated variables (Wu et al., 2014).

SFA aims to minimize the variance of the unchanged landscapes so as to highlight the real changes. If all the pixels in the remote sensing

image are used for SFA learning, the changed pixels will obviously affect the determination of the feature space. One way to solve this problem is to utilize training samples for the learning; however, in this case, the performance will be influenced by the manual selection. Therefore, the ISFA algorithm was proposed to replace the training samples with iteratively assigned weights (Wu et al., 2014; Zhang et al., 2014).

The basic idea of ISFA is to assign large weights to unchanged pixels in the image so that they play a bigger role in feature extraction over the iteration. Before the iteration, unit weights are assigned to all the pixels. In the next iteration, the weights will affect every calculation process. The input data of the multi-temporal images are normalized with the weights (Wu et al., 2014). After the normalized features are obtained, the optimization problem of SFA can be solved by a generalized eigenvalue problem $AW = BW$, where $\mathbf{W}$ is the transformation matrix, $\mathbf{A}$ is the weighted covariance matrix of the difference image, and $\mathbf{B}$ is the mean of the two weighted covariance matrices of each image (Wu et al., 2014).

The output features are sorted according to the ascending order of the eigenvalues, where the features with the lowest variance are ranked first. Finally, the changed pixels will have a high separability with the unchanged background in the difference image of the transformed features as $SFA_{i,j} = \hat{\boldsymbol{w}}_j^{\mathrm{T}} \hat{\boldsymbol{x}}_i - \hat{\boldsymbol{w}}_j^{\mathrm{T}} \hat{\boldsymbol{y}}_i$.

After the transformed features are obtained, the weights are determined according to the chi-squared distribution (Wu et al., 2014). According to SFA theory, the differences of the transformed features for unchanged pixels are normally distributed and independent, and thus their squared sum will follow an χ^2 distribution with N degrees of freedom (Lancaster and Seneta, 2005). In each iteration, the weight of each pixel can be assigned as the probability of being unchanged $p = \mathrm{P}\{\chi^2(N) > T\}$, where T is the chi-squared distance and calculated with the transformed features (Wu et al., 2014).

Pixels with a small chi-squared distance will have a high probability of being unchanged, and thus they will be assigned large weights in the learning of the invariant feature space. Due to the distribution, the weights will be limited to the range of [0,1]. In the next iteration, the updated weights will be assigned to all the pixels and will affect the feature learning.

The iteration continues until the stopping criterion is met. The stopping criterion of ISFA is that the maximum difference between the eigenvalues of the current iteration and the last iteration is smaller than a given threshold.

When the iteration converges, the change probability of every pixel in the image can be obtained accurately by $p_c = 1 - p$, where p is the probability of being unchanged.

3.3. Bayesian soft fusion

In multi-temporal remote sensing images, let us consider a corresponding pixel pair $\boldsymbol{x}$ and $\boldsymbol{y}$, where w_i and v_j are their class labels ($i = 1, ..., m_1$, and $j = 1, ..., m_2$). In traditional post-classification change detection, the objective is to determine the optimal w_i and v_j for the corresponding pixels according to their spectral features $\boldsymbol{x}$ and $\boldsymbol{y}$, and we then compare the class labels to obtain the "from-to" change information. However, due to the fact that the post-classification methods neglect the temporal correlation, they suffer from an accumulation of misclassification errors.

If we assume that the coupled class labels of the multi-temporal pixels are temporally correlated, then the optimal multi-temporal classification can be given by Bayesian theory as: find the coupled class combination (w_i, v_j) that provides the maximum a posteriori probability for the multi-temporal features $\boldsymbol{x}$ and $\boldsymbol{y}$ (Bruzzone and Serpico, 1997):

$$\max_{w_i, v_j} \{ P(w_i, v_j | \boldsymbol{x}, \boldsymbol{y}) \} \tag{1}$$

with Bayesian theory, the a posteriori probability can be reformulated as:

$$\max_{w_i, v_j} \left\{ \frac{P(\boldsymbol{x}, \boldsymbol{y} | w_i, v_j) P(w_i, v_j)}{P(\boldsymbol{x}, \boldsymbol{y})} \right\}$$
$$\Rightarrow \max_{w_i, v_j} \left\{ \frac{P(\boldsymbol{x}, \boldsymbol{y} | w_i, v_j) P(w_i | v_j) P(v_j)}{P(\boldsymbol{x}, \boldsymbol{y})} \right\} \tag{2}$$

$P(\boldsymbol{x}, \boldsymbol{y})$ is independent of w_i, v_j, and thus it makes no contribution to the determination of the class combination and can be ignored. Let us assume that the probability distribution of spectral feature $\boldsymbol{x}$ only depends on its class in the multispectral image at one time. Thus, we can write:

$$P(\boldsymbol{x}, \boldsymbol{y} | w_i, v_j) = P(\boldsymbol{x} | w_i) P(\boldsymbol{y} | v_j) \tag{3}$$

Therefore, by combining Eqs. (2) and (3), we obtain the following formulation of the a posteriori probability:

$$\max_{w_i, v_j} \{ P(\boldsymbol{x} | w_i) P(\boldsymbol{y} | v_j) P(w_i | v_j) P(v_j) \}$$
$$\Rightarrow \max_{w_i, v_j} \left\{ \frac{P(w_i | \boldsymbol{x}) P(\boldsymbol{x})}{P(w_i)} \cdot \frac{P(v_j | \boldsymbol{y}) P(\boldsymbol{y})}{P(v_j)} \cdot P(w_i | v_j) P(v_j) \right\} \tag{4}$$

$P(\boldsymbol{x})$ and $P(\boldsymbol{y})$ are also independent of their classes, and we can assume that the a priori probabilities $P(w_i)$ of each landscape class are equal in one image. Therefore, the final decision rule is obtained as follows:

$$\max_{w_i, v_j} \{ P(w_i | \boldsymbol{x}) P(v_j | \boldsymbol{y}) P(w_i | v_j) \} \tag{5}$$

where $P(w_i | \boldsymbol{x})$ and $P(v_j | \boldsymbol{y})$ are the two conditional probabilities from the independent classification, and $P(w_i | v_j)$ is the transition probability, which is estimated by the change probability as:

$$P(w_i | v_j) = \begin{cases} 1 - p_c & \text{if } w_i = v_j \\ p_c & \text{if } w_i \neq v_j \end{cases} \tag{6}$$

where p_c is the probability of being changed, which can be obtained by ISFA proposed in Section 3.3.

In summary, firstly, the class probabilities $P(w_i | \boldsymbol{x})$ and $P(v_j | \boldsymbol{y})$ are obtained by independent classification, and the change probability $P(w_i | v_j)$ is calculated by ISFA. The a posteriori probabilities (Eq. (5)) for all possible combinations (w_i, v_j) are then calculated. Finally, the coupled class combination with the maximum a posteriori probability is determined as the result of the change detection in the multi-temporal remote sensing images. The proposed process integrating the class probability with the change probability is referred to as "Bayesian soft fusion".

3.4. Class probability filter

It is worth noting that there is a problem in the proposed Bayesian soft fusion strategy. When the a posteriori probability is calculated, the landscapes with high change probabilities are more likely to be determined as changes. However, in real applications, some regions with the same class will show very different spectral features at different times (e.g., water areas). The Bayesian soft fusion may lead to some false alarms for unchanged pixels compared to the original method. Therefore, we also propose a class probability filter to avoid such errors.

The class probability filter is that, if the class probabilities of the corresponding pixels in multi-temporal images are both larger than a threshold, such as 0.7, their class labels are only determined according to their class probabilities obtained from the independent classifications. And for the other pixel pairs, the class combinations will be

determined by the proposed Bayesian soft fusion. The class probability filter can be expressed as follows:

$$(w_i, v_j) = \begin{cases} \max_{w_i,v_j}\{P(w_i|\boldsymbol{x}), P(v_j|\boldsymbol{y})\} \\ \quad\quad, \text{ if } P(w_i|\boldsymbol{x}) > f \& P(v_j|\boldsymbol{y}) > f \\ \max_{w_i,v_j}\{P(w_i|\boldsymbol{x})P(v_j|\boldsymbol{y})P(w_i|v_j)\} \\ \quad\quad, \text{ else} \end{cases} \quad (7)$$

where f indicates the class probability filter.

The basic idea is, if the class probabilities are both very high, the independent classification is definite, and it is not necessary to integrate the change probability to determine the classes for the multi-temporal pixels. The use of class probability filter differs in several ways from hard fusion with a threshold for change intensity. In hard fusion methods, the threshold is applied to the change intensities by interpreting temporal correlation to separate change and non-change candidates; while in the proposed soft fusion method, the probability threshold is applied to the class probabilities from independent classifications to exclude the definitely classified pixels. The filter setting is evaluated in the experimental part.

3.5. Comparative methods

Hard fusion with CVA is a widely used improved post-classification change detection method for updating LULC maps from multi-temporal remote sensing images (Xian and Homer, 2010; Xian et al., 2009; Yu et al., 2016), and is thus used here as a comparison method. Firstly, CVA is applied to obtain the change intensity, and a threshold is determined to identify the change and non-change candidates. In previous studies, $\mu + a\sigma$ was used as the criterion, and the parameter a was not determined automatically (Yu et al., 2016). Therefore, in order to avoid the influence of manual interpretation, we use Otsu's thresholding algorithm to segment the change map (Otsu, 1975; Wu et al., 2015). The changed areas are then obtained from the original post-classification method, and the classes of the unchanged areas are determined by the maximum class probability from the two independent classifications.

For the soft fusion approach, in addition to ISFA, supervised SFA (SSFA) without iteration (Wu et al., 2014) and iteratively reweighted multivariate alteration detection (IRMAD) (Nielsen, 2007) can also provide change probability information. For SSFA, the training samples were the same as those used in the independent classification. For the hard fusion approach, besides CVA, ISFA, SSFA, and IRMAD could also be utilized to obtain the change map. In the experiments, we compared and analyzed the proposed soft fusion method and the hard fusion approach in change detection performance.

3.6. Accuracy assessment

We conducted the accuracy assessment in several ways, to ensure a comprehensive evaluation. The change detection problem was regarded as a binary classification problem for change and non-change, and kappa coefficients were calculated to avoid the unbalanced test samples. The "from-to" transition type was then regarded as one class for the evaluation. The classification accuracies obtained by the proposed method for each image were also evaluated. Finally, the detection rate (DR) and false alarm rate (FAR) were determined for a detailed analysis. Since different methods may focus on different kinds of improvement in DR or FAR, the F-score, which has been widely used in pattern recognition, was utilized for comprehensive evaluation through the fusion of DR and FAR (Goutte and Gaussier, 2005; Sokolova et al., 2006; Wang et al., 2015). The

F-score with the balance of DR and FAR was calculated to measure the ability of the change detection as follows:

$$F\text{-}score = \frac{2 \times DR \times RR}{DR + RR} \quad (8)$$

where the recall rate (RR) is equal to $1 - FAR$.

4. Experiments

4.1. Nanjing TM dataset

The change detection results for change/non-change and the "from-to" change type are shown in Fig. 3. Fig. 3 (a) and (b) show the results of the traditional post-classification method and the proposed method with a class filter of 0.7 (which is the parameter setting that obtains the best performance). It can be observed that the proposed method obtains fewer changes than the traditional post-classification method (e.g., in the subset areas of (c)–(k)). From Fig. 3 (d), it can be seen that the detected changes in south of the lake in the city areas are removed by the proposed method. Visually, most of the removed changes are false alarms, since the city area didn't show so many changes in Fig. 3 (c). This is also shown by Fig. 3 (e), where most of the removed changes are from city to vegetation, that are impossible in practice. Fig. 3 (f)–(h) show that the road is falsely detected by the traditional method, which can be corrected by fusing the change probability information. Fig. 3 (i)–(k) show that the obvious city expansion areas are both detected by the two methods. The quantitative assessments for the performances of change detection between the traditional post-classification method and the proposed method can be seen in Appendix B.

Fig. 4 shows the results for the "from-to" change type, where it can be seen that the city expansion led to obvious land-cover transitions in the Jiangning region, the south part of Nanjing. The land-cover changes mainly consist of vegetation areas changing into city areas or soil areas ready for development. In Fig. 4 (a), the traditional post-classification method obtains numerous false alarms, showing changes from city to vegetation in the center of the city of Nanjing, which is almost impossible in a real case. Comparatively speaking, the result obtained by the proposed method in (b) shows a much better performance in suppressing the false alarms.

In order to compare the proposed method with other state-of-the-art methods, the accuracies obtained by the different methods are summarized in Table 1. In Table 1, "soft fusion" means that the change probability and the independent classification are fused based on the proposed approach, as described in Section 3.3. In addition to ISFA, SSFA and IRMAD can also provide change probability information for the Bayesian soft fusion. "Hard fusion" in Table 1 indicates the fusion approach described in Section 3.5, which is widely used to improve the performance of post-classification (Xian et al., 2009). Apart from the traditional CVA method, ISFA, SSFA, and IRMAD can also obtain a change map with Otsu's thresholding algorithm. Besides automatic thresholding with Otsu algorithm, $\mu + 1.5\sigma$, which is manually set as the threshold according to (Yu et al., 2016), was used for hard fusion with CVA (CVA-std) and evaluated in Table 1. The threshold for the class probability filter in Table 1 results in the best change detection accuracy for each method. The highest accuracy is highlighted in bold, and the second highest accuracy is denoted by underlining. In Table 1, "Kappa1" and "Kappa2" are the kappa coefficients for the two multi-temporal image classifications, and "cdKappa" and "trKappa" illustrate the accuracy for change/non-change and the "from-to" transition. "DR" and "FAR" represent the detection rate and false alarm rate, where the F-score is a comprehensive indicator of their performance.

It can be observed in Table 1 that all the improved methods have the ability to increase the accuracies for change detection and the "from-to" change type identification over the result obtained by the traditional post-classification method. The quantitative evaluation demonstrates
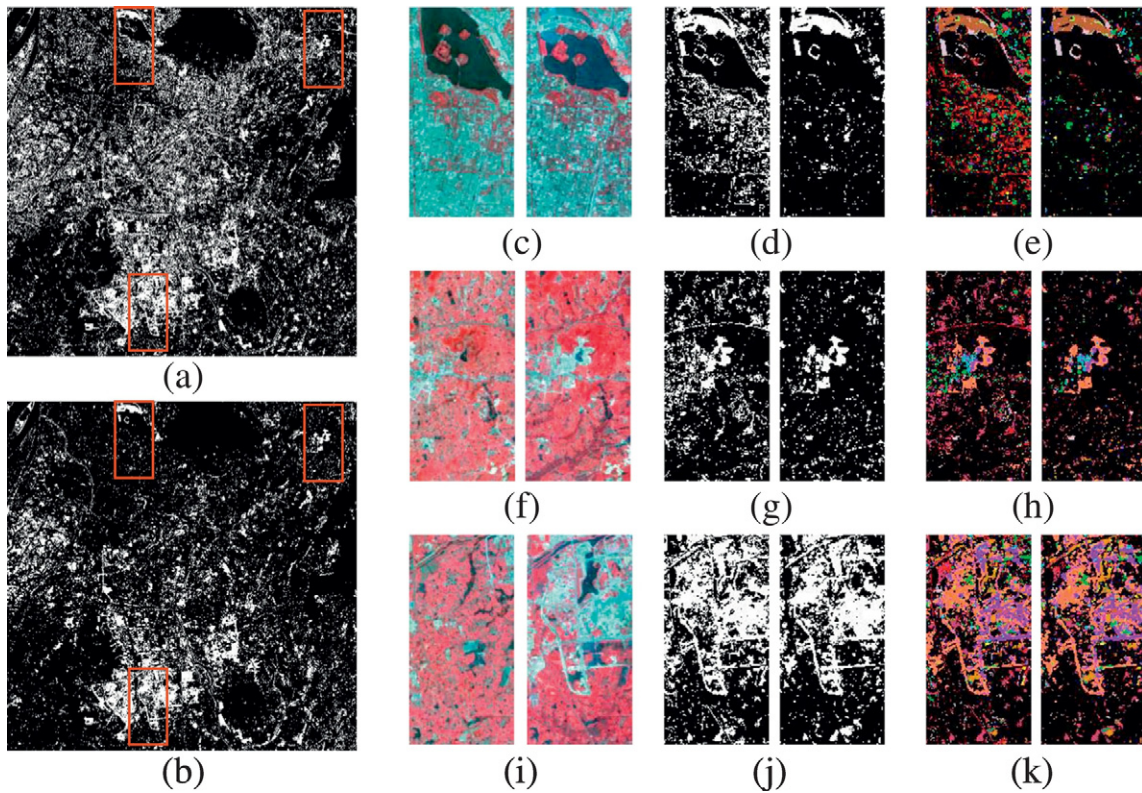
Fig. 3. Change detection results for change/non-change: (a) and (b) are the change detection results of the traditional post-classification method and the proposed method; (c), (f), and (i) are subset areas of the multi-temporal remote sensing images from 2000 and 2002; (d), (g), and (j) are the corresponding subset areas of the change detection results; (e), (h), and (k) are the corresponding subset areas of the "from-to" change type, where the different colors indicate different change types. The legends of colors are shown in Fig. 4.

that the proposed method can perform significantly better than the traditional post-classification method. Although all the improved methods show a slight decrease in classification accuracy for remote sensing image 1, they show a great improvement in the classification map for image 2. Compared with the manual threshold, hard fusion with automatic Otsu's thresholding algorithm shows higher accuracies in all cases in Table 1. It demonstrates that Otsu's thresholding method is very effective, and has the ability to replace the manual threshold in hard fusion methods. Among all the methods, the proposed method with ISFA and Bayesian soft fusion obtains the highest change detection accuracy of 0.850, and the third highest accuracy of 0.828 for land-cover transition. The proposed method also shows the maximum DR of 93.7%

and a comparatively low FAR of 17.8%, which are both better than those obtained by the traditional post-classification method. According to the F-score, the proposed method is superior to all the other methods. The quantitative evaluation of the performances with different thresholds of class probability filter can be seen in Appendix C.

4.2. Maanshan TM dataset

Fig. 5 shows the change detection results for the Maanshan TM dataset, where it can be seen that, compared with the traditional post-classification method, the proposed method obtains fewer changed regions. From the subsets shown in Fig. 5 (c)–(k), it can be seen that the
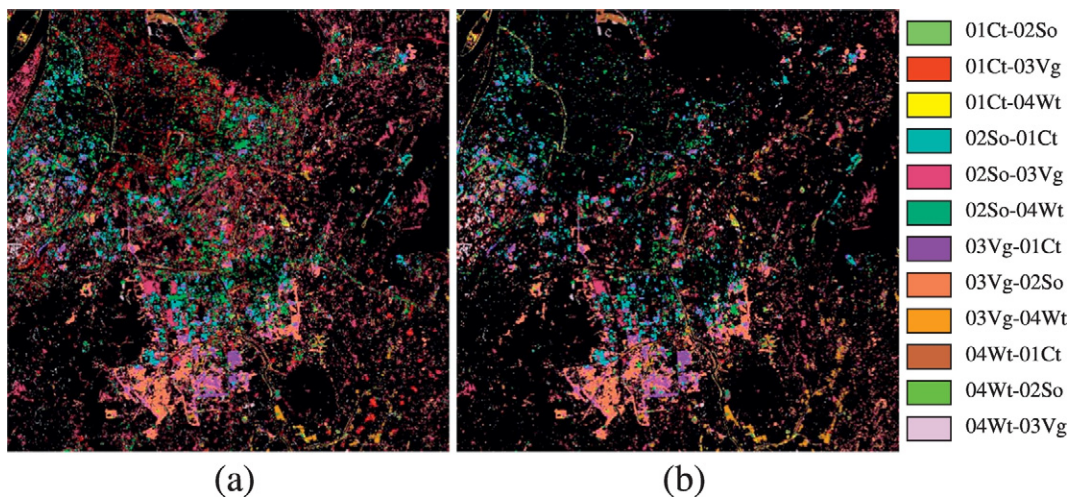

Fig. 4. Change detection results for the "from-to" change type by: (a) the traditional post-classification method; and (b) the proposed method, where the different colors indicate the different change types.

Table 1
Accuracy assessment for change detection and "from-to" transition identification.

	Method	Thresh.	Kappa1	Kappa2	cdKappa	trKappa	DR	FAR	F-score
	Post		**0.897**	0.816	0.631	0.786	0.883	0.417	0.703
Soft fusion	ISFA	0.7	0.872	0.861	**0.850**	0.828	**0.937**	0.178	**0.876**
	SSFA	0.5	0.885	0.833	0.705	0.798	<u>0.928</u>	0.356	0.760
	IRMAD	0.5	<u>0.892</u>	0.809	0.624	0.777	<u>0.928</u>	0.439	0.699
Hard fusion	ISFA	0.5	<u>0.877</u>	0.860	0.764	0.821	0.815	0.209	0.802
	SSFA	0.6	0.870	0.865	0.764	0.822	0.751	0.146	0.799
	IRMAD	0.6	0.873	<u>0.866</u>	0.782	0.825	0.777	<u>0.143</u>	0.815
	CVA	0.7	0.879	**0.866**	<u>0.822</u>	**0.833**	0.833	**0.132**	<u>0.850</u>
	CVA-std	0.6	0.876	0.865	0.799	<u>0.829</u>	0.805	0.143	0.831

city expansion is effectively detected by both methods. It is worth noting that the spectral feature of the triangular water region in Fig. 5 (i) changed very obviously, and it is correctly determined as non-change by the proposed method. This demonstrates that the proposed approach with the class probability filter has the ability to avoid inaccurate detection caused by spectral variation. The quantitative assessments for the performances of change detection between the traditional post-classification method and the proposed method can be seen in Appendix B.

Fig. 6 shows the "from-to" change detection results for this dataset. Visually, the proposed method performs well in detecting the city expansion. There are also fewer detected changes in the image obtained by the proposed method than in the image obtained by the traditional post-classification method.

The maximum accuracy of the change detection for each method (Table 2) shows that the proposed method with ISFA and Bayesian soft fusion obtains the highest kappa coefficients for the change detection (0.8800) and the "from-to" transition identification (0.884). In the comprehensive evaluation with the F-score, the proposed method also shows an obvious improvement from 0.842 to 0.902, which is the

highest among all the methods. What needs special attention is that, although the proposed method only obtains a slight increase in the accuracies of the independent classification, it shows a great improvement for change detection. It can be found that CVA with automatic threshold gets better performances in almost all accuracies than that with manual threshold for hard fusion, thus Otsu's thresholding is recommended in hard fusion. For the quantitative assessment of the performances with different thresholds of class probability filter, please refer to Appendix C.

5. Discussion

5.1. Improvement over traditional methods

The comparison between the proposed method and the traditional post-classification method indicates that the proposed method with ISFA and Bayesian soft fusion is superior to the direct comparison of independent classification. As we discussed in Section 1, the traditional post-classification method suffers from an accumulation of misclassification errors, which show as the large number of false alarms in the change detection result (Singh, 1989). This is because the unchanged
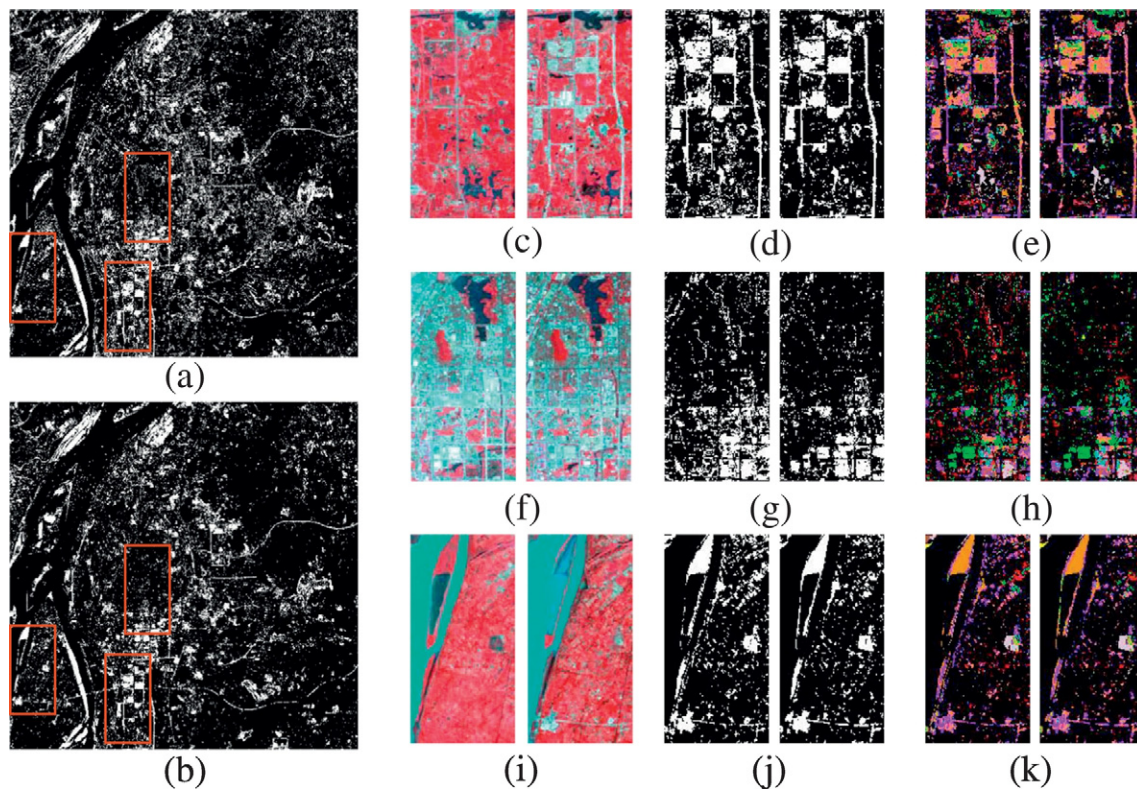

Fig. 5. Change detection results for change/non-change: (a) and (b) are the change detection results of the traditional post-classification method and the proposed method; (c), (f), and (i) are subset areas of the multi-temporal remote sensing images from 2007 and 2010; (d), (g), and (j) are the corresponding subset areas of the change detection results; (e), (h), and (k) are the corresponding subset areas of the "from-to" change type, where the different colors indicate different change types. The legends of colors are shown in Fig. 6.
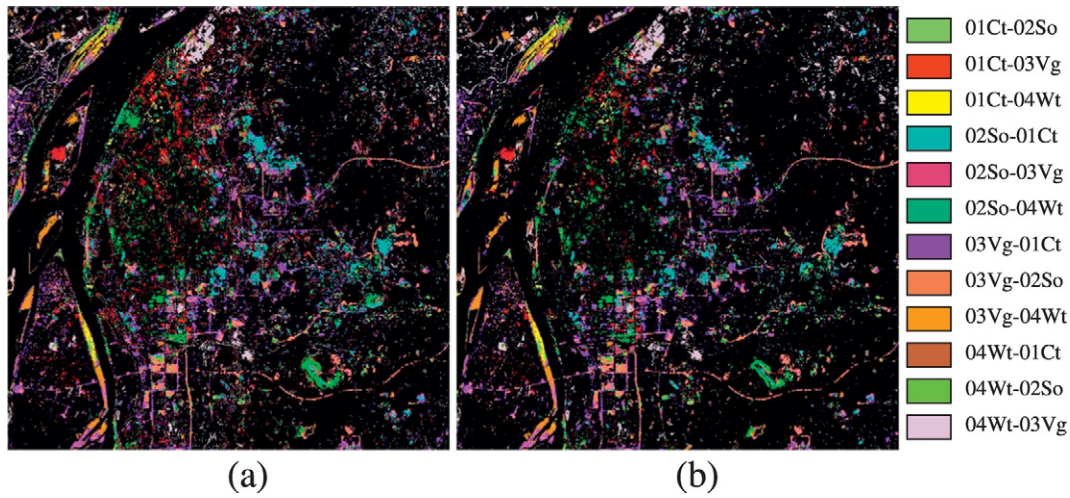
Fig. 6. Change detection results for the "from-to" change type obtained by: (a) the traditional post-classification method; and (b) the proposed method, where the different colors indicate the different change types.

landscapes with the same spectral features may be detected as changes by the traditional post-classification method, due to the variation of the classification boundaries in the multi-temporal images. Therefore, the change information should be taken into consideration to avoid this problem. This can be seen in the experimental results. Taking the Nanjing TM dataset as an example, there are numerous changes from city to vegetation in the center of the city in the result of the traditional post-classification method (Fig. 3 (a) and Fig. 4 (a)), which is almost impossible in reality. As we analyzed above, most of these false alarms are the result of misclassification. Comparatively speaking, the false alarms are effectively suppressed in the results of the proposed method shown in Fig. 3 (b) and Fig. 4 (b). In the quantitative evaluation (Table 1), the proposed method shows a decrease in FAR from 0.4169 to 0.1779 and an increase in DR from 0.8832 to 0.9365 at the same time. The improvement can also be seen in the Maanshan TM dataset.

In order to solve the problem in using the post-classification method, hard fusion has been widely used for updating LULC maps from multi-temporal images (Xian and Homer, 2010; Xian et al., 2009; Yu et al., 2016). According to our experiments, this approach using CVA is also very effective, obtaining the second highest accuracy for change/non-change and the maximum accuracy for the "from-to" change type in Table 1. The basic idea of hard fusion is to avoid false alarms in unchanged candidates, and thus it shows the lowest FAR in Table 1 (Yu et al., 2016). However, since hard fusion approach limits the unchanged candidates to belonging to the same class, the DR is also decreased compared with the traditional method, which can be observed from the accuracies by all the methods with hard fusion in Table 1. Comparatively, this paper proposes Bayesian soft fusion taking advantage of continuous change probability. As shown in Table 1, methods with soft fusion tend to obtain higher DR, while their FARs are mostly higher than those of hard fusion. With the change probability provided by ISFA, the proposed

method increases the DR as well as decreasing the FAR, and thus obtains the highest F-score in the comprehensive evaluation (Table 1).

The difference between hard fusion and soft fusion based on Bayesian theory can also be observed in Fig. C1 in Appendix C. Fig. C1 (c) and (d) show the DR and FAR of these methods with different thresholds. Compared with the DR and FAR obtained by the traditional post-classification method (indicated by the horizontal dashed line), all the methods with soft fusion show an obvious improvement in DR, while most of them show a higher FAR at the same time. For the cases of hard fusion, the improved methods obtain significantly lower FARs, while obtaining lower DRs at the same time. This can also be seen in Fig. C2 (c) and (d) in Appendix C.

Therefore, we can conclude that, as improvements of the traditional post-classification approach, hard fusion tends to suppress the false alarms in change detection, while soft fusion improves the detection rate of changed areas. If the change probability is accurate enough, such as the result of ISFA, the soft fusion method is capable of reducing the FAR at the same time, as shown in Fig. C1 and Fig. C2. This is the reason why the proposed method with ISFA and Bayesian soft fusion obtains the highest F-score for change detection evaluation.

5.2. Effect of the class probability filter

In this paper, we propose a filter for the class probability obtained by the independent classification. If the class probabilities in the multi-temporal images are both higher than a pre-defined threshold, their class labels will not be corrected by the soft fusion. The motivation for this procedure is that some landscapes with the same class show extremely different spectral signatures at different times. For example, visually, the spectral feature of the water region in the center of Fig. 7 (a) and (b) changes a lot, whereas it belongs to the water class in both

Table 2
Accuracy assessment for the change detection and the "from-to" transition identification.

|	Method	Thresh.	Kappa1	Kappa2	cdKappa	trKappa	DR	FAR	F-score
	Post		0.929	0.897	0.804	0.873	0.931	0.231	0.842
Soft fusion	ISFA	0.7	**0.930**	**0.898**	**0.880**	**0.884**	0.955	0.145	**0.902**
	SSFA	0.6	0.925	0.895	0.833	0.872	**0.963**	0.214	0.866
	IRMAD	0.5	0.926	0.896	0.783	0.867	0.947	0.267	0.826
Hard fusion	ISFA	0.5	0.926	0.888	0.785	0.867	0.843	0.194	0.824
	SSFA	0.6	0.925	0.887	0.798	0.868	0.803	0.137	0.832
	IRMAD	0.5	0.926	0.888	0.785	0.867	0.843	0.194	0.824
	CVA	0.7	0.924	0.890	0.849	0.874	0.859	**0.108**	0.875
	CVA-std	0.5	0.926	0.889	0.786	0.868	0.850	0.199	0.825
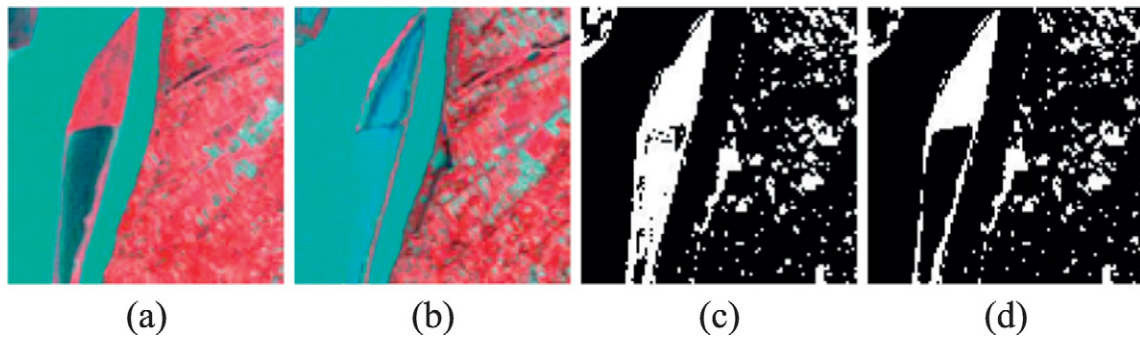
Fig. 7. The effect of the class probability filter: (a) and (b) are subsets of the Maanshan TM dataset; (c) shows the result of the proposed method without filter; and (d) shows the result of the proposed method with the threshold of 0.7.

images. With the proposed method, it will obtain a very high change probability, and will be more likely to be categorized into different classes with the *a posteriori* probability, as shown in Fig. 7 (c). Therefore, we utilize a filter for the class probability so that the class labels of the corresponding pixels are not influenced by the change probability when the independent classifications are both definite. With the threshold of 0.7, the false alarms in the water region are avoided, as shown in Fig. 7 (d).

This can also be seen in Fig. C2 in Appendix C, where the threshold of 1 indicates that the method is employed without a threshold. Fig. C2 (a), (b), and (e) show that most methods obtain an improvement in accuracy with the use of the filter. In Fig. C2 (d), with the filter, the three methods with soft fusion show an obvious decrease in FAR, which indicates our point made above. However, when the threshold is too small, such as 0.5, most of the change detection accuracies show a slight decrease, since too many pixels are excluded from the soft fusion. If the threshold is set as 0, all the methods are equal to the traditional post-classification method. This conclusion can also be verified by the experiment with the Nanjing TM dataset. Therefore, we can conclude that the filter for the class probability used in this paper is effective at reducing the false alarms caused by spectral variation within the same class.

5.3. LULC changes in urban expansion monitoring

LULC changes are essential indicators for the monitoring of urban development and expansion (Hu and Zhang, 2013; Zhang and Zhang, 2007). The "from-to" change information can illustrate the expansion tendency of a city (Yuan et al., 2005). In order to analyze the expansion

of the two study sites in this paper, the statistics of the "from-to" changes obtained by the proposed method are displayed in a 3×3 grid in Fig. 8. The statistical bars are normalized according to the maximum numbers of changes in each dataset.

From Fig. 8 (a), it can be observed that the expansion of the city of Nanjing mostly took place in the south of the study site, in the 4th, 7th, and 8th sub-regions. Compared with the developed urban areas of Nanjing (the 1st and 2nd sub-regions), the newly developing suburban areas experienced a dramatic change from agriculture to urban. This phenomenon is shown in the fact that numerous changes from vegetation to city or soil appeared in the suburban areas. It is worth noting that the changes from vegetation to soil make up a large proportion of the changes, and it is likely that these "idle" regions will be further developed in the next few years.

The city of Maanshan experienced development and expansion in its suburban areas (the 5th and 8th sub-regions), as shown in Fig. 8 (b). In the 5th sub-region, the major change type is from vegetation to city. In the suburban areas in the south of the city (the 8th sub-region), numerous idle areas (from vegetation to soil) have appeared, which suggests that the future urban expansion will take place in the south of the city of Maanshan.

From the above analysis, it can be seen that the LULC change information obtained by the proposed method is crucial information for the monitoring of urban development and expansion. The statistics of the "from-to" LULC changes can be used to illustrate the development tendency of a city. However, due to the many different applications, the unit regions for such statistics are often different, *e.g.*, total urban area (Yuan et al., 2005), spatial grids (Xiao et al., 2006), buffer zones
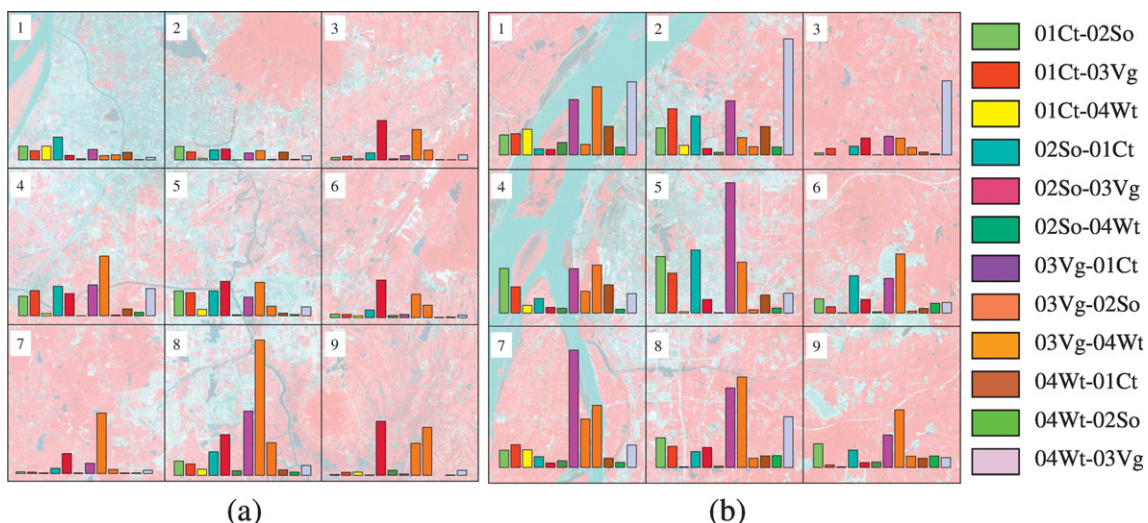

Fig. 8. Statistics of the "from-to" changes for: (a) the Nanjing TM dataset; and (b) the Maanshan TM dataset.

(Seto and Fragkias, 2005), and city planning patterns such as ring roads (Li et al., 2011).

5.4. Limitations

The proposed Bayesian soft fusion method takes advantage of the class probabilities from independent classification and the change probability from change detection to obtain the optimal class combinations of coupled pixels in multi-temporal images. Thus, the performance of the proposed method mainly depends on the accuracies of class probability and change probability.

For independent classification, if the classifier doesn't provide class probability, the *a posteriori* probability of class combinations cannot be calculated. It is worth noting that the proposed Bayesian soft fusion method improves the final performance on the basis of independent classification. If the classification methods cannot provide accurate class probabilities, the accuracies of change detection and transition identification will also be unsatisfactory.

Besides the class probability, the change probability also has significant effect on the final result. It can be observed that, if the change probability is not very accurate, such as those obtained by SSFA and IRMAD in the experiments, the results will be only slightly better, or even worse than that of the traditional post-classification method, as shown in Table 1 and Table 2. Thereforein the proposed method, we utilize ISFA algorithm to provide accurate change probability. We can conclude that, in some special cases when the changes are extremely complex, ISFA cannot get a good result, and thus the proposed method will fail to get an improvement compared with the traditional post-classification. Then, other methods, such as supervised ISFA (S-ISFA) (Zhang et al., 2014), can be considered to replace ISFA for a better performance.

In this paper, the proposed Bayesian soft fusion provides a framework for the improvement of the traditional post-classification method. ISFA is used in the proposed method since it can provide accurate change probability information. Other change detection methods, which have the ability to produce more accurate change probability, can be utilized in the framework with Bayesian soft fusion in the future.

In real cases, some landscapes with the same class will show different spectral features in multi-temporal images. These landscapes will obtain high change probabilities, and trend to belong to different classes by the proposed Bayesian soft fusion, as discussed in Section 5.2. Thus, it will bring some false alarms without the class probability filter, which can be indicated with higher false alarms rate in Fig. C1 and Fig. C2 in Appendix C. Another limitation of the class probability filter is the selection of threshold. Different thresholds will lead to different accuracies, as shown in Fig. C1 and Fig. C2. According to our experiences, we recommend that 0.7 is a suitable threshold for the balance of Bayesian soft fusion and independent classification.

6. Conclusion

Accurate and detailed land-use/land-cover (LULC) change detection is of great importance for a wide variety of applications, such as ecosystem monitoring and urban development studies. Post-classification is a widely used method to provide "from-to" change information, but it suffers from the accumulation of independent misclassification errors. Therefore, in this paper, we have proposed a new post-classification method with iterative slow feature analysis (ISFA) and Bayesian soft fusion, which integrates the class probability from independent classification with the continuous change probability from change detection. The continuous probability of pixel pairs to be changed is provided by ISFA, and the optimal class combination is obtained by fusing the class probability and change probability with Bayesian soft fusion.

The results of two experiments with multi-temporal TM images confirmed the effectiveness and advantage of the proposed method. Compared with the traditional post-classification method, the proposed method has the ability to clearly improve the performance from all aspects, including the accuracy of the change/non-change, the accuracy of the "from-to" change type, the detection rate, the false alarm rate, and the comprehensive assessment of F-score. The existing hard fusion with CVA mainly focuses on avoiding false alarms, while the proposed soft fusion based on Bayesian theory aims at increasing the detection rate, and the proposed ISFA can provide accurate change probability to decrease the false alarm rate at the same time. Therefore, the proposed method outperforms the hard fusion methods in change detection and transition identification in most cases. The qualitative and quantitative evaluations also confirmed that the class probability filter is effective in suppressing the false alarms caused by spectral variation. The manual threshold of class probability filter will have influence on the final performance, and 0.7 is recommended for the balance. In summary, the proposed method achieved the best performance in all the experiments conducted in this study, and it is superior to the other improved classification-based methods. Considering its effectiveness and usability, the proposed method has the potential to be widely applied in real applications, where an accurate LULC change map is extremely useful for monitoring and analyzing city development and expansion.

Appendix A. Reference samples of the experiment datasets

In order to evaluate the performance of the proposed method, we have selected 14,756 pixels for the Nanjing TM dataset, and 9641 pixels for the Maanshan TM dataset. The details for the reference samples are shown in Tables A1 and A2

Table A1
Reference samples of the Nanjing TM dataset (pixels).

TM 2000	TM 2002				Total
	01City	02Soil	03Veg.	04Water	
01City	3934	104	4	5	4047
02Soil	370	119	262	339	1090
03Veg.	204	577	4361	210	5352
04Water	167	24	97	3979	4267
Total	4675	824	4724	4533	14,756

Table A2
Reference samples of the Maanshan TM dataset (pixels).

TM 2007	TM 2010				Total
	01City	02Soil	03Veg.	04Water	
01City	2024	0	0	0	2024
02Soil	83	1375	62	0	1520
03Veg.	241	764	2345	339	3689
04Water	15	0	205	2188	2408
Total	2363	2139	2612	2527	9641

Appendix B. Quantitative assessment for the improvement in change detection

In the experiment of Nanjing TM dataset, the confusion matrices for the change detection by the traditional post-classification method and the proposed method are shown in Table B1 and Table B2. The producer's accuracies of non-change and change are both improved from 88.0% to 96.1%, and from 88.3% to 93.7%. The user's accuracies of non-change and change are also improved from 97.5% to 98.8%, and from 58.3% to 82.2%. The overall accuracy (OA) and kappa both show a great increase from 88.0% to 95.7%, and from 0.631 to 0.850. The quantitative evaluation demonstrates that the proposed method can perform significantly better than the traditional post-classification method.

Table B1
Change detection assessment of the traditional post-classification method (pixels).

Test data	Reference data			User's acc.
	Non-change	Change	Total	
Non-change	10,901	276	11,177	97.5%
Change	1492	2087	3579	58.3%
Total	12,393	2363	14,756	
Producer's acc.	88.0%	88.3%		
Overall acc.	88.0%			
Kappa	0.631			

Table B2
Change detection assessment of the proposed method (pixels).

Test data	Reference data			User's acc.
	Non-change	Change	Total	
Non-change	11,914	150	12,064	98.8%
Change	479	2213	2692	82.2%
Total	12,393	2363	14,756	
Producer's acc.	96.1%	93.7%		
Overall acc.	95.7%			
Kappa	0.850			

In the experiment of Maanshan TM dataset, the confusion matrices for the change detection results of the traditional post-classification method and the proposed method are shown in Table B3 and Table B4. This shows that the proposed method obtains an obvious improvement over the traditional post-classification method. The producer's accuracies increase from 94.0% to 96.5%, and from 93.1% to 95.5%. The user's accuracies increase from 98.4% to 99.0%, and from 76.9% to 85.5%. Compared with the improvement for OA (93.8% to 96.3%), the increase of the kappa coefficient is more significant (0.804 to 0.880), due to the imbalance of the test samples.

Table B3
Change detection assessment of the traditional post-classification method (pixels).

Test data	Reference data			User's acc.
	Non-change	Change	Total	
Non-change	7454	118	7572	98.4%
Change	478	1591	2069	76.9%
Total	7932	1709	9641	
Producer's acc.	94.0%	93.1%		
Overall acc.	93.8%			
Kappa	0.804			

Table B4
Change detection assessment of the proposed method (pixels).

Test data	Reference data			User's acc.
	Non-change	Change	Total	
Non-change	7656	77	7733	99.0%
Change	276	1632	1908	85.5%
Total	7932	1709	9641	
Producer's acc.	96.5%	95.5%		
Overall acc.	96.3%			
Kappa	0.880			

Appendix C. Evaluation with different thresholds of class probability filter

In order to evaluate the performance with different thresholds, the accuracy assessments for the different methods with respect to different thresholds of class probability filter are shown in Figs. C1 and C2. The "S" and "H" after the change detection represent the different fusion methods, namely soft fusion and hard fusion. The horizontal dashed line indicates the result of post-classification. When the threshold is 1, it means that the filter for the class probability is of no use. If one curve is above the other curves, it means that the corresponding method is superior to the other methods despite of the class probability filter.
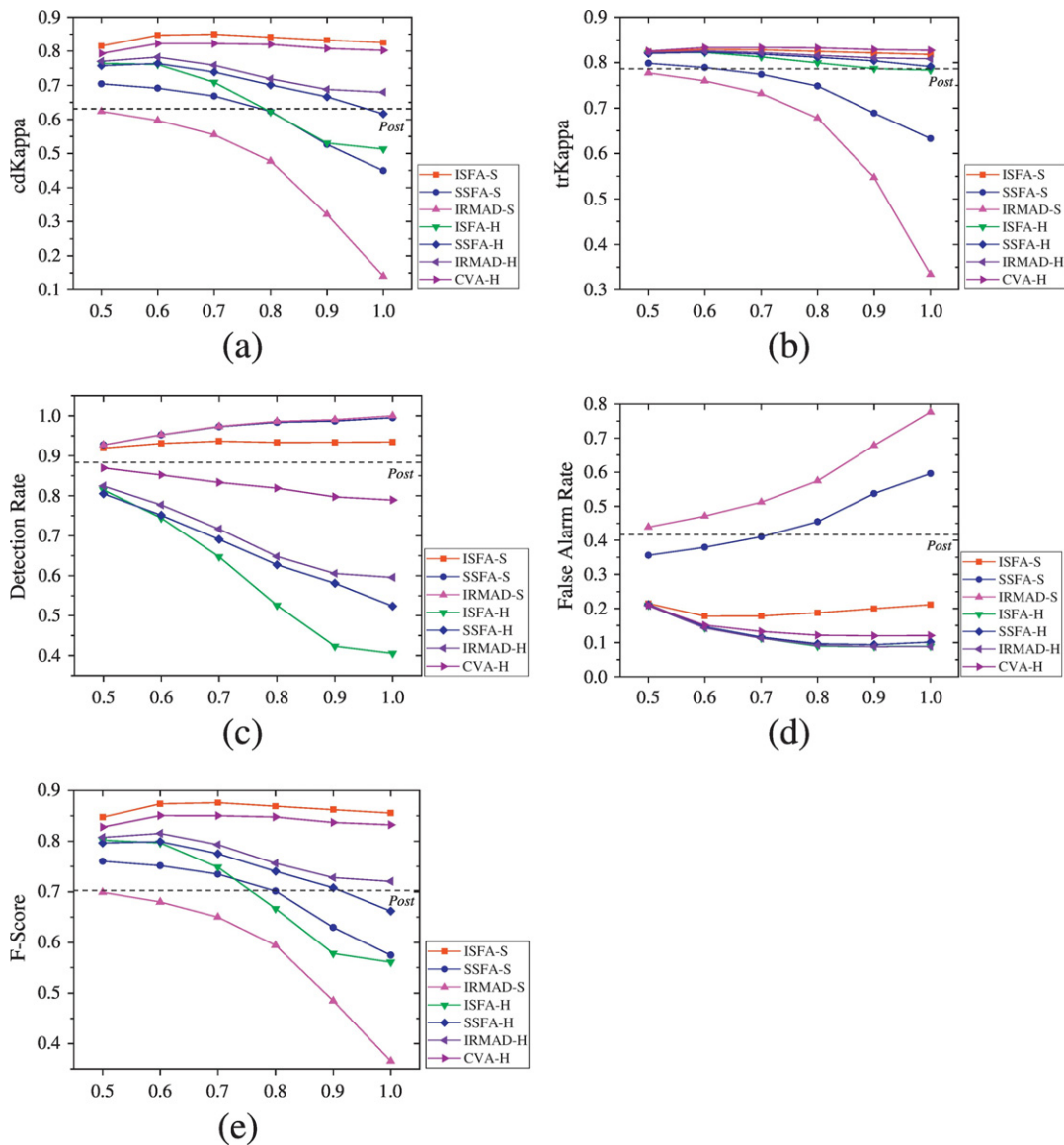

Fig. C1. Accuracy assessment with different thresholds for: (a) change/non-change; (b) "from-to" land-cover transition; (c) detection rate; (d) false alarm rate; and (e) F-score.

In Nanjing TM dataset, Fig. C1 (a) and (b) show that the proposed method obtains the highest change detection accuracy and the second highest accuracy for the "from-to" land-cover transition, in all cases. Fusing the DR and FAR in Fig. C1 (c) and (d), the F-score shown in (e) demonstrates that the proposed method performs better than all the other methods in the same condition. Fig. C1 also demonstrates that the class probability filter has the ability to improve the performance of all the methods, and the threshold has an influence on the final result.
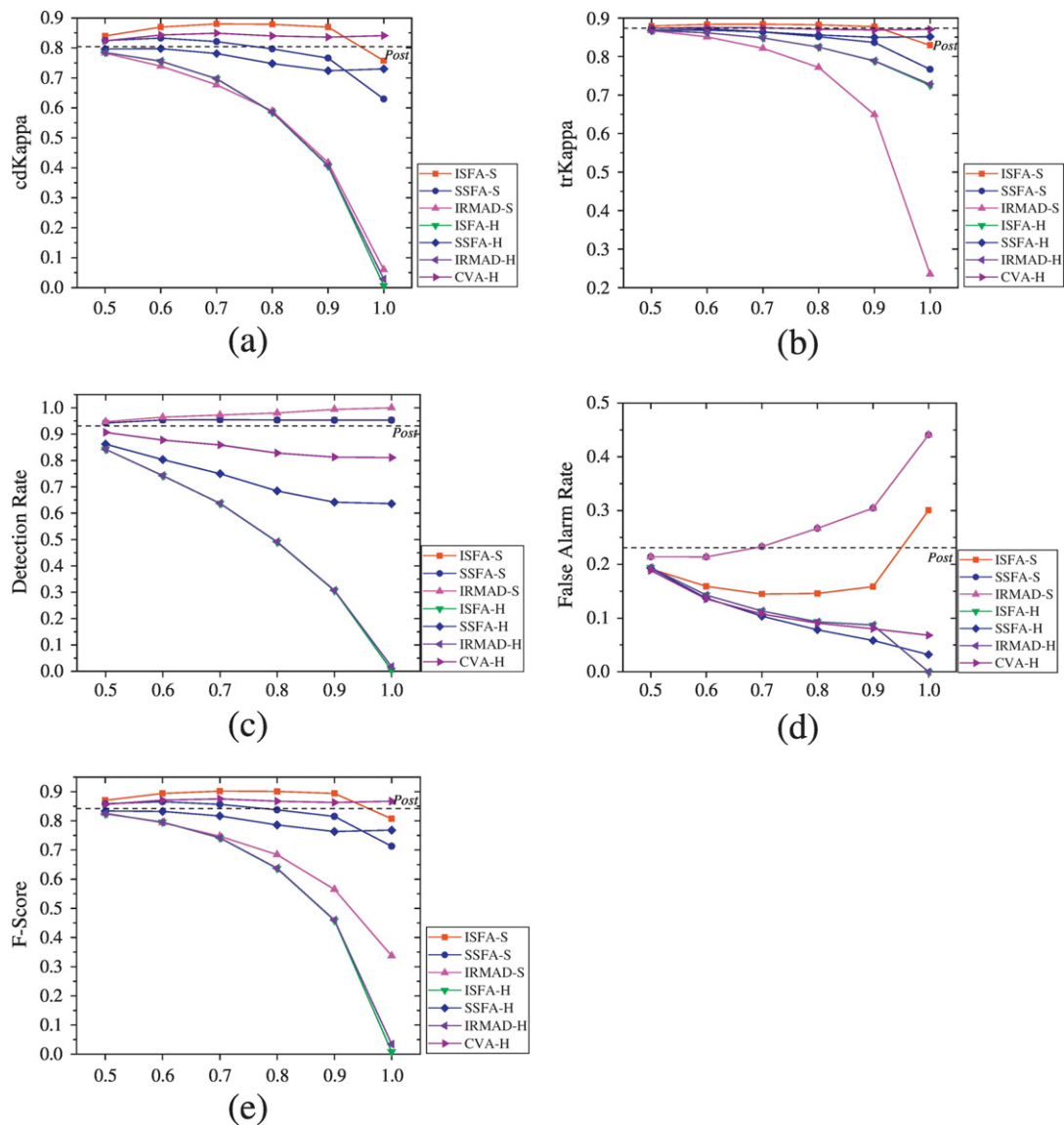

Fig. C2. Accuracy assessment with different thresholds for: (a) change/non-change; (b) "from-to" land-cover transition; (c) detection rate; (d) false alarm rate; and (e) F-score.

In Maanshan Dataset, the accuracy assessments with different thresholds of class probability filter for each improved change detection method are shown in Fig. C2. Except for the threshold of 1, the curves of the proposed method are all above the other methods in Fig. C2 (a) and (b). Furthermore, in the comprehensive evaluation with the F-score shown in Fig. C2 (e), the proposed method also shows a better ability to detect changes. Fig. C2 also illustrates that the class probability filter improves the performances of all the methods.

References

Ahlqvist, O., 2008. Extending post-classification change detection using semantic similarity metrics to overcome class heterogeneity: a study of 1992 and 2001 U.S. National Land Cover Database changes. Remote Sens. Environ. 112, 1226–1241.

Berkes, P., 2005. Pattern Recognition with Slow Feature Analysis. Cognitive Sciences EPrint Archive (CogPrint), p. v4104.

Berkes, P., Wiskott, L., 2005. Slow feature analysis yields a rich repertoire of complex cell properties. J. Vis. 5, 579–601.

Bindschadler, R.A., Scambos, T.A., Choi, H., Haran, T.M., 2010. Ice sheet change detection by satellite image differencing. Remote Sens. Environ. 114, 1353–1362.

Bovolo, F., Bruzzone, L., 2007. A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain. IEEE Trans. Geosci. Remote Sens. 45, 218–236.

Bruzzone, L., Serpico, S.B., 1997. An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images. IEEE Trans. Geosci. Remote Sens. 35, 858–867.

Canty, M.J., Nielsen, A.A., 2008. Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation. Remote Sens. Environ. 112, 1025–1036.

Cardille, J.A., Fortin, J.A., 2016. Bayesian updating of land-cover estimates in a data-rich environment. Remote Sens. Environ. 186, 234–249.

Celik, T., 2009. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. IEEE Geosci. Remote Sens. Lett. 6, 772–776.

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 1–27.

Chen, J., Gong, P., He, C., Pu, R., Shi, P., 2003. Land-use/land-cover change detection using improved change-vector analysis. Photogramm. Eng. Remote Sens. 69, 369–380.

Chen, G., Hay, G.J., Carvalho, L.M.T., Wulder, M.A., 2012. Object-based change detection. Int. J. Remote Sens. 33, 4434–4457.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., Lambin, E., 2004. Digital change detection methods in ecosystem monitoring: a review. Int. J. Remote Sens. 25, 1565–1596.

Fu, P., Weng, Q., 2016. A time series analysis of urbanization induced land use and land cover change and its impact on land surface temperature with Landsat imagery. Remote Sens. Environ. 175, 205–214.

Goutte, C., Gaussier, E., 2005. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In: Losada, D.E., Fernández-Luna, J.M. (Eds.), Advances in Information Retrieval: 27th European Conference on IR Research, ECIR 2005, Santiago de Compostela, Spain, March 21–23, 2005. Proceedings (pp. 345–359). Berlin, Heidelberg: Springer Berlin Heidelberg.

Hu, J., Zhang, Y., 2013. Seasonal change of land-use/land-cover (LULC) detection using MODIS data in rapid urbanization regions: A Case Study of the Pearl River Delta Region (China). IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 1–8.

Huang, Z., Jia, X.P., Ge, L.L., 2010. Sampling approaches for one-pass land-use/land-cover change mapping. Int. J. Remote Sens. 31, 1543–1554.

Huang, X., Schneider, A., Friedl, M.A., 2016. Mapping sub-pixel urban expansion in China using MODIS and DMSP/OLS nighttime lights. Remote Sens. Environ. 175, 92–108.

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.

Kennedy, R.E., Townsend, P.A., Gross, J.E., Cohen, W.B., Bolstad, P., Wang, Y.Q., Adams, P., 2009. Remote sensing change detection tools for natural resource managers: understanding concepts and tradeoffs in the design of landscape monitoring projects. Remote Sens. Environ. 113, 1382–1396.

Lancaster, H.O., Seneta, E., 2005. Chi-Square Distribution. Wiley Online Library.

Li, W., Ouyang, Z., Zhou, W., Chen, Q., 2011. Effects of spatial resolution of remotely sensed data on estimating urban impervious surfaces. J. Environ. Sci. 23, 1375–1383.

Lin, S., Kui, J., Tsung-Han, C., Yuqiang, F., Gang, W., Shuicheng, Y., 2014. DL-SFA: deeply-learned slow feature analysis for action recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2625–2632.

Lu, D., Mausel, P., Brondizio, E., Moran, E., 2004. Change detection techniques. Int. J. Remote Sens. 25, 2365–2401.

Nemmour, H., Chibani, Y., 2006. Multiple support vector machines for land cover change detection: an application for mapping urban extensions. ISPRS J. Photogramm. Remote Sens. 61, 125–133.

Nielsen, A.A., 2007. The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data. IEEE Trans. Image Process. 16, 463–478.

Nielsen, E.M., Prince, S.D., Koeln, G.T., 2008. Wetland change mapping for the U.S. mid-Atlantic region using an outlier detection technique. Remote Sens. Environ. 112, 4061–4074.

Otsu, N., 1975. A threshold selection method from gray-level histograms. Automatica 11, 23–27.

Serra, P., Pons, X., Saurí, D., 2003. Post-classification change detection with data from different sensors: some accuracy considerations. Int. J. Remote Sens. 24, 3311–3340.

Seto, C.K., Fragkias, M., 2005. Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics. Landsc. Ecol. 20, 871–888.

Singh, A., 1989. Review article digital change detection techniques using remotely-sensed data. Int. J. Remote Sens. 10, 989–1003.

Sokolova, M., Japkowicz, N., Szpakowicz, S., 2006. Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In: Sattar, A., Kang, B.-H. (Eds.), AI 2006: Advances in Artificial Intelligence: 19th Australian Joint Conference on Artificial Intelligence, Hobart, Australia, December 4–8, 2006. Proceedings (pp. 1015–1021). Berlin, Heidelberg: Springer Berlin Heidelberg.

Song, C., Huang, B., Ke, L., Richards, K.S., 2014. Remote sensing of alpine lake water environment changes on the Tibetan plateau and surroundings: a review. ISPRS J. Photogramm. Remote Sens. 92, 26–37.

Tang, Y., Zhang, L., 2017. Urban change analysis with multi-sensor multispectral imagery. Remote Sens. 9, 252.

Wang, Z., Zhang, L., Fang, T., Mathiopoulos, P.T., Tong, X., Qu, H., Xiao, Z., Li, F., Chen, D., 2015. A multiscale and hierarchical feature extraction method for terrestrial laser scanning point cloud classification. IEEE Trans. Geosci. Remote Sens. 53, 2409–2425.

Wilbert, N., Sprekeler, H., Franzius, M., Berkes, P., Wiskott, L., 2011. Slow feature analysis. Scholarpedia 6, 5282.

Wiskott, L., Sejnowski, T.J., 2002. Slow feature analysis: unsupervised learning of invariances. Neural Comput. 14, 715–770.

Wu, C., Du, B., Zhang, L., 2014. Slow feature analysis for change detection in multispectral imagery. IEEE Trans. Geosci. Remote Sens. 52, 2858–2874.

Wu, C., Zhang, L., Du, B., 2015. Hyperspectral anomaly change detection with slow feature analysis. Neurocomputing 151, 175–187 (Part 1).

Xian, G., Homer, C., 2010. Updating the 2001 National Land Cover Database impervious surface products to 2006 using Landsat imagery change detection methods. Remote Sens. Environ. 114, 1676–1686.

Xian, G., Homer, C., Fry, J., 2009. Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods. Remote Sens. Environ. 113, 1133–1147.

Xiao, J., Shen, Y., Ge, J., Tateishi, R., Tang, C., Liang, Y., Huang, Z., 2006. Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing. Landsc. Urban Plan. 75, 69–80.

Yang, J., Weisberg, P.J., Bristow, N.A., 2012. Landsat remote sensing approaches for monitoring long-term tree cover dynamics in semi-arid woodlands: comparison of vegetation indices and spectral mixture analysis. Remote Sens. Environ. 119, 62–71.

Yao, X., Han, J., Cheng, G., Qian, X., Guo, L., 2016. Semantic annotation of high-resolution satellite images via weakly supervised learning. IEEE Trans. Geosci. Remote Sens. 54, 3660–3671.

Yu, W., Zhou, W., Qian, Y., Yan, J., 2016. A new approach for land cover classification and change analysis: integrating backdating and an object-based method. Remote Sens. Environ. 177, 37–47.

Yuan, F., Sawaya, K.E., Loeffelholz, B.C., Bauer, M.E., 2005. Land cover classification and change analysis of the twin cities (Minnesota) metropolitan area by multitemporal Landsat remote sensing. Remote Sens. Environ. 98, 317–328.

Zhang, Z., Tao, D., 2012. Slow feature analysis for human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. 34, 436–450.

Zhang, J., Zhang, Y., 2007. Remote sensing research issues of the National Land use change program of China. ISPRS J. Photogramm. Remote Sens. 62, 461–472.

Zhang, L., Wu, C., Du, B., 2014. Automatic radiometric normalization for multitemporal remote sensing imagery with iterative slow feature analysis. IEEE Trans. Geosci. Remote Sens. 52, 6141–6155.

Zhao, X., Stein, A., Chen, X.-L., 2011. Monitoring the dynamics of wetland inundation by random sets on multi-temporal images. Remote Sens. Environ. 115, 2390–2401.

Zillmann, E., Gonzalez, A., Montero Herrero, E.J., van Wolvelaer, J., Esch, T., Keil, M., Weichelt, H., Garzon, A.M., 2014. Pan-European grassland mapping using seasonal statistics from multisensor image time series. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 7, 3461–3472.