

# Kernel Slow Feature Analysis for Scene Change Detection

Chen Wu, *Member, IEEE*, Liangpei Zhang, *Senior Member, IEEE*, and Bo Du, *Senior Member, IEEE*

**Abstract**—Scene change detection between multitemporal image scenes can be used to interpret the variation of regional land use, and has significant potential in the application of urban development monitoring at the semantic level. The traditional methods directly comparing the independent semantic classes neglect the temporal correlation, and thus suffer from accumulated classification errors. In this paper, we propose a novel scene change detection method via kernel slow feature analysis (KSFA) and postclassification fusion, which integrates independent scene classification with scene change detection to accurately determine scene changes and identify the “from-to” transition type. After representation with the bag-of-visual-words model, KSFA is proposed to extract the nonlinear temporally invariant features, to better measure the change probability between corresponding multitemporal image scenes. Two postclassification fusion methods, which are based on Bayesian theory and predefined rules, respectively, are then employed to identify the optimal coupled class combinations of multitemporal scene pairs. Furthermore, in addition to identifying semantic changes, the proposed method can also improve the performance of scene classification, since the unchanged scenes are more likely to belong to the same class. Two experiments with high-resolution remote sensing image scene data sets confirm that the proposed method can increase the accuracy of scene change detection, scene transition identification, and scene classification.

**Index Terms**—Bayesian theory, kernel slow feature analysis (KSFA), rule-based method, remote sensing, scene change detection.

## I. INTRODUCTION

SCENE classification for classifying image scenes into semantic categories is one of the most important topics in the field of computer vision. It has been widely used in many practical applications, including image retrieval, content-based image filtering, and image understanding [1]–[3]. Scene classification for remote sensing imagery has also attracted a lot of attention due to its significant potential

Manuscript received April 30, 2016; revised September 2, 2016 and November 21, 2016; accepted December 10, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61601333, Grant 61471274, and Grant 41431175, and in part by the China Postdoctoral Science Foundation under Grant 2015M580667 and Grant 2016T90733.

C. Wu is with the International School of Software, and the Spatial Information and Digital Engineering Research Center, Wuhan University, Wuhan 430079, China (e-mail: chen.wu@whu.edu.cn).

L. Zhang is with the Remote Sensing Group, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zlp62@whu.edu.cn).

B. Du is with the School of Computer Science, and Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430072, China (e-mail: gunspace@163.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2016.2642125

in interpreting urban land-use distribution at the semantic level [4]–[7]. Compared with classifying pixels or objects, high-resolution remote sensing data can enable us to assign image scenes with semantic land-use labels, such as industrial region, commercial region, and residential region, based on the landscape distribution and structural patterns encoded in the image [5]. Many methods have been proposed to identify the semantic labels of remote sensing image scenes, such as the bag-of-visual-words (BOVW) model [8], [9], sparse coding [5], [6], the topic model [7], [10], [11], and deep learning [12], [13].

Although numerous studies have addressed scene classification with a single image, few studies have focused on scene change detection with multitemporal image scenes. The main reason for this may be that there is no specific need in real applications to compare the differences in the semantic labels of two natural image scenes, since natural images do not have specific corresponding relationships with other images. However, for remote sensing data, the label change of multitemporal image scenes covering the same region and acquired at different times can indicate the land-use variation at the semantic level. For example, the appearance of residential and commercial regions may indicate the development of a city. Thus, detecting scene changes and identifying “from-to” transition types are both extremely useful approaches in urban development research and expansion monitoring [14]. Furthermore, the temporal correlation information between multitemporal remote sensing image scenes can also be taken into consideration to improve the performance of scene classification, since the unchanged scenes are more likely to belong to the same category.

For multitemporal remote sensing imagery, pixel-wise and object-wise change detection has been intensively studied as one of the most important research topics in earth observation technology [15]–[18]. The existing methods can be mainly categorized into three groups: 1) image algebra, which is used to calculate the difference between multitemporal images [19], [20]; 2) image transformation, which transforms the multitemporal data into a new feature space to extract information and make a comparison [21]–[24]; and 3) classification-based methods, which compare the class labels of the corresponding landscapes after independent classification, or classify the landscape transition as one class [25], [26]. However, the traditional change detection methods only detect the transitions of landscapes, and cannot effectively identify the change of an image scene, since the variation of the thematic objects inside the image does not modify the semantic classes of the image scenes directly.

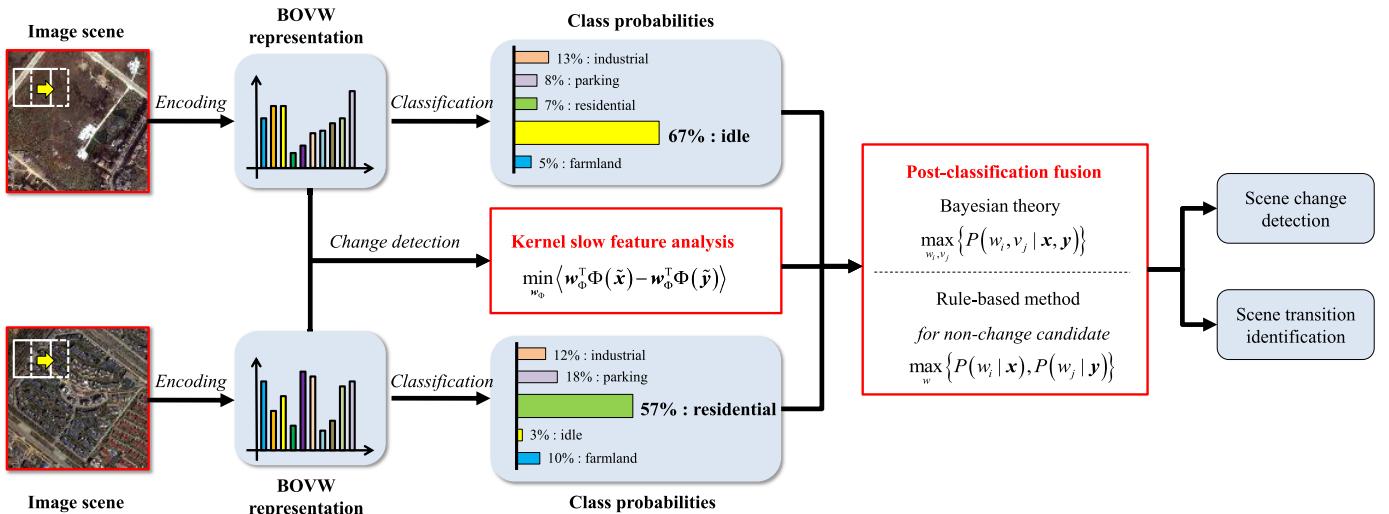


Fig. 1. Procedure of the proposed scene change detection method. The processes in the red squares are the two main contributions of this paper. Details are presented in the main text.

Therefore, it is necessary to develop a scene change detection method at the semantic level.

Wu *et al.* [27] proposed a scene change detection approach utilizing the BOVW model and a postclassification method. This straightforward approach involves comparing the class labels from independent scene classification. However, the performance of the scene change detection is limited since the temporal correlation is ignored and the misclassification errors accumulate. In previous studies, a simple and effective method was proposed to improve the performance of postclassification. This method utilizes change vector analysis (CVA) to obtain the change information of multitemporal images, and keeps the class consistency of the unchanged pixels [25], [28], [29]. Therefore, inspired by this approach, we focus on developing an effective change detection algorithm to measure the change probability of multitemporal scene pairs, and integrate independent scene classification with change probability to improve the performance of scene change detection, transition identification, and classification.

Slow feature analysis (SFA) is an effective unsupervised learning algorithm that aims to extract temporally invariant features from quickly varying input signals [30], [31]. It has been successfully applied in robust object recognition [32], dynamic scene classification [33], and action recognition [34], [35]. SFA has also been utilized in traditional change detection for multitemporal remote sensing data [22], [36], where the basic idea is to learn the most invariant features of the unchanged samples so as to highlight the feature differences in real changed landscapes. However, the representative features of image scenes are much more complicated in a high-dimensional feature space. Therefore, in this paper, we propose kernel slow feature analysis (KSFA) with supervised training samples to determine a nonlinear high-level representation for the accurate change detection of semantic scene classes.

After the change probabilities are determined by KSFA, we then propose two postclassification fusion methods to integrate independent scene classification with the

temporal information. The first method is based on Bayesian theory, where the *a posteriori* probability of each coupled class combination is calculated to determine the final semantic classes of the multitemporal scene pairs. The second method is a rule-based method, in which the multitemporal scene sets are first distinguished as change or nonchange, and then the labels of the unchanged candidates are corrected based on predefined rules, according to the class probabilities from the independent classification.

Therefore, in this paper, we propose a novel scene change detection method with KSFA and postclassification fusion. The image scenes are first represented with the BOVW model. Support vector machine (SVM) is then used for the independent scene classification. KSFA with a histogram intersection kernel (HIK) is then employed to obtain the change probabilities of the multitemporal scene pairs. Finally, postclassification fusion with Bayesian theory or a rule-based method is applied to accurately determine the coupled class combinations of scene pairs for change detection, fusing the independent classification and temporal correlation interpreted by KSFA.

The rest of this paper is organized as follows. Section II details the proposed scene change detection method. The experiments and discussions are presented in Section III. Finally, the conclusion is drawn in Section IV.

## II. METHODOLOGY

In this section, we elaborate how to integrate KSFA and postclassification fusion with independent scene classification to improve the performance of scene change detection. The whole procedure of the proposed method is shown in Fig. 1. The main steps are as follows:

- 1) encode the multitemporal image scenes with the BOVW model to obtain the frequency features;
- 2) classify the image scenes with the representative features and SVM algorithm to obtain the class probabilities;
- 3) implement the KSFA change detection algorithm to obtain the change probabilities of the multitemporal scene pairs;

- 4) employ postclassification fusion to integrate the class probabilities with the change probabilities for the optimal class combinations;
- 5) output the results of the scene change detection and scene transition identification.

#### A. BOVW Model and Classification

Machine learning has been widely used in the field of computer vision [37]–[40]. The BOVW model is a classical representation approach for scene classification, which is aimed at encoding image scenes with the statistical frequency of the contained visual words [41]. The BOVW model is very robust in practical cases, and is thus selected as the foundation of the proposed approach [9].

For an image scene, a small patch is regarded as one visual word. The features of the visual words will obviously influence the final classification performance. The mean value and standard deviation of each band are simple but effective descriptors in scene classification [11], [27]. Therefore, the feature vector of each visual word is constructed by stacking the mean values and standard deviations. In the experiments, we compared the performance of the BOVW model with those of the raw pixel vector and the color dense scale-invariant feature transform (SIFT) descriptor. The raw pixel vector is a widely used low-level feature that arranges the pixels in the image patch into a vector band by band [42]. The color dense SIFT descriptor is very robust and effective, and is obtained by stacking the dense SIFT feature of each spectral band [43], [44].

The dictionary needs to be comprehensive and representative to quantize the features of the visual words. For computational efficiency, we randomly select numerous sample patches from the image scenes in the data set, and we then learn the dictionary with  $k$ -means clustering. The input data for the  $k$ -means clustering are preprocessed by zero component analysis (ZCA). ZCA is performed, respectively, for the training patch acquired from each data set to remove the redundancy and reduce the spectral variance [45]. In the proposed method, since the multitemporal image scenes are acquired at different times, they should be encoded into the same feature space to calculate the accurate change probability. Therefore, the union dictionary is used as the unique dictionary for feature encoding [27]. It is learned by gathering equal quantities of training patches from the multitemporal data sets, and is then used to encode all the training and test scenes.

After the dictionary is obtained, all the training and test scenes are represented with the frequencies of the visual words encoded in the image. The visual words are extracted with overlapping patches, which cover the whole image in squared windows, moving in steps of  $s$  horizontally and vertically. The ZCA whitening matrix obtained from the dictionary learning is also applied to these visual words. The visual words are then quantized to the most similar elements in the dictionary, where the similarity is measured by the L2-norm.

Classification is implemented with the SVM algorithm and LIBSVM code [46], since it can provide accurate *a posteriori* probabilities of each class for the test scenes. The kernel used

in the SVM classification has great influence on the final performance. In this paper, the HIK is selected, which has been proven to be effective in scene classification with BOVW representation, and has no parameters to set [47]–[50]. For the two frequency features  $\mathbf{x}$  and  $\mathbf{y}$  of the image scene, where its dimension is  $k$ , the HIK can be expressed as follows:

$$K(\mathbf{x}, \mathbf{y}) = \sum_{b=1}^k \min\{\mathbf{x}_b, \mathbf{y}_b\} \quad (1)$$

where  $\mathbf{x}_b$  indicates the value of feature vector  $\mathbf{x}$  in dimension  $b$ .

#### B. KSFA Change Detection

1) *Slow Feature Analysis*: In most cases, original input signals or data are very sensitive to small changes in the environment over time, so that they vary very quickly. However, when observing certain phenomena or objects continuously, they can be seen to have some common structures, which lead to statistical regularities in the signals or data [30]. Therefore, a high-level representation of temporal observation data usually changes very slowly.

The basic idea of SFA is to extract the slowly varying features from the original quickly varying input signals to discover the underlying essential pattern. Mathematically, SFA can be explained as an optimization problem: given a multidimensional temporal signal  $\mathbf{s}(t) = [s_1(t), \dots, s_M(t)]^T$ , where  $t \in [t_0, t_1]$  indicates the time, we want to find a set of functions  $g_1(s), \dots, g_M(s)$  to confirm that the transformed output signals will be temporally invariant, which means that

$$\min_{g_j} \langle (\dot{g}_j(s))^2 \rangle_t \quad (2)$$

under the constraints

$$\langle g_j(s) \rangle_t = 0 \text{ zero mean} \quad (3)$$

$$\langle (g_j(s))^2 \rangle_t = 1 \text{ unit variance} \quad (4)$$

$$\forall i < j : \langle g_i(s) g_j(s) \rangle_t = 0 \text{ decorrelation and order} \quad (5)$$

where the bracket  $\langle \cdot \rangle_t$  indicates the mean value over time, and  $\dot{g}$  represents the first-order derivative of the output signal. The objective of this optimization problem is to minimize the temporal variance of the output signal measured by the mean of the power of its first-order derivative, shown as (2). Constraints (3) and (4) normalize all the output signals to the same scale to ensure that their temporal variances are comparable. Constraint (3) also simplifies the optimization problem. Constraint (4) forces the output features to contain some information, and constraint (5) ensures that the different output signals are mutually uncorrelated.

In a linear case, the transformation function can be shown as  $g_j(s) = \mathbf{w}_j^T \mathbf{s}$ . The objective and constraints can be expressed as

$$\langle (\mathbf{w}_j^T \dot{\mathbf{s}})^2 \rangle_t = \mathbf{w}_j^T \langle \dot{\mathbf{s}} \dot{\mathbf{s}}^T \rangle_t \mathbf{w}_j = \mathbf{w}_j^T \mathbf{A} \mathbf{w}_j \quad (6)$$

$$\langle (\mathbf{w}_j^T \mathbf{s})(\mathbf{w}_j^T \mathbf{s})^T \rangle_t = \mathbf{w}_j^T \langle \mathbf{s} \mathbf{s}^T \rangle_t \mathbf{w}_j = \mathbf{w}_j^T \mathbf{B} \mathbf{w}_j = 1 \quad (7)$$

$$\langle (\mathbf{w}_i^T \mathbf{s})(\mathbf{w}_j^T \mathbf{s})^T \rangle_t = \mathbf{w}_i^T \langle \mathbf{s} \mathbf{s}^T \rangle_t \mathbf{w}_j = \mathbf{w}_i^T \mathbf{B} \mathbf{w}_j = 0. \quad (8)$$

In the theory of SFA, constraint (4) can be integrated into the optimization objective as [51]

$$\frac{\langle (\mathbf{w}_j^T \dot{\mathbf{s}})^2 \rangle_t}{\langle (\mathbf{w}_j^T \mathbf{s})(\mathbf{w}_j^T \mathbf{s}) \rangle_t} = \frac{\mathbf{w}_j^T \mathbf{A} \mathbf{w}_j}{\mathbf{w}_j^T \mathbf{B} \mathbf{w}_j}. \quad (9)$$

Thus, the optimization problem can then be solved by the generalized eigenvalue problem

$$\mathbf{A}\mathbf{W} = \mathbf{B}\mathbf{W}\Lambda \quad (10)$$

where  $\mathbf{A} = \langle \dot{\mathbf{s}}\dot{\mathbf{s}}^T \rangle_t$  and  $\mathbf{B} = \langle \mathbf{s}\mathbf{s}^T \rangle_t$  are the covariance matrices of the first-order derivative and original input signals. The output signals are sorted by the order of the corresponding eigenvalues  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_M$ , where the most invariant component has the lowest index.

For multitemporal data sets, such as multitemporal remote sensing images, the environmental changes between the different acquisition times will lead to differences in the observation features, even for the same landscape. This feature variance will result in pseudo-changes, which are one of the main error sources in change detection. SFA is thus proposed to extract the temporally invariant component from the original multitemporal images, so that the separability of the real changes can be improved. Since it was originally designed for use with continuous signals, SFA can be reformulated for discrete cases by replacing the derivate with the finite differences of the samples in the data set [22].

The reconstructed SFA can be expressed as follows: the multidimensional feature vector pair are represented as  $\mathbf{x}_i$  and  $\mathbf{y}_i$ , where  $i$  is the number of samples. After normalization into zero mean and unit variance, the SFA algorithm can be rewritten as

$$\min_{\mathbf{w}_j} \frac{1}{n} \sum_{i=1}^n (\mathbf{w}_j^T \mathbf{x}_i - \mathbf{w}_j^T \mathbf{y}_i)^2 \quad (11)$$

under the constraints

$$\frac{1}{2n} \left[ \sum_{i=1}^n \mathbf{w}_j^T \mathbf{x}^i + \sum_{i=1}^n \mathbf{w}_j^T \mathbf{y}^i \right] = 0 \quad (12)$$

$$\frac{1}{2n} \left[ \sum_{i=1}^n (\mathbf{w}_j^T \mathbf{x}^i)^2 + \sum_{i=1}^n (\mathbf{w}_j^T \mathbf{y}^i)^2 \right] = 1 \quad (13)$$

$$\frac{1}{2n} \left[ \sum_{i=1}^n (\mathbf{w}_j^T \mathbf{x}^i)(\mathbf{w}_l^T \mathbf{x}^i) + \sum_{i=1}^n (\mathbf{w}_j^T \mathbf{y}^i)(\mathbf{w}_l^T \mathbf{y}^i) \right] = 0. \quad (14)$$

In this algorithm, the optimization problem can also be solved with (10), where matrices  $\mathbf{A}$  and  $\mathbf{B}$  are reformulated as

$$\mathbf{A} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \mathbf{y}_i)(\mathbf{x}_i - \mathbf{y}_i)^T = \Sigma_\Delta \quad (15)$$

$$\mathbf{B} = \frac{1}{2n} \left[ \sum_{i=1}^n (\mathbf{x}_i)(\mathbf{x}_i)^T + \sum_{i=1}^n (\mathbf{y}_i)(\mathbf{y}_i)^T \right] = \frac{1}{2}(\Sigma_X + \Sigma_Y) \quad (16)$$

where  $\Sigma_X$  and  $\Sigma_Y$  are the covariance matrices of the multitemporal data set, and  $\Sigma_\Delta$  is the covariance matrix of their difference.

After the transformation vectors are obtained, they can be normalized to fulfill the constraints as follows:

$$\hat{\mathbf{w}}_j = \frac{\mathbf{w}_j}{\sqrt{\mathbf{w}_j^T \mathbf{B} \mathbf{w}_j}}. \quad (17)$$

Finally, the difference between the transformed features can be calculated to extract the change information as  $\mathbf{d}_j = \hat{\mathbf{w}}_j^T \mathbf{x} - \hat{\mathbf{w}}_j^T \mathbf{y}$ .

2) *Kernel Slow Feature Analysis*: Although SFA has been proven to be effective in computer vision and machine learning, the representation pattern of image scenes may be too complex to analyze in a linear feature space. Therefore, we propose KSFA to map the original features into a nonlinear high-dimensional space, to better extract the essential high-level representation of the multitemporal image scenes.

Compared with the kernel version proposed in [52] and [53], the proposed KSFA is especially formulated for the discrete case of multitemporal scene data. Furthermore, the optimization solution and its implementation, which is more commonly used in SFA theory [34], [35], differs from that of the existing kernel method [52], [53].

For a given nonlinear mapping function  $\Phi$ , the original multitemporal input data  $\mathbf{x}$  and  $\mathbf{y}$  can be mapped into a high-dimensional nonlinear feature space

$$\mathbf{x} \rightarrow \Phi(\mathbf{x}), \quad \mathbf{y} \rightarrow \Phi(\mathbf{y}). \quad (18)$$

The idea of KSFA is to solve the problem of SFA in the new nonlinear feature space. This can be achieved by solving the following eigenvalue problem (10):

$$\mathbf{A}_\Phi \mathbf{W}_\Phi = \mathbf{B}_\Phi \mathbf{W}_\Phi \Lambda_\Phi. \quad (19)$$

Since the eigenvectors can be expressed by a linear combination of the samples [54], we can rewrite the transformation vector  $\mathbf{w}_\Phi$  as

$$\mathbf{w}_\Phi = \sum_{i=1}^n \alpha_i \tilde{\Phi}(\mathbf{x}_i) + \sum_{i=1}^n \beta_i \tilde{\Phi}(\mathbf{y}_i) \quad (20)$$

where  $\tilde{\Phi}(\mathbf{x}_i)$  is the centralized mapped feature, and  $n$  is the number of training samples.  $\mathbf{X}$  and  $\mathbf{Y}$  have the same number of samples in the change detection problem. Thus, combining (15) and (20),  $\mathbf{A}_\Phi \mathbf{W}_\Phi$  in (19) can be reformulated as

$$\begin{aligned} \mathbf{A}_\Phi \mathbf{W}_\Phi &= \frac{1}{n} \left[ \sum_{i=1}^n (\tilde{\Phi}(\mathbf{x}_i) - \tilde{\Phi}(\mathbf{y}_i))(\tilde{\Phi}(\mathbf{x}_i) - \tilde{\Phi}(\mathbf{y}_i))^T \right] \\ &\quad \cdot \left[ \sum_{j=1}^n \alpha_j \tilde{\Phi}(\mathbf{x}_j) + \sum_{j=1}^n \beta_j \tilde{\Phi}(\mathbf{y}_j) \right]. \end{aligned} \quad (21)$$

We can then instead consider the following formula by expanding the equations [54], [55]:

$$\begin{aligned} &[\tilde{\Phi}(\mathbf{x}_1), \dots, \tilde{\Phi}(\mathbf{x}_n)]^T n \mathbf{A}_\Phi \mathbf{W}_\Phi \\ &= \tilde{\mathbf{K}}_{XX} \tilde{\mathbf{K}}_{XX} \alpha - \tilde{\mathbf{K}}_{XY} \tilde{\mathbf{K}}_{XX} \alpha - \tilde{\mathbf{K}}_{XY} \tilde{\mathbf{K}}_{YY} \alpha + \tilde{\mathbf{K}}_{XY} \tilde{\mathbf{K}}_{YX} \alpha \\ &\quad + \tilde{\mathbf{K}}_{XX} \tilde{\mathbf{K}}_{XY} \beta - \tilde{\mathbf{K}}_{XY} \tilde{\mathbf{K}}_{XY} \beta - \tilde{\mathbf{K}}_{XX} \tilde{\mathbf{K}}_{YY} \beta + \tilde{\mathbf{K}}_{XY} \tilde{\mathbf{K}}_{YY} \beta \end{aligned} \quad (22)$$

$$\begin{aligned} &[\tilde{\Phi}(\mathbf{y}_1), \dots, \tilde{\Phi}(\mathbf{y}_n)]^T n \mathbf{A}_\Phi \mathbf{W}_\Phi \\ &= \tilde{\mathbf{K}}_{YY} \tilde{\mathbf{K}}_{XX} \alpha - \tilde{\mathbf{K}}_{YY} \tilde{\mathbf{K}}_{XX} \alpha - \tilde{\mathbf{K}}_{YX} \tilde{\mathbf{K}}_{YY} \alpha + \tilde{\mathbf{K}}_{YY} \tilde{\mathbf{K}}_{YX} \alpha \\ &\quad + \tilde{\mathbf{K}}_{YX} \tilde{\mathbf{K}}_{XY} \beta - \tilde{\mathbf{K}}_{YY} \tilde{\mathbf{K}}_{XY} \beta - \tilde{\mathbf{K}}_{YX} \tilde{\mathbf{K}}_{YY} \beta + \tilde{\mathbf{K}}_{YY} \tilde{\mathbf{K}}_{YY} \beta. \end{aligned} \quad (23)$$

Thus, we can obtain the following formula:

$$\begin{aligned} & [\tilde{\Phi}(x_1), \dots, \tilde{\Phi}(x_n), \tilde{\Phi}(y_1), \dots, \tilde{\Phi}(y_n)]^T n \mathbf{A}_\Phi \mathbf{W}_\Phi \\ &= \begin{bmatrix} \tilde{K}_{XX}\tilde{K}_{XX} - \tilde{K}_{XY}\tilde{K}_{XX} & \tilde{K}_{XX}\tilde{K}_{XY} - \tilde{K}_{XY}\tilde{K}_{XY} \\ -\tilde{K}_{XY}\tilde{K}_{YX} + \tilde{K}_{XY}\tilde{K}_{YX} & -\tilde{K}_{XX}\tilde{K}_{YY} + \tilde{K}_{XY}\tilde{K}_{YY} \\ \tilde{K}_{YX}\tilde{K}_{XX} - \tilde{K}_{YY}\tilde{K}_{XX} & \tilde{K}_{YX}\tilde{K}_{XY} - \tilde{K}_{YY}\tilde{K}_{XX} \\ -\tilde{K}_{YX}\tilde{K}_{YX} + \tilde{K}_{YY}\tilde{K}_{YX} & -\tilde{K}_{YX}\tilde{K}_{YY} + \tilde{K}_{YY}\tilde{K}_{YY} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \\ &= \begin{bmatrix} \tilde{K}_{XX} - \tilde{K}_{XY} \\ \tilde{K}_{YX} - \tilde{K}_{YY} \end{bmatrix} [\tilde{K}_{XX} - \tilde{K}_{YX} \quad \tilde{K}_{XY} - \tilde{K}_{YY}] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (24) \end{aligned}$$

where  $\alpha = [\alpha_1, \dots, \alpha_n]^T$  and  $\beta = [\beta_1, \dots, \beta_n]^T$ , and we can define an  $n \times n$  kernel matrix  $\mathbf{K}$  as

$$(\tilde{\mathbf{K}}_{XY})_{ij} = (\tilde{\Phi}(x_i) \cdot \tilde{\Phi}(y_j)). \quad (25)$$

The same process can be implemented on  $\mathbf{B}_\Phi \mathbf{W}_\Phi$  as follows:

$$\begin{aligned} & [\tilde{\Phi}(x_1), \dots, \tilde{\Phi}(x_n), \tilde{\Phi}(y_1), \dots, \tilde{\Phi}(y_n)]^T n \mathbf{B}_\Phi \mathbf{W}_\Phi \\ &= \frac{1}{2} \begin{bmatrix} \tilde{K}_{XX}\tilde{K}_{XX} + \tilde{K}_{XY}\tilde{K}_{YX} & \tilde{K}_{XX}\tilde{K}_{XY} + \tilde{K}_{XY}\tilde{K}_{YY} \\ \tilde{K}_{YX}\tilde{K}_{XX} + \tilde{K}_{YY}\tilde{K}_{YX} & \tilde{K}_{YX}\tilde{K}_{XY} + \tilde{K}_{YY}\tilde{K}_{YY} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} \tilde{K}_{XX} & \tilde{K}_{XY} \\ \tilde{K}_{YX} & \tilde{K}_{YY} \end{bmatrix} \begin{bmatrix} \tilde{K}_{XX} & \tilde{K}_{XY} \\ \tilde{K}_{YX} & \tilde{K}_{YY} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \quad (26) \end{aligned}$$

Therefore, the generalized eigenvalue problem of KSFA can be rewritten as

$$\mathbf{A}_\mathbf{K} \mathbf{w}_\mathbf{K} = \mathbf{B}_\mathbf{K} \mathbf{w}_\mathbf{K} \Lambda_\mathbf{K} \quad (27)$$

where the matrices  $\mathbf{A}_\mathbf{K}$ ,  $\mathbf{B}_\mathbf{K}$ , and  $\mathbf{w}_\mathbf{K}$  can be expressed as

$$\mathbf{A}_\mathbf{K} = \begin{bmatrix} \tilde{K}_{XX} - \tilde{K}_{XY} \\ \tilde{K}_{YX} - \tilde{K}_{YY} \end{bmatrix} [\tilde{K}_{XX} - \tilde{K}_{YX} \quad \tilde{K}_{XY} - \tilde{K}_{YY}] \quad (28)$$

$$\mathbf{B}_\mathbf{K} = \frac{1}{2} \begin{bmatrix} \tilde{K}_{XX} & \tilde{K}_{XY} \\ \tilde{K}_{YX} & \tilde{K}_{YY} \end{bmatrix} \begin{bmatrix} \tilde{K}_{XX} & \tilde{K}_{XY} \\ \tilde{K}_{YX} & \tilde{K}_{YY} \end{bmatrix} \quad (29)$$

$$\mathbf{w}_\mathbf{K} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \quad (30)$$

In the above procedure, the centralized kernel matrix can be calculated by the original kernel matrix [56]

$$\tilde{K}_{XX} = K_{XX} - \frac{1}{n} \mathbf{1} K_{XX} - \frac{1}{n} K_{XX} \mathbf{1} + \frac{1}{n^2} \mathbf{1} K_{XX} \mathbf{1} \quad (31)$$

$$\tilde{K}_{YY} = K_{YY} - \frac{1}{n} \mathbf{1} K_{YY} - \frac{1}{n} K_{YY} \mathbf{1} + \frac{1}{n^2} \mathbf{1} K_{YY} \mathbf{1} \quad (32)$$

$$\tilde{K}_{XY} = K_{XY} - \frac{1}{n} \mathbf{1} K_{XY} - \frac{1}{n} K_{XY} \mathbf{1} + \frac{1}{n^2} \mathbf{1} K_{XY} \mathbf{1} \quad (33)$$

where  $\mathbf{1}$  is an  $n \times n$  matrix with unit elements.

After the eigenvectors are obtained by (27), they should be normalized so that constraint (4) can be fulfilled [51]. The normalization is achieved as [57]

$$\tilde{\mathbf{w}}_\mathbf{K} = \frac{\mathbf{w}_\mathbf{K}}{\sqrt{\mathbf{w}_\mathbf{K}^T \mathbf{B}_\mathbf{K} \mathbf{w}_\mathbf{K}}} \quad (34)$$

where the denominator can be calculated as

$$\mathbf{w}_\mathbf{K}^T \mathbf{B}_\mathbf{K} \mathbf{w}_\mathbf{K} = \frac{1}{2n} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^T \begin{bmatrix} \tilde{K}_{XX} & \tilde{K}_{XY} \\ \tilde{K}_{YX} & \tilde{K}_{YY} \end{bmatrix} \begin{bmatrix} \tilde{K}_{XX} & \tilde{K}_{XY} \\ \tilde{K}_{YX} & \tilde{K}_{YY} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \quad (35)$$

Finally, the projection of the KSFA component for the test samples  $\mathbf{Z}_X$  and  $\mathbf{Z}_Y$  is obtained by the equation  $\mathbf{d} = \mathbf{w}_\Phi^T \tilde{\Phi}(\mathbf{Z}_X) - \mathbf{w}_\Phi^T \tilde{\Phi}(\mathbf{Z}_Y)$ . For each test sample  $z_j$ , the projection is as follows:

$$\begin{aligned} \mathbf{w}_\Phi^T \tilde{\Phi}(z_j) &= \left( \sum_{i=1}^n \alpha_i \tilde{\Phi}(x_i) + \sum_{i=1}^n \beta_i \tilde{\Phi}(y_i) \right)^T \tilde{\Phi}(z_j) \\ &= \left( \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^T \begin{bmatrix} \tilde{K}_{XZ} \\ \tilde{K}_{YZ} \end{bmatrix} \right)_j. \end{aligned} \quad (36)$$

Therefore, the difference between the projected features is

$$\mathbf{w}_\Phi^T \tilde{\Phi}(\mathbf{Z}_X) - \mathbf{w}_\Phi^T \tilde{\Phi}(\mathbf{Z}_Y) = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^T \begin{bmatrix} \tilde{K}_{XZ_X} - \tilde{K}_{XZ_Y} \\ \tilde{K}_{YZ_X} - \tilde{K}_{YZ_Y} \end{bmatrix}. \quad (37)$$

In the above projection, the element of the centralized kernel matrix between the training and test samples is expressed as follows:

$$(\tilde{\mathbf{K}}_{XZ_X})_{ij} = \left( \Phi(x_i) - \frac{1}{n} \sum_{l=1}^n \Phi(x_l) \right)^T \left( \Phi(z_{Xj}) - \frac{1}{n} \sum_{l=1}^n \Phi(x_l) \right). \quad (38)$$

Therefore, the centralized kernel matrix  $\mathbf{K}_{XZ}$  can be written as

$$\tilde{\mathbf{K}}_{XZ_X} = \mathbf{K}_{XZ_X} - \frac{1}{n} \mathbf{1} \mathbf{K}_{XZ_X} - \frac{1}{n} \mathbf{K}_{XX} \mathbf{1} + \frac{1}{n^2} \mathbf{1} \mathbf{K}_{XX} \mathbf{1} \quad (39)$$

$$\tilde{\mathbf{K}}_{XZ_Y} = \mathbf{K}_{XZ_Y} - \frac{1}{n} \mathbf{1} \mathbf{K}_{XZ_Y} - \frac{1}{n} \mathbf{K}_{YY} \mathbf{1} + \frac{1}{n^2} \mathbf{1} \mathbf{K}_{YY} \mathbf{1} \quad (40)$$

$$\tilde{\mathbf{K}}_{YZ_X} = \mathbf{K}_{YZ_X} - \frac{1}{n} \mathbf{1} \mathbf{K}_{YZ_X} - \frac{1}{n} \mathbf{K}_{YX} \mathbf{1} + \frac{1}{n^2} \mathbf{1} \mathbf{K}_{YX} \mathbf{1} \quad (41)$$

$$\tilde{\mathbf{K}}_{YZ_Y} = \mathbf{K}_{YZ_Y} - \frac{1}{n} \mathbf{1} \mathbf{K}_{YZ_Y} - \frac{1}{n} \mathbf{K}_{YY} \mathbf{1} + \frac{1}{n^2} \mathbf{1} \mathbf{K}_{YY} \mathbf{1}. \quad (42)$$

Finally, KSFA can be achieved by all of the above equations from (18) to (42). It is worth noting that, in all the procedure, we only need the dot products of the nonlinear mapped features  $\Phi(x)$  and  $\Phi(y)$ . Therefore, we are able to use the kernel functions to compute their dot products without actually employing the nonlinear mapping  $\Phi$  [58].

There are several common kernel functions that can be used in KSFA, including the polynomial kernel and the Gaussian radial basis function (RBF) kernel. In this paper, we prefer the HIK, which is also used in the SVM classifier for accurate class probabilities. The details of the HIK can be found in Section II-A.

*3) Algorithm Implementation:* There are several important issues in the implementation of the KSFA algorithm.

Although SFA is an unsupervised learning algorithm, it can also be trained with selected unchanged training samples, to learn a more accurate invariant feature space [22]. Especially for scene change detection, *a priori* information is needed for the definition of semantic scene classes. In this paper, the training samples for scene classification are selected from the multitemporal data sets. Since it is difficult to directly extract unchanged scene pairs for all the classes, we generate simulated training scene pairs by fully connecting the multitemporal training scenes with the same semantic class in each data set.

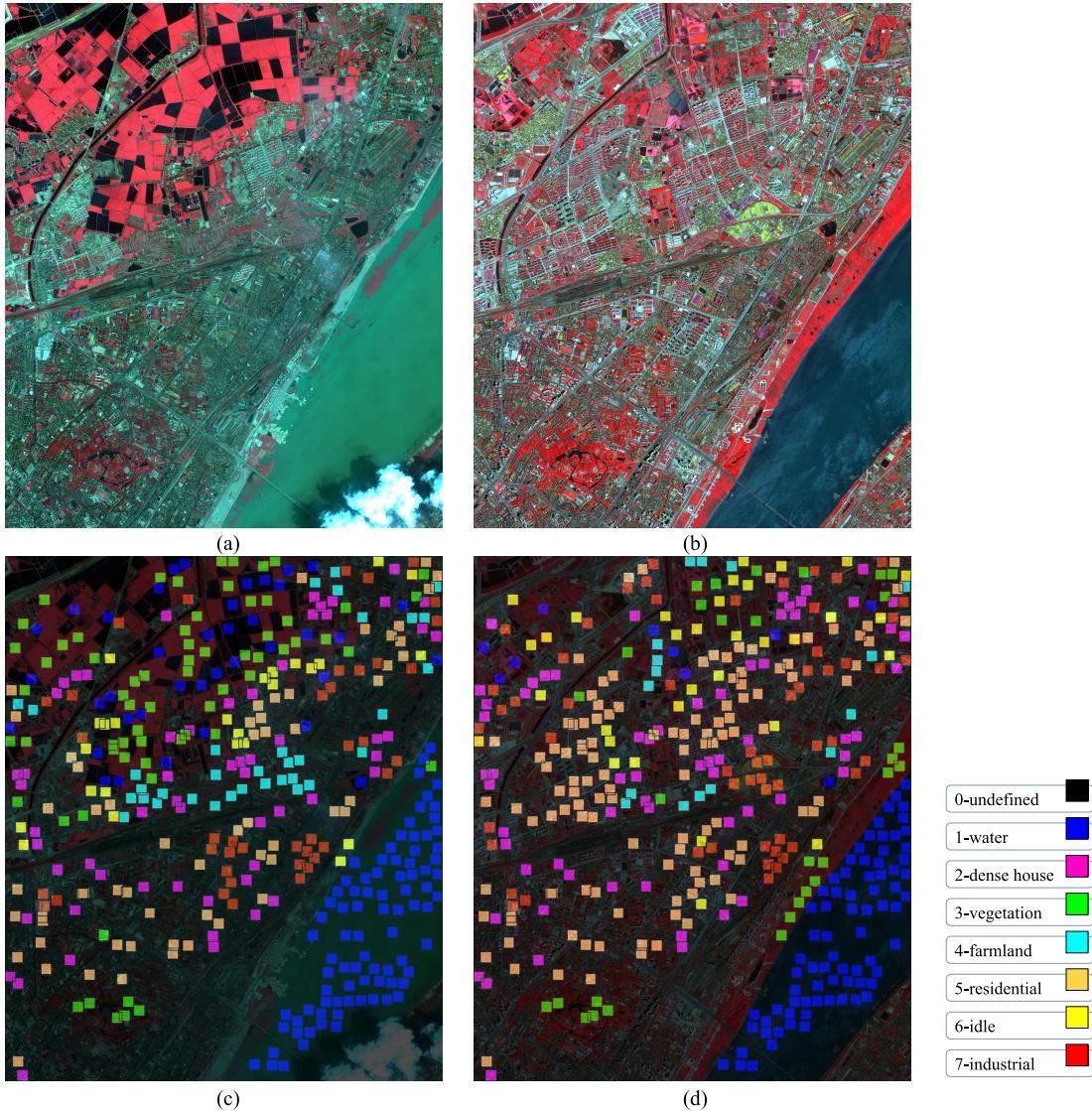


Fig. 2. Pseudocolor images of the Hanyang area of the city of Wuhan, acquired in (a) 2002 and (b) 2008. Reference maps for the test samples in (c) 2002 and (d) 2008, where the different colors represent different scene classes.

As mentioned above, the number of training samples for KSFA may be very large, so that matrices  $\mathbf{A}_K$  and  $\mathbf{B}_K$  in (27) will be singular in the eigenvalue problem. The solution is regularization [59], i.e., adding a small number to the diagonal of matrix  $\mathbf{B}_K$ . However, in KSFA, the most important component with the lowest index will have the smallest eigenvalue, approximately equaling 0, which will be greatly disturbed by the regularization, even though the regularization parameter  $\mu$  is very small. To overcome this problem, we exchange the places of  $\mathbf{A}_K$  and  $\mathbf{B}_K$  so that the most important component takes the highest eigenvalue  $\lambda$ . Theoretically, these two solutions are equivalent, and the output features will be sorted with the descending order of eigenvalues. The regularization is thus employed on matrix  $\mathbf{A}_K$ , i.e.,  $\mathbf{A}_K \leftarrow \mathbf{A}_K + \mu \mathbf{I}$ . To adaptively determine the regularization parameter for different kernel functions, we replace the fixed regularization parameter by the mean of the values in the diagonal of matrix  $\mathbf{A}_K$  multiplied by  $10^{-5}$ . In this way, KSFA can be successfully implemented to obtain the nonlinear invariant features.

After the differences between the transformed features are obtained, we need to calculate a distance to measure the change probabilities. Since most of the change information is concentrated in the first few bands, not all the transformed feature bands are needed for the measurement, due to the fact that the other bands may contain some noise and decrease the accuracy. As we know, the eigenvalue  $\lambda$  in our proposed method will indicate the contained information. In the proposed approach, the bands with eigenvalues lower than 0.5% of the total sum of all the eigenvalues are excluded from the calculation.

Two distances are used in the proposed method. The first is the Euclidean distance, which is calculated by the L2-norm. The other is the chi-squared distance, which is calculated as

$$\text{chi2} = \sum_{i=1}^m (\mathbf{d}_i^2 / \sigma_i^2) \in \chi^2(m) \quad (43)$$

where  $m$  is the number of feature bands. The chi-squared distance follows an  $\chi^2$  distribution with  $m$  degrees of freedom.

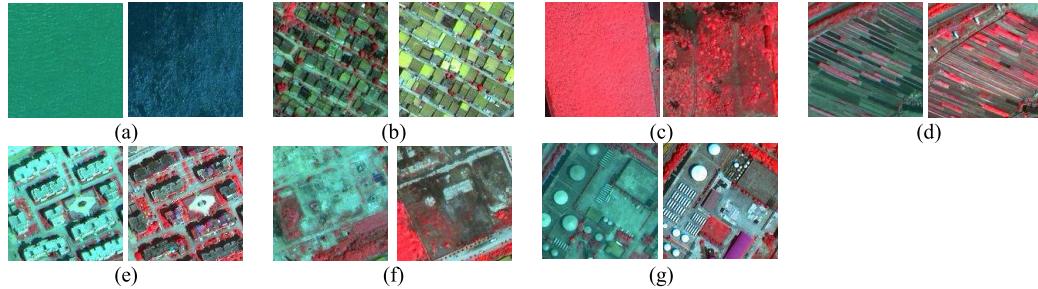


Fig. 3. Examples of the multitemporal scene pairs acquired in 2002 and 2008 belonging to (a) 1-water, (b) 2-dense houses, (c) 3-vegetation region, (d) 4-agricultural region, (e) 5-residential region, (f) 6-idle region, and (g) 7-industrial region.

TABLE I  
FORMULATION OF THE RULE-BASED METHOD

Rule 1	satisfied	unsatisfied
Rule 2	unsatisfied	Rule-based method 1
	satisfied	Rule-based method 3
		Rule-based method 2
		Rule-based method 4

In SFA,  $\sigma^2$  is the variance of the unchanged samples for the purpose of calculating the change probability according to the chi-squared distance [22]. Therefore, in order to exclude the changed samples and achieve accurate estimation of the variance, we use the test samples that are determined as unchanged according to the independent classification to estimate the variance.

### C. Postclassification Fusion

In this section, we introduce the two postclassification fusion methods, to integrate the change probability by KSFA with independent scene classification.

1) *Bayesian Theory*: In most classification-based change detection, the two classes of corresponding multitemporal samples are independent. However, if we assume that the identification of multitemporal scenes is correlated, the classification problem can be regarded as finding the couple of classes ( $w_i, v_j$ ) that provide the maximum *a posteriori* probability with the given observation features  $x$  and  $y$  [60]

$$\max_{w_i, v_j} \{P(w_i, v_j | x, y)\}. \quad (44)$$

According to Bayesian theory, the *a posteriori* probability can be rewritten as

$$\max_{w_i, v_j} \left\{ \frac{P(x, y | w_i, v_j) P(v_j | w_i) P(w_i)}{P(x, y)} \right\}. \quad (45)$$

We can assume that the probability of feature  $x$  depends only on the scene class  $w$ , as well as feature  $y$  and class  $v$  [60]. Furthermore,  $P(x, y)$  can be neglected since it has no influence on the semantic classes. We therefore obtain the following:

$$\max_{w_i, v_j} \{P(x | w_i) P(y | v_j) P(v_j | w_i) P(w_i)\}. \quad (46)$$

One can rewrite (46) as

$$\begin{aligned} & \max_{w_i, v_j} \left\{ \frac{P(w_i | x) P(x) P(v_j | y) P(y) P(v_j | w_i)}{P(v_j)} \right\} \\ & \Rightarrow \max_{w_i, v_j} \left\{ \frac{P(w_i | x) P(v_j | y) P(v_j | w_i)}{P(v_j)} \right\} \end{aligned} \quad (47)$$

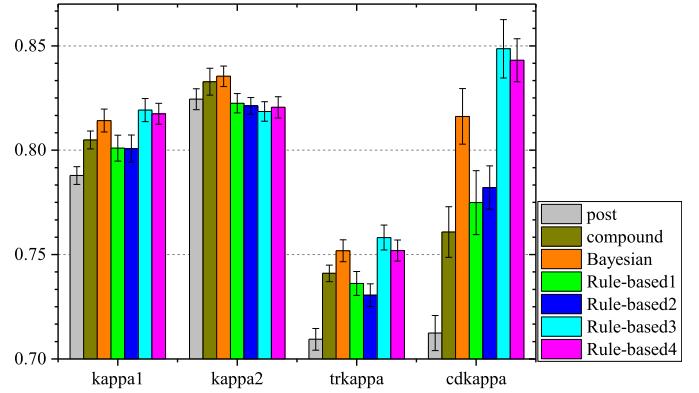


Fig. 4. Kappa coefficients of the different methods for scene classification, transition identification, and change detection.

where  $P(x)$  and  $P(y)$  are independent of  $w$  and  $v$ .

Another assumption is that *a priori* probability  $P(v_j)$  of each semantic class is equal; thus, it can also be neglected. Finally, we can obtain the decision rule according to Bayesian theory

$$\max_{w_i, v_j} \{P(w_i | x) P(v_j | y) P(v_j | w_i)\} \quad (48)$$

where  $P(w_i | x)$  and  $P(v_j | y)$  are the two conditional probabilities of the scene classes, which can be obtained by independent SVM classification.  $P(v_j | w_i)$  can be obtained with the change probability as

$$P(v_j | w_i) = \begin{cases} P_c & \text{if } w_i \neq v_j \\ 1 - P_c & \text{if } w_i = v_j. \end{cases} \quad (49)$$

The change probability  $P_c$  is assigned as the probability of being smaller than the chi-squared distance by KSFA in the  $\chi^2$  distribution [22] as

$$P_c = P\{\chi^2(m) < \text{chi2}\}. \quad (50)$$

The final semantic classes of the multitemporal scenes are determined as the coupled class combination with the maximum *a posteriori* probability (48).

### D. Rule-Based Method

In the traditional classification-based method, one of the main problems is that the unchanged samples with similar features are falsely detected as changes due to the variation of classification boundaries in the multitemporal data. So, the basic idea of the proposed rule-based method is to limit the false alarms in nonchange candidates. If the change probability of two multitemporal scenes is very small, they should belong

TABLE II

NUMBERS OF TRAINING AND TEST SAMPLES FOR EACH SCENE CLASS IN 2002 AND 2008 AND THE NUMBERS OF EACH “FROM-TO” TRANSITION TYPE

	Test (2008)							<b>SUM</b> (2002)	<b>Training</b> (2002)
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>		
<b>Test (2002)</b>	<b>1</b>	83	0	15	1	11	8	7	125
	<b>2</b>	0	59	0	0	2	0	0	61
	<b>3</b>	4	3	11	6	20	16	6	66
	<b>4</b>	0	3	2	8	7	2	12	34
	<b>5</b>	0	2	0	1	56	1	0	60
	<b>6</b>	0	1	2	0	17	3	2	25
	<b>7</b>	0	0	0	0	6	3	31	40
<b>SUM (2008)</b>		87	68	30	16	119	33	58	401
<b>Training (2008)</b>		13	14	11	9	20	11	14	

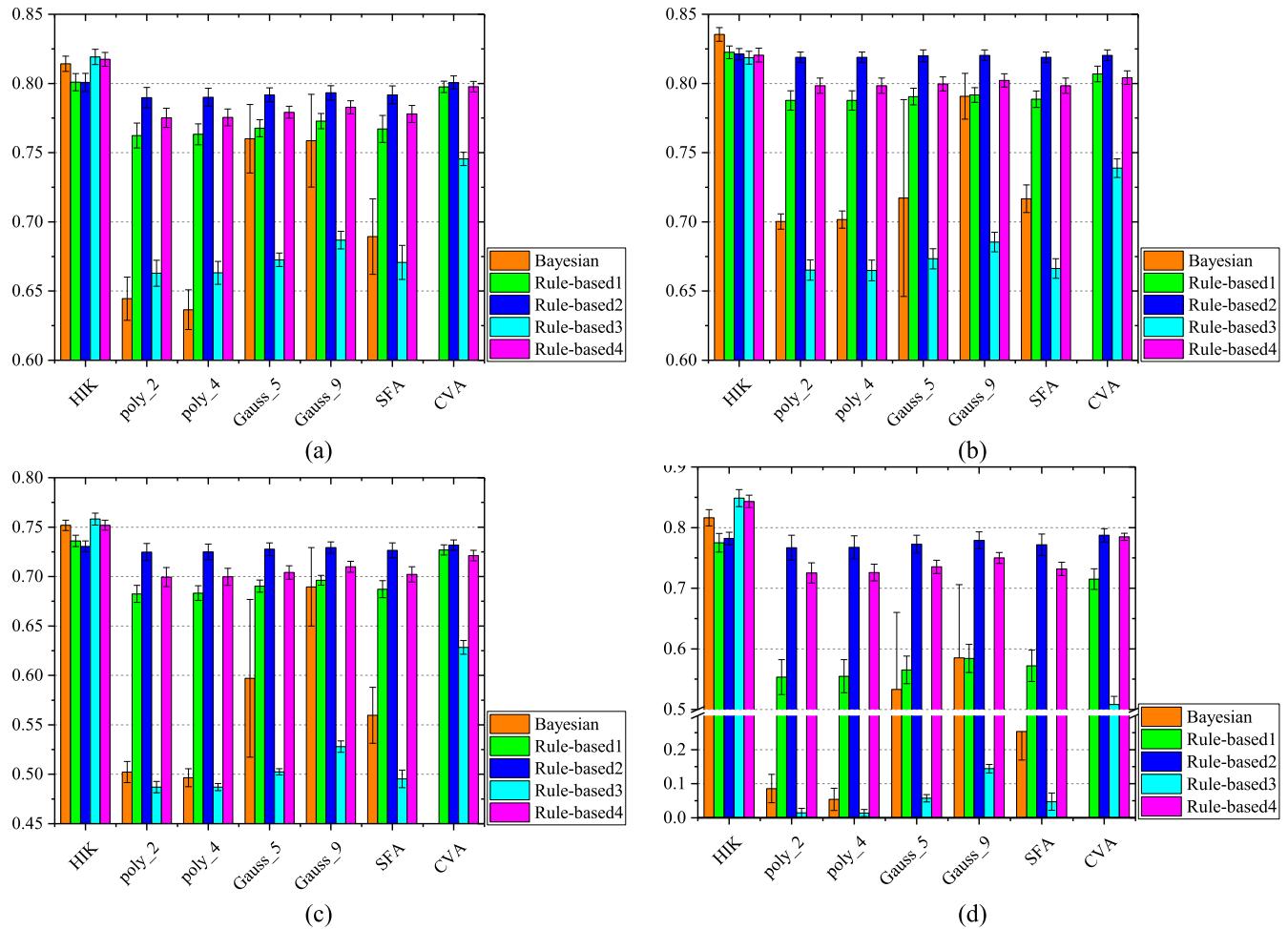


Fig. 5. Kappa coefficients after postclassification fusion with the proposed KSFA (different kernels) and the comparative methods. (a) Kappa for the first data set from 2002. (b) Kappa for the second data set from 2008. (c) Kappa for scene transition identification. (d) Kappa for scene change detection.

to the same semantic class. Therefore, we propose several rules for the nonchange candidates. If the rules are satisfied, the nonchange candidates with different classes are assigned as the class with the higher class probability between the two multitemporal scenes.

The nonchange candidates are determined according to the L2-norm of the KSFA features. Otsu’s thresholding method [61] is applied to automatically determine the threshold. The scene pair with an L2-norm that is smaller than the threshold is regarded as a nonchange candidate.

Rule 1 is that at least one of the maximum class probabilities of the corresponding multitemporal scenes is lower than 50%.

If they are both higher than 50%, it means that the scene classification is definite, and there is no need to make a correction.

Rule 2 is that the class  $w_{2nd}$  with the second maximum probability for  $\mathbf{x}$  should be the same as class  $v$  of  $\mathbf{y}$  when the class label of  $\mathbf{x}$  is changed to  $v$ . This rule assumes that although the classification of  $\mathbf{x}$  is not accurate, the correct class should also have a comparatively high class probability.

We finally construct four rule-based methods from the combination of the above two rules. Rule-based methods 1–4 are formed with rule 1 (satisfied/unsatisfied) and rule 2 (unsatisfied/satisfied), respectively, and are shown in Table I.

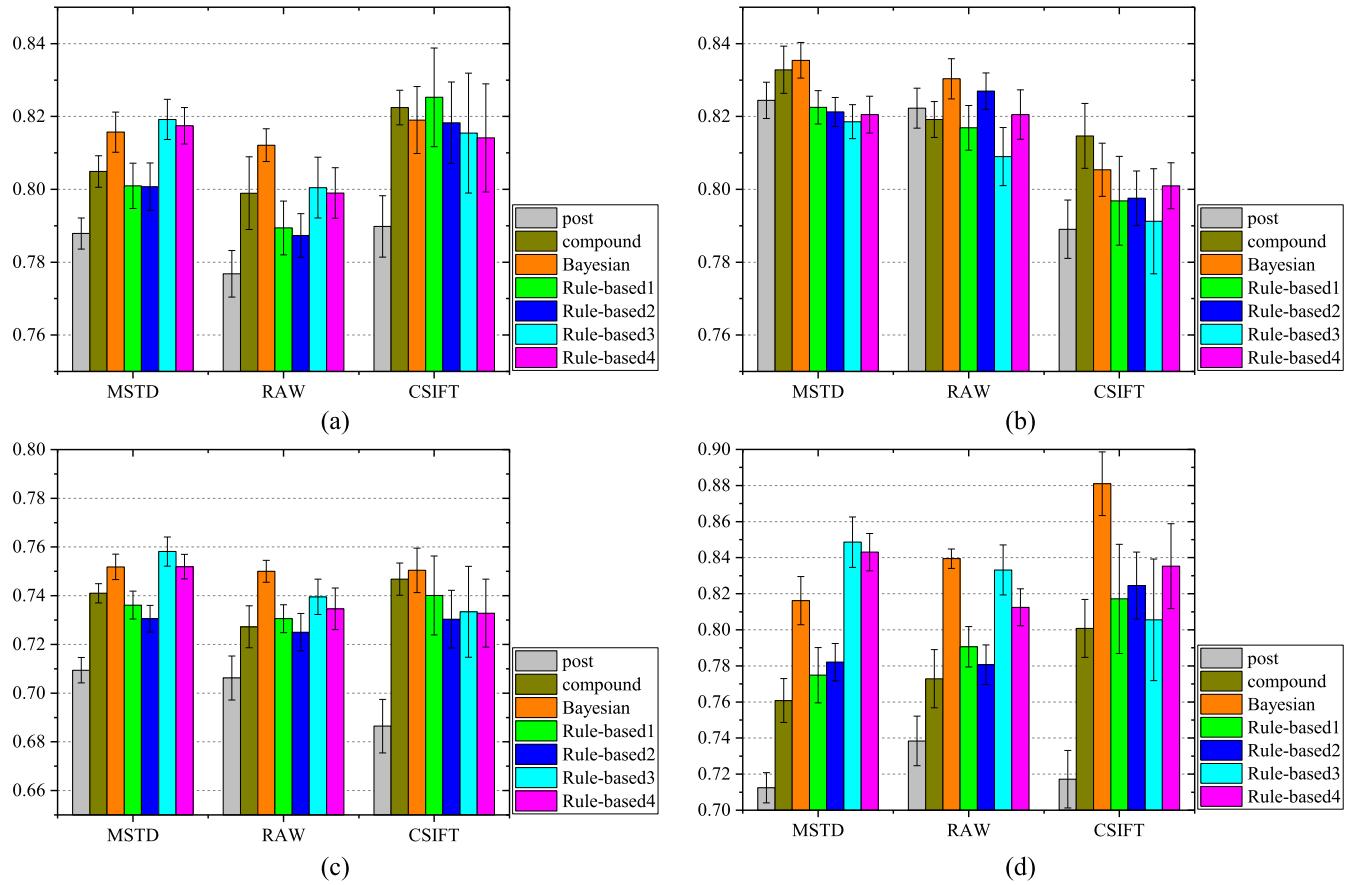


Fig. 6. Kappa coefficients for scene classification and change detection by the proposed method with different features. (a) Kappa for the first data set from 2002. (b) Kappa for the second data set from 2008. (c) Kappa for scene transition identification. (d) Kappa for scene change detection.

### III. EXPERIMENTS

Two multitemporal high-resolution remote sensing image scene data sets were used for the evaluation. They were both acquired by the IKONOS satellite sensor, and have a spatial resolution of 1 m after fusion of the pan and multispectral images by the Gram–Schmidt algorithm. Both data sets have four bands, which are red, green, blue, and near-infrared. Georeferencing and coregistration were performed to guarantee that the corresponding multitemporal image scenes covered the same region. The scene size was  $150 \times 150$  in the two data sets. In the experiments, the number of training patches for dictionary learning in the BOVW model was 100 000, the patch size was  $10 \times 10$ , the step for the patch overlap was 5, and the size of dictionary was 1000. Since the training patches were randomly selected for the dictionary learning, we repeated each experiment 10 times.

#### A. Hankou Data Set

The first data set features multitemporal high-resolution images covering the area of Hankou in the city of Wuhan, China, as shown in Fig. 2(a) and (b). The two large images were acquired on August 11, 2002, and May 11, 2008, respectively. A total of 411 scene pairs covering the same areas were extracted as test samples from the large images, as shown in Fig. 2(c) and (d). The images contain seven classes of water, dense houses, vegetation region, agricultural region, residential region, idle region, and industrial region, whose

TABLE III  
AVERAGE ACCURACY OF THE CHANGE PROBABILITY

		OA	kappa
post	post	0.8567	0.7124
	compound	0.8832	0.7608
KSFA(HIK)	L2-norm	0.9392	0.8710
	chi2	<b>0.9406</b>	<b>0.8741</b>
KSFA(poly_2)	L2-norm	0.7217	0.4437
	chi2	0.6929	0.3995
KSFA(poly_4)	L2-norm	0.6411	0.2394
	chi2	0.6377	0.2202
KSFA(RBF_5)	L2-norm	0.6270	0.1888
	chi2	0.6309	0.2073
KSFA(RBF_9)	L2-norm	0.6360	0.2206
	chi2	0.6436	0.2385
SFA	L2-norm	0.6642	0.3169
	chi2	0.6684	0.2694
CVA	L2-norm	0.6345	0.2623
	chi2	0.0000	-0.9065

representative examples are shown in Fig. 3. The numbers of test samples and training samples (which differed from the test samples) for each class and the numbers of each

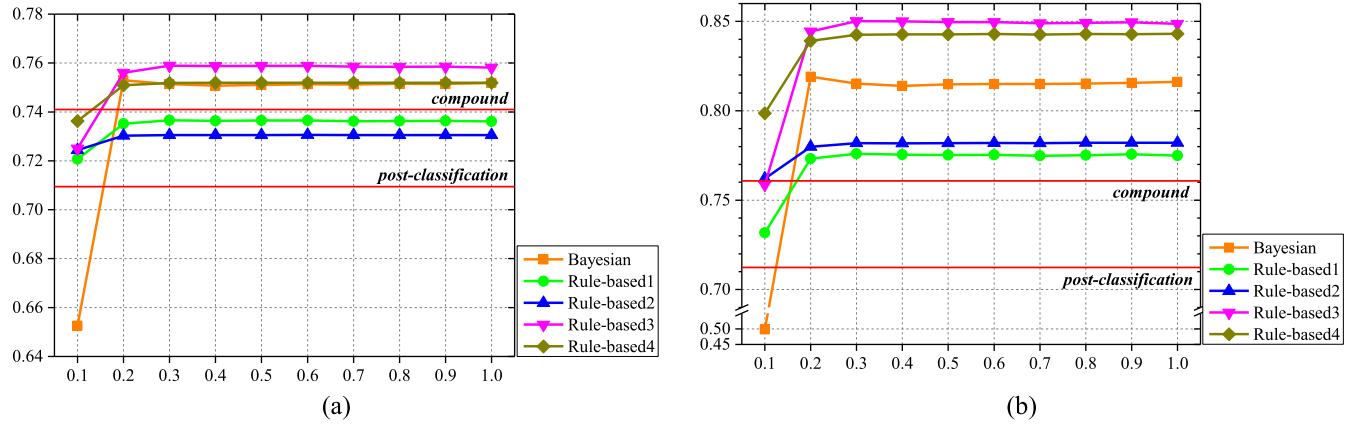


Fig. 7. Kappa coefficients of the proposed method with different percentages of simulated training samples for (a) scene transition identification and (b) scene change detection.

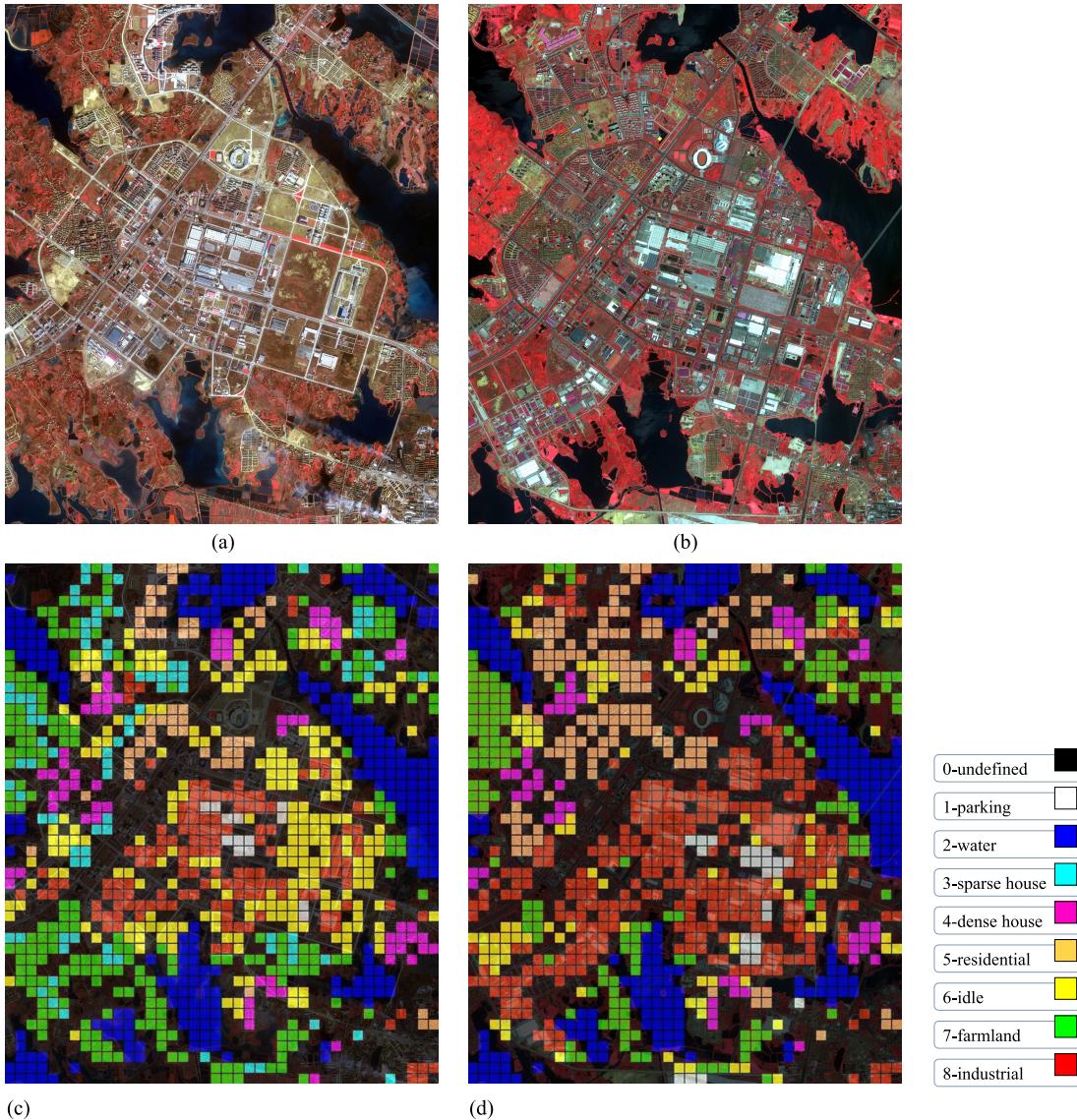


Fig. 8. Pseudocolor images of the Hankou area of the city of Wuhan, acquired in (a) 2002 and (b) 2009. Reference maps for the test samples in (c) 2002 and (d) 2009, where the different colors represent different scene classes.

“from-to” transition type are shown in Table II. Differing from the test samples, the representative training samples for each semantic class were selected from the large multitemporal

images, respectively. The training samples were then used for independent classification, and fully connected to simulate unchanged scene pairs for KSFA learning.

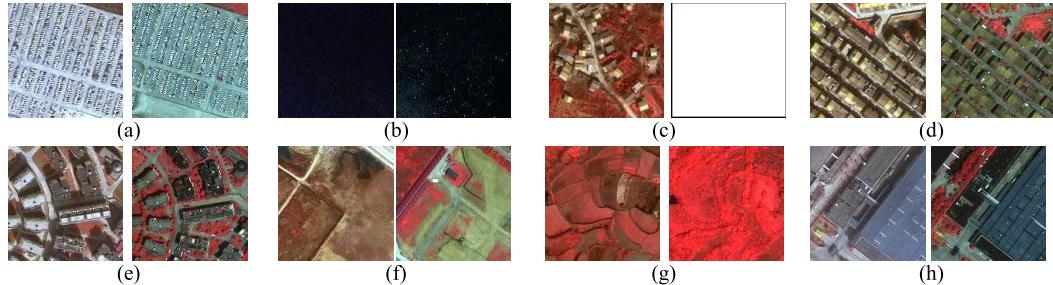


Fig. 9. Examples of multitemporal scene pairs acquired in 2002 and 2009 belonging to (a) 1-parking, (b) 2-water, (c) 3-sparse houses, (d) 4-dense houses, (e) 5-residential region, (f) 6-idle region, (g) 7-vegetation region, and (h) 8-industrial region.

TABLE IV  
NUMBERS OF TRAINING AND TEST SAMPLES FOR EACH SCENE CLASS IN 2002 AND 2009 AND THE NUMBERS OF EACH “FROM-TO” TRANSITION TYPE

		Test (2009)								<b>SUM</b> (2002)	<b>Training</b> (2002)
		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>		
<b>Test</b> (2002)	<b>1</b>	6	0	0	0	0	1	0	6	13	9
	<b>2</b>	0	240	0	0	1	3	2	0	246	27
	<b>3</b>	3	0	0	0	25	15	27	19	89	26
	<b>4</b>	0	0	0	63	1	4	0	0	68	22
	<b>5</b>	0	0	0	0	53	0	0	1	54	19
	<b>6</b>	11	0	0	0	44	40	21	109	225	28
	<b>7</b>	7	0	0	0	11	37	104	54	213	28
	<b>8</b>	0	0	0	0	4	5	2	131	142	31
<b>SUM (2009)</b>		27	240	0	63	139	105	156	320	1050	
<b>Training (2009)</b>		11	30	0	29	37	25	22	43		

Table III shows the change detection accuracy (change or nonchange) of the change probability obtained by the proposed KSFA method and the comparative methods. In each experimental run, all the methods use the same frequency features obtained by the BOVW model. “Post” indicates the postclassification change detection method directly comparing the classes of scene pairs, where “compound” in the second column means that all the multitemporal training samples were used in the training process for both the classifications. “HIK,” “poly,” and “RBF” represent the different kernel functions used in KSFA, while the following numbers are their kernel parameters. “SFA” indicates the original linear SFA method. “CVA” indicates change vector analysis, which calculates the difference between the original multitemporal feature vectors to detect changes. The change detection accuracy in Table III is the average value in ten runs of the maximum accuracy with different thresholds for the change probability obtained by KSFA, SFA, and CVA. Table III shows that the proposed KSFA with HIK can obtain better change detection performance than the other methods, while the chi-squared distance obtains the best result in both overall accuracy (OA) and kappa coefficient.

Fig. 4 shows the accuracy evaluation for the scene change detection, scene transition identification, and scene classification with the proposed KSFA and postclassification fusion, where the kernel function used in KSFA is HIK. In Fig. 4, “kappa1” and “kappa2” indicate the classification accuracies for the multitemporal scenes acquired in two years. “trkappa” regards the scene transition from class A to class B as one combination for accuracy evaluation, and “cdkappa” assesses the performance of detecting change or nonchange. In Fig. 4, “post” and “compound” are the two traditional postclas-

TABLE V  
AVERAGE ACCURACY OF THE CHANGE PROBABILITY

		<b>OA</b>	<b>kappa</b>
post	post	0.8867	<b>0.7685</b>
	compound	<b>0.8870</b>	0.7683
KSFA(HIK)	L2-norm	0.8529	0.6898
	chi2	0.8760	0.7374
KSFA(poly_2)	L2-norm	0.6324	0.2974
	chi2	0.6433	0.3143
KSFA(poly_4)	L2-norm	0.6028	0.1753
	chi2	0.6050	0.1909
KSFA(RBF_5)	L2-norm	0.5927	0.1501
	chi2	0.6248	0.1776
KSFA(RBF_9)	L2-norm	0.6065	0.2003
	chi2	0.6175	0.2014
SFA	L2-norm	0.6063	0.0910
	chi2	0.6089	0.1114
CVA	L2-norm	0.5786	0.0051
	chi2	0.1572	-0.6164

sification methods with independent and compound training samples, respectively. The other methods are combinations of the proposed KSFA with HIK and different postclassification fusion methods. Fig. 4 shows that the method with compound training samples obtains a higher accuracy than the straightforward postclassification method, which indicates that more training samples can provide a better estimation

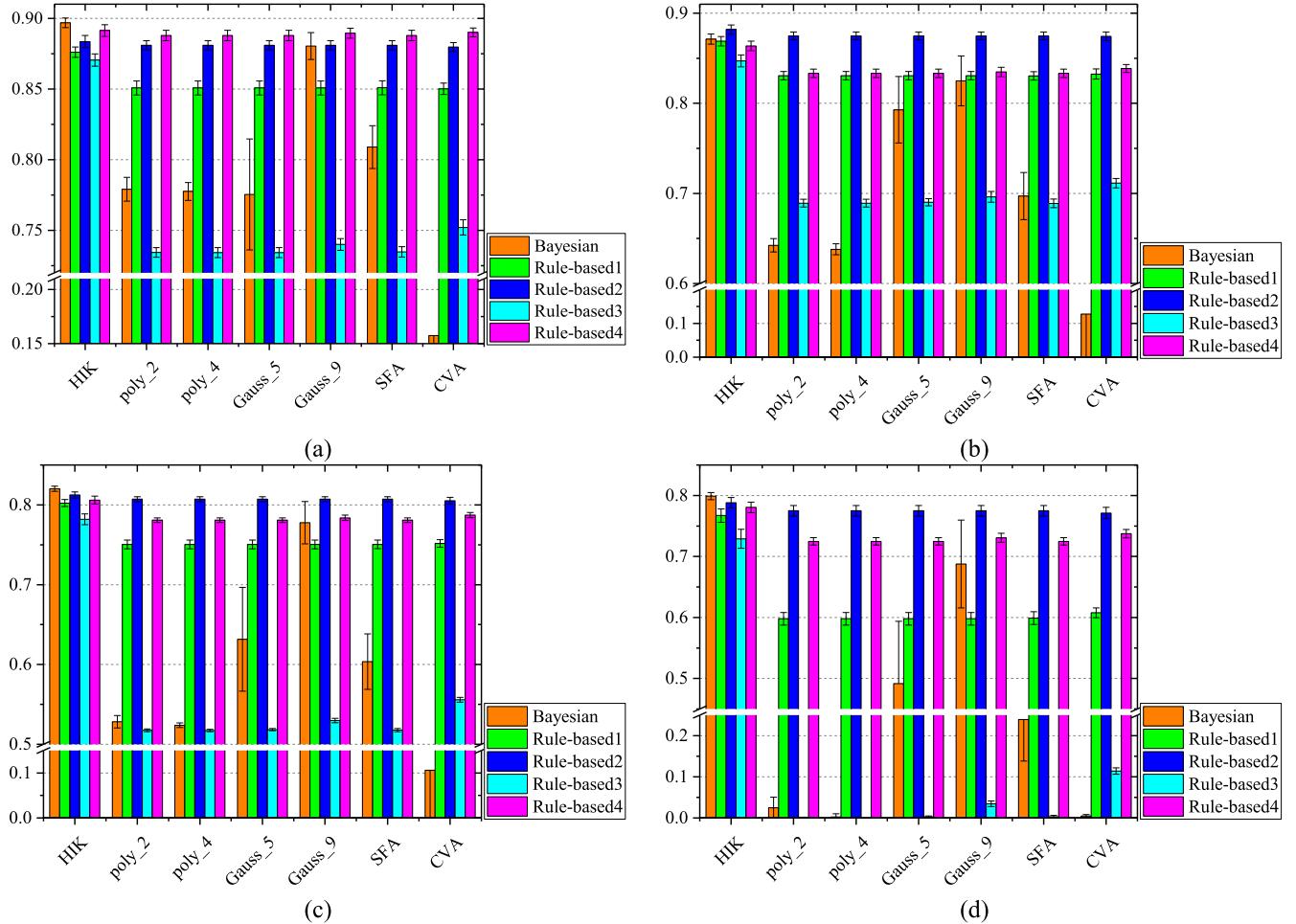


Fig. 10. Kappa coefficients after postclassification fusion with the proposed KSFA (different kernels) and the comparative methods. (a) Kappa for the first data set from 2002. (b) Kappa for the second data set from 2009. (c) Kappa for scene transition identification. (d) Kappa for scene change detection.

for the distribution of scene features. The proposed method with KSFA and postclassification fusion shows an obvious improvement in scene transition identification and change detection. It also improves the identification of the scene labels, since the unchanged scenes are more likely to belong to the same semantic class. Among the postclassification fusion methods, the Bayesian method and rule-based methods 3 and 4 all obtain satisfactory results. It is worth noting that the proposed method outperforms the method with compound training samples, which illustrates that the temporal correlation information interpreted by the proposed approach is superior to the simple approach of collecting multitemporal samples.

Fig. 5 shows the kappa coefficients after postclassification fusion with the proposed KSFA and the other comparative change detection methods. It shows that KSFA with HIK outperforms all the other methods with higher accuracies in all the postclassification fusion methods, especially for scene transition identification and change detection. Combined with Table III, it can be observed that when the change probabilities are not very accurate, such as KSFA with other kernels, SFA and CVA, the rule-based method can obtain comparatively high accuracies, as shown in Fig. 5. However, for these methods, the accuracies obtained by the postclassification

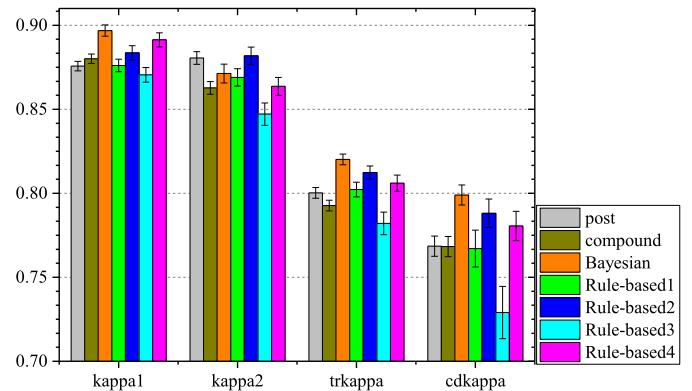


Fig. 11. Kappa coefficients of the different methods for scene classification, transition identification, and change detection.

fusion of Bayesian method are not satisfactory. Therefore, we can conclude that the rule-based method is more robust, even if the change detection information is not very accurate, whereas the Bayesian method is more dependent on the accuracy of the change probability. Among the rule-based methods, rule-based method 2 with both the rules obtains the best result in almost all the cases. Rule-based methods 1 and 4 with one rule are

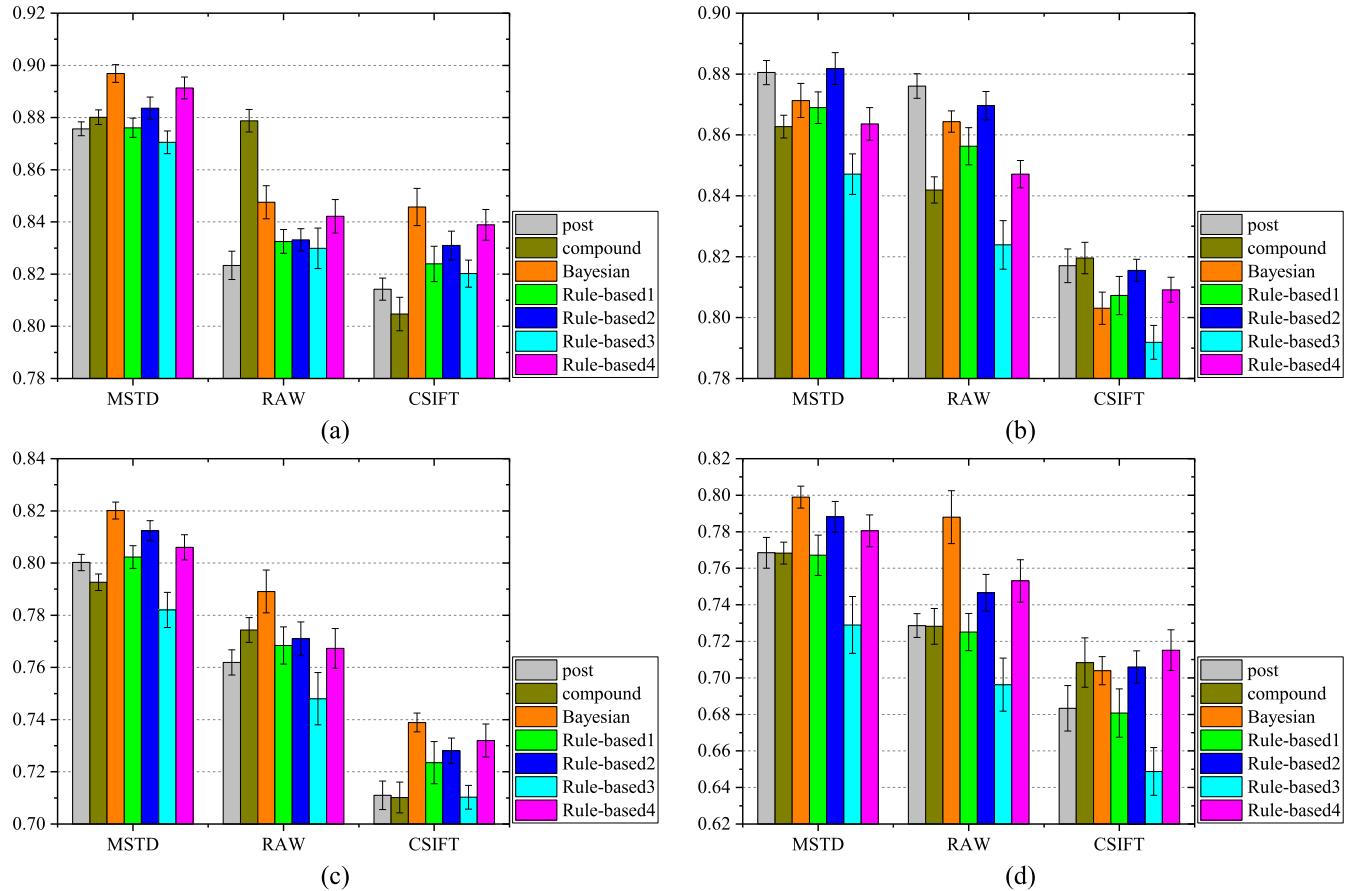


Fig. 12. Kappa coefficients for scene classification, transition identification, and change detection by the proposed method with different features. (a) Kappa for the first data set from 2002. (b) Kappa for the second data set from 2009. (c) Kappa for scene transition identification. (d) Kappa for scene change detection.

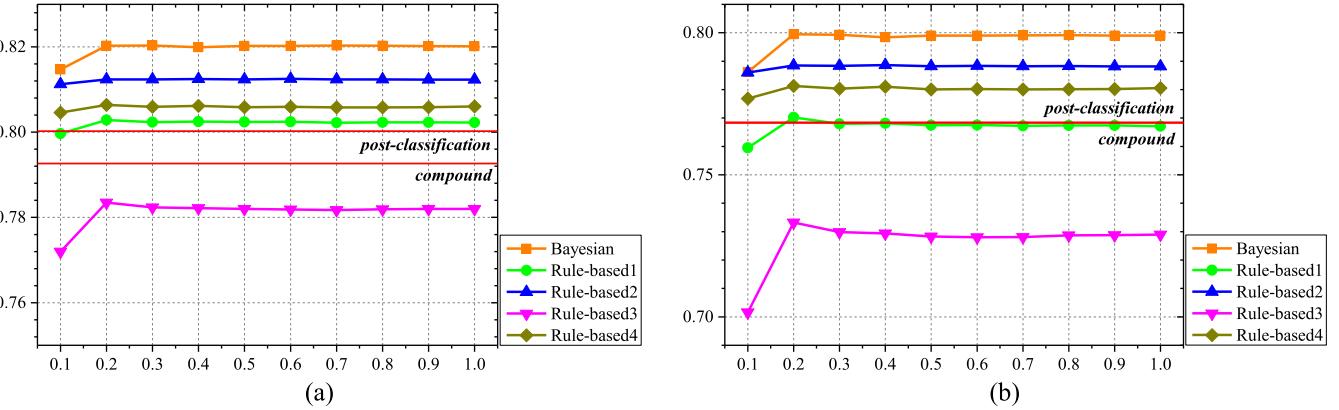


Fig. 13. Kappa coefficients of the proposed method with different percentages of simulated training samples for (a) scene transition identification and (b) scene change detection.

also robust, while rule-based method 3 without any rules does not obtain good accuracy. This demonstrates that the proposed rules are effective in the class correction of unchanged scenes.

Fig. 6 shows the kappa coefficients of different features with the proposed method. “MSTD” means the mean and standard deviation of the image patch, “RAW” is the raw pixel vector, and “CSIFT” indicates the dense SIFT descriptor. It can be observed that the proposed method with any of the features can obtain better performance than the simple comparison of scene classification, especially for the scene transition identification

and change detection. Among the three features, it is difficult to determine which one performs the best. However, when combined with the result of the next experiment, it can be concluded that “MSTD” is the most robust feature.

In the proposed method, the simulated training samples for KSFA come from the full connection of multitemporal training samples for scene classification. Since the computational cost is very high with a large number of training samples, we evaluated the performance of the proposed method with different percentages of simulated samples, as shown in Fig. 7. The

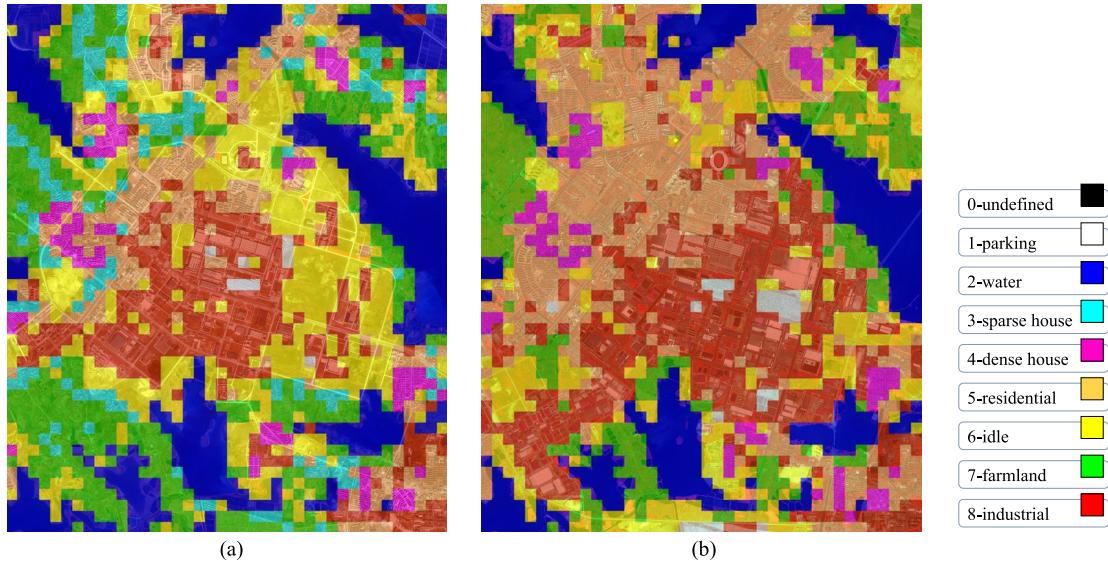


Fig. 14. Scene classification maps of the images acquired in (a) 2002 and (b) 2009.

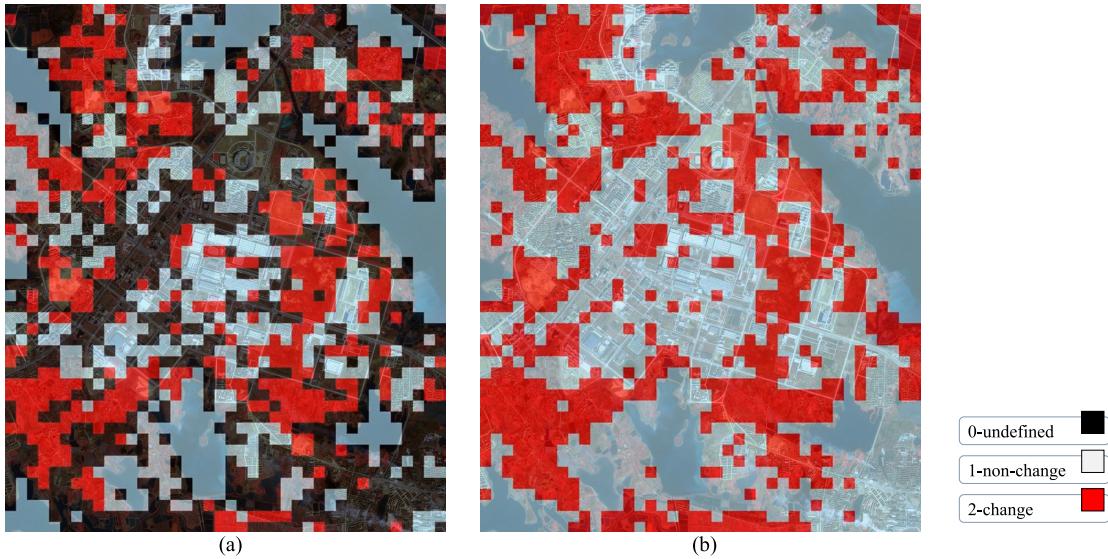


Fig. 15. (a) Reference map for change detection. (b) Scene change detection result obtained by the proposed method.

horizontal red lines indicate the kappa coefficients obtained by the postclassification method. It can be seen that with only 20% simulated training samples, the proposed method can obtain better results than the traditional method, and stable results can be obtained with only 30% simulated samples.

#### B. Hanyang Data Set

The second data set covers the Hanyang area of the city of Wuhan, China. The two large images were acquired on February 11, 2002, and June 24, 2009, respectively, as shown in Fig. 8(a) and (b). Fig. 8(c) and (d) shows a total of 1050 scene pairs covering the same areas as the test samples. In this data set, in order to evaluate the performance of the proposed scene change detection method in monitoring land-use changes in a given region, we selected the test scene pairs from the large image by a nonoverlapping grid. Examples of the test scenes for the semantic land-use classes are shown in Fig. 9. Since there are no scenes labeled as sparse houses

in 2009, the third class of scenes in Fig. 9 is filled by a blank image. The numbers of training and test samples for each scene class and the numbers of each “from-to” transition type are shown in Table IV. The training samples were all selected independently and differed from the test scene pairs.

Table V shows the accuracies of the change detection results obtained by the proposed method and the comparative methods. Differing from the Hankou experiment, the best accuracies are obtained by the traditional postclassification method. However, it is worth noting that the proposed KSFA with HIK obtains a clearly better performance than the other change detection methods, except for the traditional postclassification method. The chi-squared distance also obtains a higher accuracy than the L2-norm in most cases.

The kappa coefficients of the scene classification, transition identification, and change detection are shown in Fig. 11, where the kernel used in KSFA is HIK. Here, it can be seen that the proposed method with Bayesian fusion and

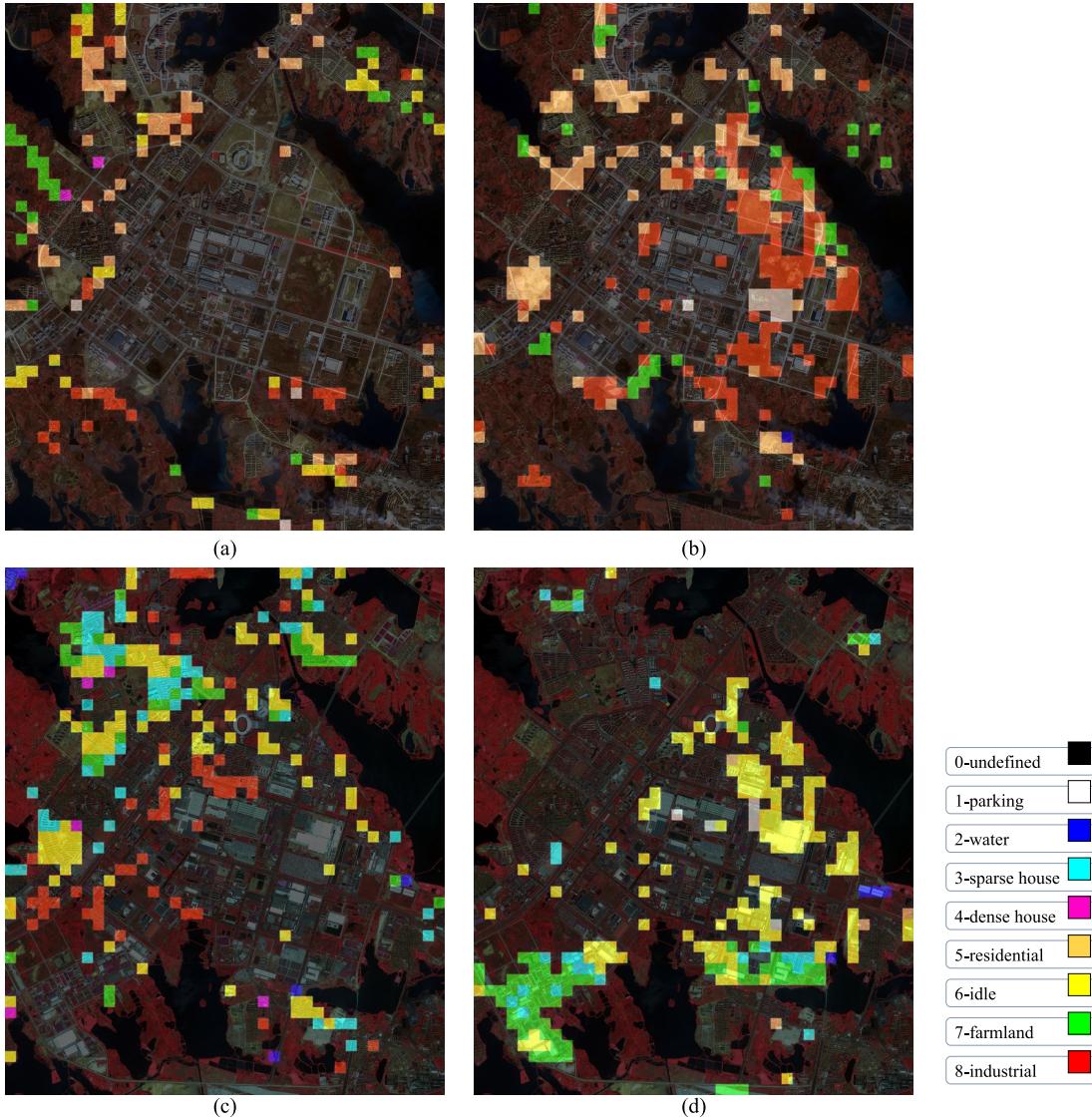


Fig. 16. “From-to” transition map of the multitemporal image scenes for the (a) change from sparse houses, (b) change from idle region, (c) change to residential region, and (d) change to industrial region. The different colors represent the different scene classes before and after scene change.

rule-based method 2 increases the accuracy of the transition identification and change detection, as well as the scene classification. When combined with Table V, we can conclude that even though the accuracy of the change information is not higher than that of direct comparison, the proposed method has the ability to improve the final performance. Among the postclassification fusion methods, the Bayesian method is very effective, and rule-based methods 2 and 4 outperform the other rule combinations. Furthermore, it can be observed that the method with compound training samples does not obtain a better result than postclassification, in most cases. This demonstrates that the approach of collecting multitemporal scene samples for independent classification does not always work. Comparatively, the proposed method of interpreting and utilizing temporal correlation information is more worthy of attention.

Fig. 10 shows the kappa coefficients after postclassification fusion with the different change detection methods. The same

conclusion can be made as the last experiment. The rule-based methods are more robust, even when the change probabilities are inaccurate, as shown in Table V. The Bayesian method can obtain the best performance if the change probability is very accurate, such as the proposed KSFA with HIK. The rules presented in this paper are effective in guaranteeing the performance of postclassification fusion, due to the fact that rule-based methods 1, 2, and 4 all obtain robust performances compared with rule-based method 3.

Fig. 12 shows the kappa coefficients with different features for scene classification, transition identification, and change detection. This shows that, in most cases, the proposed method can improve the performance for scene change detection and transition type identification, even with different features. The “MSTD” feature obtains a higher accuracy than the other features. When combined with Fig. 6, we can conclude that “MSTD” is more robust in real applications. It can also be observed that a better representative feature for scene

TABLE VI  
STATISTICS OF THE SCENE CLASSES IN 2002 AND 2009

Class	2002	2009	+/-
1-parking	9	25	+16
2-water	384	331	-53
3-sparse houses	141	0	-141
4-dense houses	78	76	-2
5-residential	221	466	+245
6-idle	460	318	-142
7-farmland	328	231	-97
8-industrial	299	473	+174

classification leads to a better performance in scene change detection.

Fig. 13 shows the kappa coefficients with different percentages of simulated training samples for scene transition identification and change detection. As in Fig. 7, the proposed method obtains a satisfactory and stable result with about 20%–30% training samples. Therefore, in a practical application, we could use 30% simulated samples for a balance between accuracy and efficiency.

Fig. 14 shows the scene classification maps after postclassification fusion with Bayesian theory. The kernel used in KSFA is HIK. The OA and kappa coefficients are 91.62%/0.8960 and 88.67%/0.8627, respectively. Compared with the independent scene classifications with kappa coefficients of 0.8785 and 0.8799, the multitemporal scene maps after postclassification fusion show an increase and a decrease in accuracy, respectively. However, the result of the proposed method shows an obvious improvement in scene change detection accuracy from 0.7759 to 0.8116, and scene transition identification from 0.8045 to 0.8205. By interpreting the multitemporal scene maps, it is easy to observe the pattern of urban expansion. Most of the sparse houses regions and idle regions have been developed into residential regions and industrial regions. From the statistics of the scene classes in 2002 and 2009, as shown in Table VI, it can be observed that the numbers of water regions, sparse houses regions, and idle regions reduced, while the residential regions and industrial regions expanded. This is a very typical pattern of urban expansion.

Fig. 15 shows the reference map for change detection and the result obtained by the proposed method. In Fig. 15, the red regions indicate the changed scene pairs, white regions indicate the unchanged scene pairs, and the undefined scene pairs are shown in black. The scene change detection result is obtained with the HIK and the “MSTD” feature. The OA and kappa for scene change detection are 91.14% and 0.8116%, respectively. By visual interpretation with the images in Fig. 8 and the quantitative assessment, we can conclude that the scene change detection result in Fig. 15(b) very accurately indicates the regions with land-use changes. The proposed method could therefore be applied in the monitoring of urban development, where the land-use changes are more important and meaningful than the changes of simple landscapes.

Fig. 16 shows the “from-to” transition maps of the multitemporal image scenes obtained by the proposed method. The different colors represent the different scene classes before and after scene change. The OA and kappa for the scene “from-to” transition type are 83.90% and 0.8205. The “from-to” transition maps shown in Fig. 16 highlight the most important land-use variations for urban development monitoring. Fig. 16(a) shows that the small villages disappeared due to the expansion of the city. Fig. 16(b) shows that most of the idle regions in 2002 were developed into residential regions and industrial regions by 2009. In Fig. 16(c) and (d), it can be observed that the residential regions appeared in the left-top, and the industrial regions expanded to the right-bottom. These two land-use regions mostly changed from sparse houses regions, idle regions, and vegetation regions. From the experimental results and the corresponding analysis of the land-use variation, it can be seen that scene change detection has great potential in the monitoring and analysis of urban development.

#### IV. CONCLUSION

Scene change detection has significant potential in the interpretation of land-use transition for urban development monitoring, but it has not been previously studied in-depth. In this paper, we have proposed a novel scene change detection method with KSFA and two postclassification fusion methods (Bayesian theory and a rule-based method). KSFA with an HIK is used to extract temporally invariant features in the nonlinear high-dimensional space for multitemporal image scenes. The two postclassification fusion methods are then used to integrate the change probability interpreted by KSFA with independent scene classification, to achieve better performance in scene change detection, scene transition identification, and scene classification.

Two experiments with high-resolution remote sensing scenes were undertaken for quantitative evaluation. The results indicate that the proposed KSFA with HIK has the ability to provide accurate change probability information. The proposed method with KSFA and postclassification fusion can clearly improve the final performance of scene change detection, scene transition identification, and scene classification. Compared with collecting multitemporal scene samples for both the classifications, the proposed method is a better way to interpret and utilize temporal correlation information.

Among the postclassification fusion methods, the Bayesian method is more dependent on the quality of the change probability, and the rule-based method is more robust to different change detection methods. The experimental results for different combinations of presented rules demonstrate that the rules are effective in correcting the scene classes of unchanged candidates.

The experiments with different scene features indicate that the mean and standard deviation of the image patch are robust features for scene change detection and classification. Finally, for a balance between accuracy and efficiency, around 30% simulated training samples is enough for a stable and satisfactory performance.

In our future work, we will focus on two aspects: 1) how to obtain more accurate change probability for the Bayesian post-classification fusion and 2) how to encode the multitemporal image scenes more effectively for better feature representation.

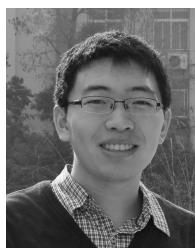
### ACKNOWLEDGMENT

The authors thank the editor and the reviewers for their instructive comments that helped to improve this paper.

### REFERENCES

- [1] T. Deselaers, L. Pimenidis, and H. Ney, "Bag-of-visual-words models for adult image classification and filtering," in *Proc. 19th Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2008, pp. 1–4.
- [2] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognit.*, vol. 37, no. 9, pp. 1757–1771, Sep. 2004.
- [3] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1794–1801.
- [4] X. Zhang and S. Du, "A linear Dirichlet mixture model for decomposing scenes: Application to analyzing urban functional zonings," *Remote Sens. Environ.*, vol. 169, pp. 37–49, Nov. 2015.
- [5] A. M. Cheriyadat, "Unsupervised feature learning for aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 439–451, Jan. 2014.
- [6] X. Zheng, X. Sun, K. Fu, and H. Wang, "Automatic annotation of satellite images via multifeature joint sparse coding with spatial relation constraint," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 652–656, Apr. 2013.
- [7] L. Wang, L. Hongliang, and L. Guanghui, "Automatic annotation of multispectral satellite images using author-topic model," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 634–638, Apr. 2012.
- [8] K. Qi, H. Wu, C. Shen, and J. Gong, "Land-use scene classification in high-resolution remote sensing images using improved correlatons," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2403–2407, Dec. 2015.
- [9] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," presented at the 18th Int. Conf. Adv. Geogr. Inf. Syst. (SIGSPATIAL), San Jose, CA, USA, 2010.
- [10] G. Cheng, L. Guo, T. Zhao, J. Han, H. Li, and J. Fang, "Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA," *Int. J. Remote Sens.*, vol. 34, no. 1, pp. 45–59, 2013.
- [11] B. Zhao, Y. Zhong, and L. Zhang, "Scene classification via latent Dirichlet allocation using a hybrid generative/discriminative strategy for high spatial resolution remote sensing imagery," *Remote Sens. Lett.*, vol. 4, no. 12, pp. 1204–1213, 2013.
- [12] F. P. S. Luus, B. P. Salmon, F. van den Bergh, and B. T. J. Maharaj, "Multiview deep learning for land-use classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2448–2452, Dec. 2015.
- [13] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior-based salient object detection via deep reconstruction residual," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1309–1321, Aug. 2015.
- [14] X. Huang, H. Liu, and L. Zhang, "Spatiotemporal detection and analysis of urban villages in mega city regions of China using high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3639–3657, Jul. 2015.
- [15] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogram. Remote Sens.*, vol. 80, pp. 91–106, Jun. 2013.
- [16] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, pp. 989–1003, Jul. 1989.
- [17] X. Lu, Y. Yuan, and X. Zheng, "Joint dictionary learning for multispectral change detection," *IEEE Trans. Cybern.*, to be published.
- [18] Y. Yuan, H. Lv, and X. Lu, "Semi-supervised change detection method for multi-temporal hyperspectral images," *Neurocomputing*, vol. 148, pp. 363–375, Jan. 2015.
- [19] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Review articledigital change detection methods in ecosystem monitoring: A review," *Int. J. Remote Sens.*, vol. 25, pp. 1565–1596, May 2004.
- [20] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2007.
- [21] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 463–478, Feb. 2007.
- [22] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2858–2874, May 2014.
- [23] A. Ertürk and A. Plaza, "Informative change detection by unmixing for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1252–1256, Jun. 2015.
- [24] X. Lu, H. Wu, and Y. Yuan, "Double constrained NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2746–2758, May 2014.
- [25] G. Xian, C. Homer, and J. Fry, "Updating the 2001 national land cover database land cover classification to 2006 by using landsat imagery change detection methods," *Remote Sens. Environ.*, vol. 113, pp. 1133–1147, Jun. 2009.
- [26] B. Demir, F. Bovolo, and L. Bruzzone, "Detection of land-cover transitions in multitemporal remote sensing images with active-learning-based compound classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1930–1941, May 2012.
- [27] C. Wu, L. Zhang, and L. Zhang, "A scene change detection framework for multi-temporal very high resolution remote sensing images," *Signal Process.*, vol. 124, pp. 184–197, Jul. 2015.
- [28] W. Yu, W. Zhou, Y. Qian, and J. Yan, "A new approach for land cover classification and change analysis: Integrating backdating and an object-based method," *Remote Sens. Environ.*, vol. 177, pp. 37–47, May 2016.
- [29] G. Xian and C. Homer, "Updating the 2001 national land cover database impervious surface products to 2006 using landsat imagery change detection methods," *Remote Sens. Environ.*, vol. 114, pp. 1676–1686, Aug. 2010.
- [30] N. Wilbert, H. Sprekeler, M. Franzius, P. Berkes, and L. Wiskott, "Slow feature analysis," *Scholarpedia*, vol. 6, p. 5282, Apr. 2011.
- [31] L. Wiskott and T. J. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural Comput.*, vol. 14, no. 4, pp. 715–770, Apr. 2002.
- [32] M. Franzius, N. Wilbert, and L. Wiskott, "Invariant object recognition and pose estimation with slow feature analysis," *Neural Comput.*, vol. 23, no. 9, pp. 2289–2323, Sep. 2011.
- [33] C. Theriault, N. Thome, and M. Cord, "Dynamic scene classification: Learning motion descriptors with slow features analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Oct. 2013, pp. 2603–2610.
- [34] Z. Zhang and D. Tao, "Slow feature analysis for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 436–450, Mar. 2012.
- [35] S. Lin, J. Kui, C. Tsung-Han, F. Yuqiang, W. Gang, and Y. Shuicheng, "DL-SFA: Deeply-learned slow feature analysis for action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2625–2632.
- [36] C. Wu, L. Zhang, and B. Du, "Hyperspectral anomaly change detection with slow feature analysis," *Neurocomputing*, vol. 151, pp. 175–187, Mar. 2015.
- [37] X. Dong, J. Shen, L. Shao, and L. V. Gool, "Sub-Markov random walk for image segmentation," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 516–527, Feb. 2016.
- [38] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 712–727, Apr. 2008.
- [39] L. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Compression of hyperspectral remote sensing images by tensor approach," *Neurocomputing*, vol. 147, pp. 358–363, Jan. 2015.
- [40] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern Recognit.*, vol. 48, pp. 3102–3112, Apr. 2015.
- [41] B. Fernando, E. Fromont, D. Muselet, and M. Sebban, "Discriminative feature fusion for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Oct. 2012, pp. 3434–3441.

- [42] A. Bosch *et al.*, "Scene classification via pLSA," in *Computer Vision—ECCV*, vol. 3954, A. Leonardis, Ed. Berlin, Germany: Springer, 2006, pp. 517–530.
- [43] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1596, Sep. 2010.
- [44] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Apr. 2006, pp. 2169–2178.
- [45] A. Coates and A. Y. Ng, "Learning feature representations with K-means," in *Neural Networks: Tricks Trade*. Berlin, Germany: Springer, 2012, pp. 561–580.
- [46] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.
- [47] A. Vedaldi and A. Zisserman, "Efficient additive kernels via explicit feature maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 480–492, Mar. 2012.
- [48] A. Barla, F. Odone, and A. Verri, "Histogram intersection kernel for image classification," in *Proc. Int. Conf. Image Process. (ICIP)*, vol. 2, Sep. 2003, pp. III-513–III-516.
- [49] B. Zhao, Y. Zhong, L. Zhang, and B. Huang, "The Fisher kernel coding framework for high spatial resolution scene classification," *Remote Sens.*, vol. 8, no. 2, p. 157, 2016.
- [50] B. Zhao, Y. Zhong, and L. Zhang, "A spectral–structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 73–85, Jun. 2016.
- [51] P. Berkes and L. Wiskott, "Slow feature analysis yields a rich repertoire of complex cell properties," *J. Vis.*, vol. 5, pp. 579–601, Sep. 2005.
- [52] W. Böhmer *et al.*, "Regularized sparse kernel slow feature analysis," in *Machine Learning and Knowledge Discovery in Databases*, vol. 6911, D. Gunopulos, Ed. Berlin, Germany: Springer, 2011, pp. 235–248.
- [53] S. Liwicki, S. P. Zafeiriou, and M. Pantic, "Online kernel slow feature analysis for temporal video segmentation and tracking," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 2955–2970, Oct. 2015.
- [54] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
- [55] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.*, vol. 12, no. 10, pp. 2385–2404, 2000.
- [56] J. Yang, Z. Jin, J.-Y. Yang, D. Zhang, and A. F. Frangi, "Essence of kernel Fisher discriminant: KPCA plus LDA," *Pattern Recognit.*, vol. 37, pp. 2097–2100, Oct. 2004.
- [57] L. Zhang, C. Wu, and B. Du, "Automatic radiometric normalization for multitemporal remote sensing imagery with iterative slow feature analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6141–6155, Oct. 2014.
- [58] B. Schölkopf *et al.*, "Kernel principal component analysis," in *Artificial Neural Networks—ICANN*, vol. 1327, W. Gerstner, Ed. Berlin, Germany: Springer, 1997, pp. 583–588.
- [59] J. Yang, A. F. Frangi, J.-Y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: A complete kernel Fisher discriminant framework for feature extraction and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 230–244, Feb. 2005.
- [60] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 858–867, Jul. 1997.
- [61] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, nos. 285–296, pp. 23–27, 1975.



**Chen Wu** (M'16) received the B.S. degree in surveying and mapping engineering from Southeast University, Nanjing, China, in 2010, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China, in 2015.

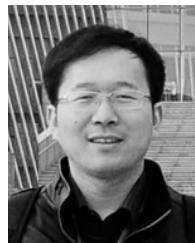
He is currently a Lecturer with the International School of Software, Wuhan University. His research interests include multitemporal remote sensing image change detection and analysis in multispectral and hyperspectral images.



**Liangpei Zhang** (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently the Head of the Remote Sensing Division with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He is also a Chang-Jiang Scholar Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China state key basic research project (2011–2016), appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has authored more than 500 research papers and five books. He holds 15 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the Institution of Engineering and Technology, Executive Member (board of governor) of the China National Committee of International Geosphere-Biosphere Program, and Executive Member of the China Society of Image and Graphics. He serves as the Co-Chair of the series SPIE Conferences on Multispectral Image Processing and Pattern Recognition, Conference on Asia Remote Sensing, and many other conferences. He is the Founding Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter. He was the General Chair for the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing and the Guest Editor of the JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING (JSTARS). He was a recipient of the 2010 Best Paper Boeing Award and the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing. He was also a recipient of the Best Reviewer Awards from the IEEE GRSS for his service to the IEEE JSTARS in 2012 and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2014. His research teams won the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium student paper contest in recent years. He is the Editor of several conference proceedings, issues, and geoinformatics symposiums. He serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, *International Journal of Image and Graphics*, *International Journal of Digital Multimedia Broadcasting*, *Journal of Geo-Spatial Information Science*, *Journal of Remote Sensing*, and the *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, and the Guest Editor of the *Journal of Applied Remote Sensing* and *Journal of Sensors*.



**Bo Du** (M'10–SM'15) received the B.S. and Ph.D. degrees in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2005 and 2010, respectively.

He is currently a Professor with the School of Computer, Wuhan University. He has authored more than 40 research papers in the *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* (TGRS), *IEEE TRANSACTIONS ON IMAGE PROCESSING* (TIP), *IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING* (JSTARS), and *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS* (GRSL). Five of them are ESI hot papers or highly cited papers. His research interests include pattern recognition, hyperspectral image processing, and signal processing.

Dr. Du was a recipient of the Best Reviewer Awards from the IEEE Geoscience and Remote Sensing Society (GRSS) for his service to the IEEE JSTARS in 2011 and the ACM Rising Star Awards for his academic progress in 2015. He was the Session Chair for both International Geoscience And Remote Sensing Symposium in 2016 and the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing. He also serves as a Reviewer for 20 Science Citation Index magazines including IEEE TGRS, TIP, JSTARS, and GRSL.