IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 30, 2018, accepted November 17, 2018, date of publication November 29, 2018,
date of current version December 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2883254

Multi-Scale Feature Based Land Cover Change Detection in Mountainous Terrain Using Multi-Temporal and Multi-Sensor Remote Sensing Images

FEI SONG[1,2,3], ZHUOQIAN YANG[4], XUEYAN GAO[1,2,3], TINGTING DAN[1,2,3], YANG YANG[1,2,3], WANJING ZHAO[1,2,3], AND RUI YU[1,2,3]

[1]School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China
[2]Engineering Research Center of GIS Technology in Western China, Ministry of Education of the People's Republic of China, Kunming 650500, China
[3]Laboratory of Pattern Recognition and Artificial Intelligence, School of Information Science, Yunnan Normal University, Kunming 650500, China
[4]College of Software, Beihang University, Beijing 100083, China

Corresponding author: Yang Yang (yyang_ynu@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 41661080 and in part by the Scientific Research Foundation of Yunnan Provincial Department of Education under Grant 2018Y037.

ABSTRACT Land use and land cover (LULC) change is frequent in mountainous terrain of southern China. Although remote sensing technology has become an important tool for gathering and monitoring LULC dynamics, image pairs can occur scale changes, noises, geometrical distortions, and illuminated variations if these are acquired from different types of sensors (e.g., satellites). Meanwhile, how to design an efficient land cover change detection algorithm that ensures a high detection rate remains a critical and challenging step. To address these problems, we propose a robust multi-temporal change detection framework for land cover change in mountainous terrain which contains the following contributions. i) To transform multi-temporal remote sensing image pairs acquired by different type of sensors into the same coordinate system by image registration, a multi-scale feature description is generated using layers formed via a pretrained VGG network. ii) A gradually increasing selection of inliers is defined for improving the robustness of feature points registration, and L_2-minimizing estimate (L_2E)-based energy optimization is formulated to calculate a reasonable position in a reproducing kernel Hilbert space. iii) Fuzzy C-Means classifier is adopted to generate a similarity matrix between image pair of geometric correction, and a robust and contractive change map is built through feature similarity analysis. Extensive experiments on multi-temporal image pairs taken by different type of satellites (e.g., Chinese GF and Landsat) or small unmanned aerial vehicles are conducted. Experimental results show that our method provides better performances in most cases after comparing with the five state-of-the-art image registration methods and the four state-of-the-art change detection methods.

INDEX TERMS LULC change, multi-scale feature description, inliers, L_2E, fuzzy C-Means classifier.

I. INTRODUCTION

Under the special natural conditions (e.g., overcast and foggy) and fragile ecological environment in mountainous terrain of southern China, land use and land cover (LULC) change have occurred frequently. In addition, land covered composition and its change serve as a crucial role as agricultural production, food security and sustainable development in mountainous terrain. Therefore, accurate and up-to-date information on land cover and its dynamic change are increasingly necessary at different spatial and temporal scales.

In recent years, many change detection methods [1]–[10], [13]–[15], [20], [26] have been developed to derive land cover change information from remote sensing image, such as, principal component analysis and k-means (PCA_Kmeans) clustering based change detection method [14], spectral variance and slow feature analysis (SSFA) based change detection method algorithm [15], local estimation and global search based deep network (LEGS) [4], and semi-supervised fuzzy C-means (Semi_FCM) clustering based change detection method [5]. However, most of these methods only focus

on the remote sensing image acquired by satellite sensors (e.g., Landsat, MODIS and SPOT-VGT), and the relatively low spatial resolution limited the identification of them due to the small size and scattered distribution of land cover in mountainous terrain. Compared with the above-mentioned methods, Wei *et al.* [9] and Milas *et al.* [11] capture more land distribution details than satellite remote sensing images by small unmanned aerial vehicle (UAV) with a small digital camera. There has to exist visual difference in camera viewpoints, although they were captured from a same location and were matched using GPS data. Deep networks are robust (i.e., invariant) to differences in viewpoints and illumination condition, and nevertheless are sensitive to highly-abstract, semantic differences of images. Specifically, recently popular convolutional neural networks (CNNs) are particularly well suited this task, Li and Yu [12] supposed a high-quality visual saliency model can be learned from multi-scale features extracted using CNNs.

Moreover, some factors cause that these image pairs acquired cannot apply directly to identify regions of change since scale changes, noises, geometrical distortions, and discontinuous rotated images with illuminated variations may also be produced in such multi-temporal images. These factors are as follows: (i) when satellite revolves around its orbit, image acquired can have geometrical distortions due to the modeling inaccuracy of the sensor geometry, and the jitter of the instruments platform during image acquisition. (ii) when collecting multi-temporal images for the same location, the imaging perspective of small UAVs is often easily affected by wind speed/direction, complex terrain, aircraft posture (pitch, roll, yaw), flying height and other human factors. In order to effectively improve the matching degree between the image and the actual terrain, the preprocessing of these image pairs is an essential step, i.e., image registration method can align these image pairs of the same scene taken from different viewpoints, from different times or with different sensors. However, most of the current registration methods are only suitable for a type of sensor, and are not sensitive enough to multi-temporal image pairs. Therefore, our goal focus on multi-temporal remote sensing image pairs acquired by different type of sensors, and transforms them into the same coordinate system.

Numerous algorithms [21], [24], [25], [27]–[31] for different registration scenarios have been presented in the last few decades. The coherent point drift (CPD) algorithm for both rigid and non-rigid point set registration [21] treated one point set as centroids of a Gaussian mixture model, and then fitted it to the other. It applied a fast Gaussian transform [22] and low-rank matrix [23] approximation techniques to reduce a large computational burden. Recently, in order to estimate correspondence relationship between two images, GLMDTPS [24] proposed a global and local mixture distance. PRGLS [25] used the point registration as the estimation of a mixture of densities to preserve both global and local structures during matching. More recently, Zhang *et al.* [28], [29] introduced an effective method that maintains a high matching ratio on inliers while taking advantage of outliers for varying the warping grids.

In this paper, we present a robust set of change detection framework for monitoring land cover change in mountainous terrain with multi-temporal remote sensing images. In the preprocessing stage of change detection, a multi-scale feature based image registration method is proposed to align image pairs acquired by different type of sensors. Compared with the current methods, the major contributions of our work include: (i) multi-scale feature descriptor (MFD) constructed by CNN-based feature descriptor (CFD) and shape context (SC). CFD is generated by layers formed a pretrained VGG network; (ii) to estimate correspondence and transformations, a gradually increasing selection of inliers is realized instead of using a stationary distinction of inliers and outliers. At the early stage of registration, the rough transformation is quickly determined by the most reliable feature points. After which the registration details are optimized by increasing the number of feature points. Then, L_2-minimizing estimate (L_2E) based energy optimization is formulated to calculate a reasonable position in a reproducing kernel Hilbert space; (iii) fuzzy C-Means classifier is adopted to generate a similarity matrix between transformed image pairs.

The rest of the paper is organized as follows. Section II introduces a novel deep learning based framework, which is infused with the CNN feature and the deep neural network (DNN), to detect land cover change in mountainous terrain. Section III demonstrates our experiments; and Section IV draws conclusions.

II. METHODOLOGY

In this section, we first give the details of three contributions

- multi-scale feature description;
- dynamic inlier selection;
- fuzzy C-means classifier based pre-classification.

Second, we give the details of the proposed land cover change detection framework. Figure 1 shows the framework of the proposed method. Finally, our algorithm and parameter settings are discussed in the latter part of this section.

Let us consider a image pair I_{t_1} and I_{t_2}, acquired over the same geographical area at two different time t_1 and t_2. The feature point sets $\mathbf{A}$ and $\mathbf{B}$ first are extracted from I_{t_1} and I_{t_2} respectively. Next the transformed image I_t is obtained by our registration algorithm. Note that I'_{t_1} and I'_{t_2} is obtained by equal split of I_t and I_{t_2} according to a certain ratio. Finally, we input I'_{t_1} and I'_{t_2} into the model as change detection, and a change detection map S_{map} will be generated.

Throughout the paper we use the following notations:

- $\mathbf{A}_{N\times D} = \{a_1, ..., a_N\}^T, \mathbf{B}_{M\times D} = \{b_1, ..., b_M\}^T$ - feature point sets are extracted from a image pair I_{t_1} and I_{t_2}, respectively. D denote the dimension of feature point sets, and $D = 2$.
- τ - the transformation function.
- $\mathbf{B}^*$ - transformed locations of source point set $\mathbf{B}$.
- I_t - the transformed image.
- I'_{t_1}, I'_{t_2} - obtained by equal split of I_t and I_{t_2}.
- S_{map} - a change detection map.

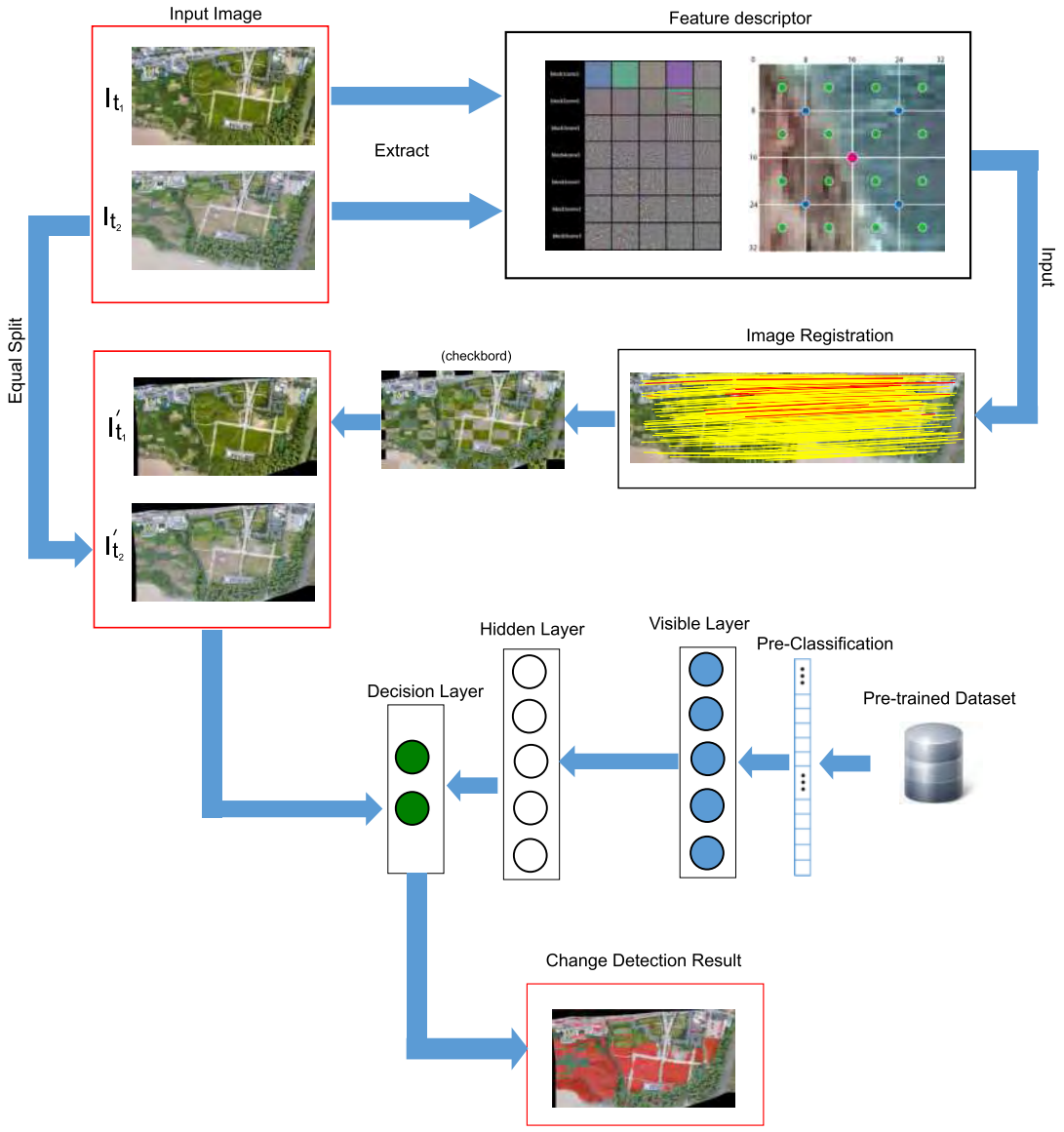

FIGURE 1. Flowchart of the proposed land cover change detection framework, consisting of three main phases: (1) a multi-scale feature description, (2) an effective registration processing, and (3) an effective change detection processing. Note that correct feature point matches are denoted by yellow lines, incorrect ones are denoted by red lines.

A. MULTI-SCALE FEATURE DESCRIPTOR

The mountainous terrain is the major geomorphic structure in the south of China, and have special natural conditions (e.g., overcast and foggy), a fragile ecological environment. Therefore, it is sometimes hard to perform precise image registration since images acquired by different type of sensors can aggravate the non-rigid geometric distortions of images. We will attempt the features extracted by the convolutional neural networks (CNNs) to improve the feature expression.

1) CNN-BASED FEATURE DESCRIPTOR (CFD).

CFD is constructed by one of the state-of-the-art CNN which used the VGG-16 architecture and pre-trained on ImageNet dataset for image classification [32]. VGG-16 has sixteen layers (as shown Figure 2) including 5 blocks of convolution computation, each with 2-3 convolution layers and a max-pooling layer at the end of each block, from which we select one of its pool3, pool4 and pool5_1 layers. We lay a 28×28 grid over the input image dividing our patches, each corresponding to a $256-d$ vector in the pool3 output, a descriptor is generated in every 8×8 square. The center of each patch is regarded as a feature point. The $256-d$ vector is defined as the pool3 feature descriptor. The pool3 layer output directly forms our pool3 feature map f_1, which is of size $28 \times 28 \times 256$. The pool4 layer output, which is of size $14 \times 14 \times 512$, is handled slightly differently. In every 16×16 area we obtain a pool4 descriptor, and pool4 feature map f_2

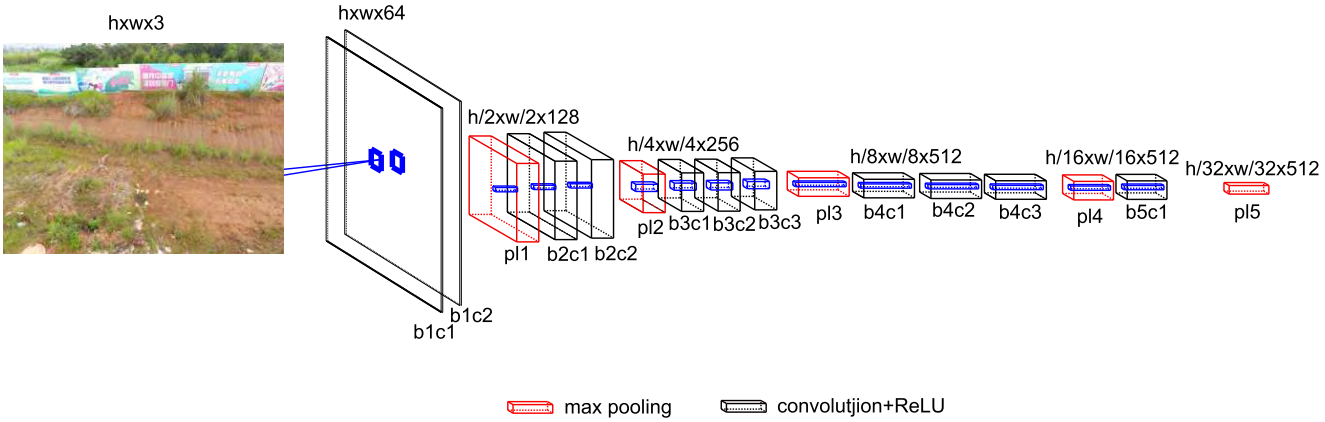

FIGURE 2. Architecture of modified VGG-16 network. *h* and *w* denote the height and width of the input image, respectively. Since we only use convolution layers to extract features, the input image will not adjust the size to keep the feature of the original image as long as *h* and *w* are multiples of 32.

is written as:

$$f_2 = \mathbf{O_{p4}} \bigotimes \mathbf{I_{2\times 2\times 1}} \tag{1}$$

where $\bigotimes$ denotes Kronecker product. $I_{2\times 2\times 1}$ presents a tensor of subscripted shape and filled with 1*s*. Note that f_2 is shared by 4 feature points.

The pool 5_1 layer output is of size $7\times 7\times 512$, and pool5_1 feature map f_3 takes the form:

$$f_3 = \mathbf{O_{p5\_1}} \bigotimes \mathbf{I_{4\times 4\times 1}} \tag{2}$$

Similarly, every pool5_1 descriptor is shared by 16 feature points. The distribution of feature descriptors is shown is Figure 3. After producing f_1, f_2 and f_3, the feature maps are

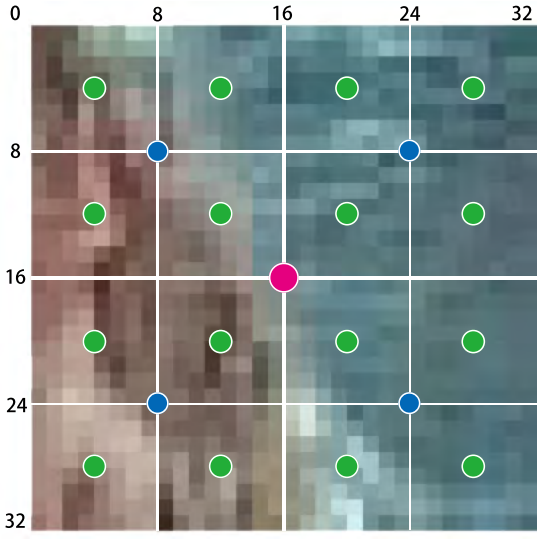

FIGURE 3. Distribution of feature descriptor. It is shown in a 32 × 32 squared region. Green dots represent pool3 descriptors, generated in a 8 × 8 squared region. Blue dots represent pool4 descriptors, each shared by 4 feature points. The cyan dot represent a pool5_1 descriptor, shared by 16 feature points.

normalized to unit variance using

$$f_k = \frac{f_k}{\sigma(f_k)}, \quad k = 1, 2, 3 \tag{3}$$

where $\sigma(.)$ calculates the standard deviation of elements in a matrix. Therefore, the pool3, pool4 and pool5_1 descriptors of point sets **A** and **B** are represented by $F_1(a)$, $F_2(a)$ and $F_3(a)$, $F_1(b)$, $F_2(b)$ and $F_3(b)$ respectively.

CFD is used to measure the feature distance of between two feature point sets **A** and **B**. CFD is a weighted sum of three distance values, and is written as:

$$F_{cfd} = \sqrt{2}F_1(a, b) + F_2(a, b) + F_3(a, b) \tag{4}$$

where each component distance value $F_k(a, b)$ is the Euclidean distance between the respective feature descriptors

$$F_k(a, b) = Euclidean - disance(F_k(a), F_k(b)) \tag{5}$$

The distance computed with pool3 descriptors $F_1(a, b)$ is compensated with a weight $\sqrt{2}$ because F_1 is 256-d whereas F_2 and F_3 are 512-d. C_{cfd} is a cost matrix of CFD, and the matrix form is as follows

$$C^{\theta}_{cfd}(m, n) = \begin{cases} \dfrac{F(b_m, a_n)}{F^{max}_{\theta}} & c_1, \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

where c_1 denotes a valid match of b_m and a_n under threshold θ, F^{max} is the maximum distance of all matched feature point pairs under threshold θ.

2) SHAPE CONTEXT (SC).

The SC [33]–[35] could play such a role in shape matching. Consider a center point a_n on the first shape and a center point b_m on the second shape. Firstly, the SC constructs a polar coordinate system (see Figure 4). Then, we compute a histogram h^a_n or h^b_m of the relative coordinates of the remaining $n-1$ or $m-1$ points for a_n or b_m on

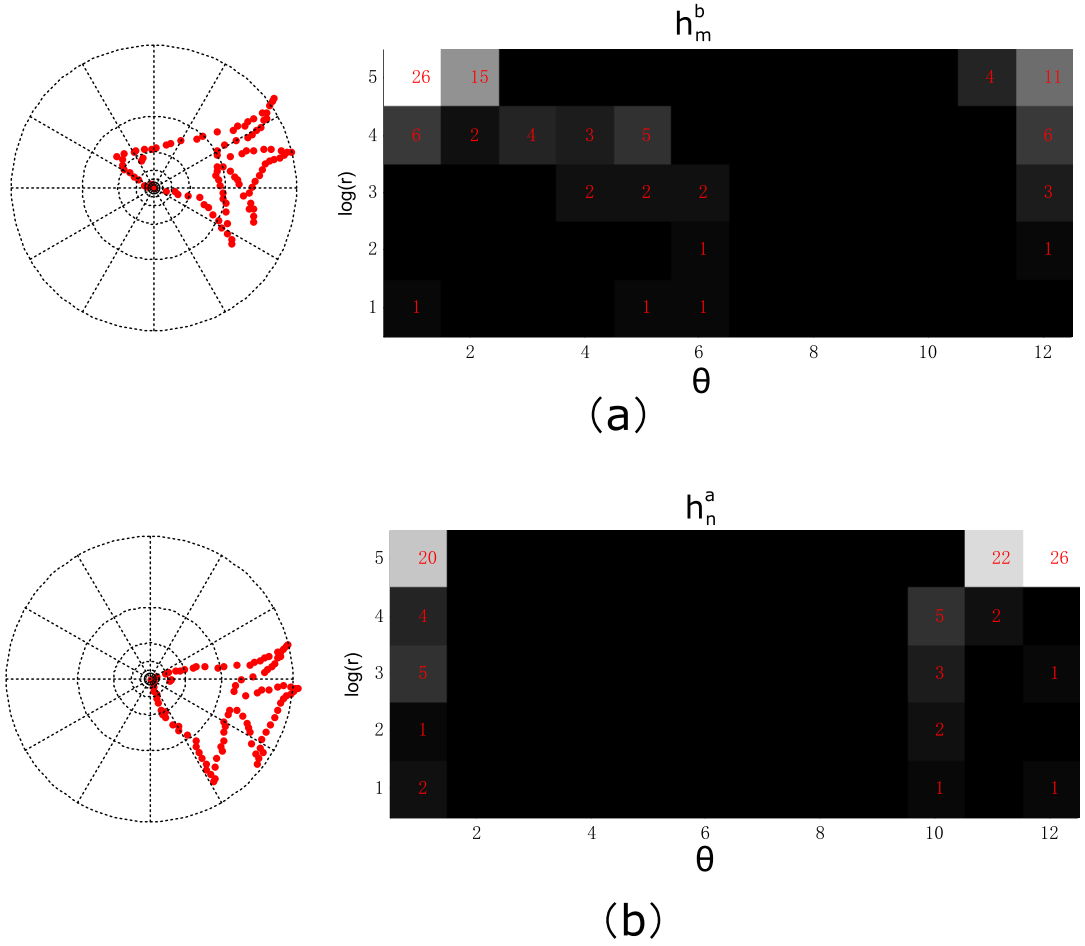

FIGURE 4. Shape context (SC) computation and matching. Left of (a) and (b): diagrams of log-polar histogram bins centered at a_n and b_m used in computing the shape contexts. We use 5 bins for $log(r)$ and 12 bins for θ. Right of (a) and (b): each shape context e.g., h^b_m or h^a_n is a log-polar histogram of the coordinates of the rest of the point set measured using the centered point as the origin.

the shape respectively. $C_{sc}(m, n)$ denotes the cost of matching these two point sets, and is measured using Chi-square distribution as:

$$C_{sc}(m,n) = \frac{1}{2}\sum_{x=1}^{X}\frac{[h^b_m(x) - h^a_n(x)]^2}{h^b_m(x) + h^a_n(x)} \tag{7}$$

where $h^b_m(x)$ and $h^a_n(x)$ are two $1 \times X$ sets, and denote the number of points within each bin surrounding b_m and a_n, respectively.

3) MIXTURE FEATURE DESCRIPTOR (MFD).

We first compute a integrated cost matrix C_{mfd} using a element-wise Hadamard product (denoted by $\odot$), and is written as:

$$C_{mfd} = C^{\theta}_{cfd} \bigodot C_{sc} \tag{8}$$

where C^{θ}_{cfd} and C^{θ}_{sc} are value in $[0, 1]$. Then, we apply Jonker-Volgenant algorithm [37] to solve the linear assignment on cost matrix C_{mfd}. Assigned point pairs are regarded as putatively corresponding.

B. DYNAMIC INLIER SELECTION

Our feature points are acquired at the center of square shaped image patches. Due to reasons of large rotation angles and deformation, corresponding feature points may have their image patches overlapping partly or completely. Thus, to improve the effect of the registration, feature points with large overlapping ratios should have a better degree of alignment, where as partly overlapping patches should have a small distance between their centers. Therefore, the degree of alignment is determined using our dynamic inlier selection.

In point set registration, there are several ways to estimate the parameters of the mixture model, such as the EM algorithm, gradient descent and variational inference. Our point set registration mainly contains the following two steps: (i) correspondence estimation, the corresponding target point set $\mathbf{A}_{\psi}$ is estimated between $\mathbf{B}$ and $\mathbf{A}$; (ii) transformation Updating, the transformation function τ is established to update the position of $\tau(\mathbf{B})$ constantly, until $\tau(\mathbf{B})$ and $\mathbf{A}_{\psi}$ can overlap as much as possible. Note that $\tau(\mathbf{B})$ (initial $\tau(\mathbf{B}) = \mathbf{B}$) indicates the transformed set $\mathbf{B}$ in each iteration.

Therefore, the inliers of selection are assigned in every k iteration to iteratively address $\mathbf{B}$. Note that these inliers guide for bundle adjustment of point locations whereas outliers are moved coherently. At the feature prematching stage, a low threshold θ_0 is applied to filter out irrelevant points and select coarsely a large number of feature points. Then, a large starting threshold $\tilde{\theta}$ is adopted to select confident inliers satisfy. In the rest of registration process, threshold θ is subtracted by a step-length ι in every k iterations, allowing a few more feature points with high similarity to affect the estimating correspondence and transformation. Such technology enables feature points with high similarity to complete the overall transformation while other feature points optimize registration accuracy.

C. PRE-CLASSIFICATION

The pre-classification step chooses the pixels that are best suited to train the deep neural network. The Fuzzy C-Means (FCM) is a popular image segmentation technique that segments an image by discovering cluster centers. Suppose a'_{ij} and b'_{ij} denote gray levels of the image pixels at the corresponding positions (i, j) in I'_{t_1} and I'_{t_2}, respectively. We use FCM classifier to provide jointly classify for the two input images, and a similarity matrix s'_{ij} is established.

$$s'_{ij} = \frac{|a'_{ij} - b'_{ij}|}{a'_{ij} + b'_{ij}} \tag{9}$$

where $0 \leq s'_{ij} \leq 1$. Then, a global threshold value of similarity T will be applied to s'_{ij} by the iterative threshold method. Iterate over all a'_{ij} and b'_{ij}, if $s'_{ij} > T$, then jointly label a'_{ij} and b'_{ij} by FCM based on the principle of minimum variance δ^2_{ij}. Otherwise label a_{ij} and b_{ij} separately. δ^2_{ij} is written as:

$$\delta^2_{ij} = a'_{ij}\frac{a'_{ij}b'_{ij}}{a'_{ij} + b'_{ij}}[s'_{ij}]^2 \tag{10}$$

The gray-level of each pixel in the same position of the corresponding two original images are compared to label the pixels. The label of a pixel and its surrounding neighborhood can be used to determine if a pixel is either part of an edge or noise. The results are then passed to the neural network for training.

D. MAIN PROCESS

1) IMAGE REGISTRATION

To effectively eliminate the geometric error and improve the matching degree between the image and the actual terrain, image registration is an essential step in the preprocessing of remote sensing image including two processes: feature point set registration and image transformation. Firstly, we carry out feature point set registration.

- Correspondence Estimation. Gaussian mixture model (GMM) has been proven the popular model in computer vision and pattern recognition. Thus, the set $\mathbf{B}$ are used as GMM centroids, and the set $\mathbf{A}$ as the data points generated by the GMM. The GMM probability density function is

$$p(a_n|b_m) = -\frac{1}{2\pi\sigma^2}\exp\frac{1}{2\sigma^2}\|a_n - b_m\|^2 \tag{11}$$

Then, the outlier and noise distribution is supposed an additional uniform distribution $p(a|M+1) = \frac{1}{N}$, which is added to the mixture model. Thus, the mixture model takes the form

$$p(a_n) = (1-\varepsilon)\sum_{n=1}^{N}\log\sum_{m=1}^{M}p(m)p(a_n|b_m) + \varepsilon\frac{1}{N} \tag{12}$$

where $p(m) = \frac{1}{M}$ denotes the mixed weight that are nonnegativity and sum-to-one. We use equal isotropic covariances σ^2 and equal membership probabilities $p(m)$ for all GMM components ($m = 1, ..., M$). ε denotes the weight of the uniform distribution, with $0 \leq \varepsilon \leq 1$. We compute the revised parameter as:

$$\varepsilon = 1 - \frac{\sum_{n=1}^{N}\sum_{m=1}^{M}p(m|a_n)}{N} \tag{13}$$

Subsequently, inlier selection calculate a $m \times n$ prior probability matrix p_{mn} which is then taken by our Gaussian mixture model (GMM) based transformation solver.

$$p_{mn} = \begin{cases} 1 & \text{if } b_m \text{ and } a_n \text{ are corresponding,} \\ \dfrac{1-\upsilon}{N} & \text{otherwise} \end{cases} \tag{14}$$

where $\upsilon \in (0, 1)$ should be designated according to our confidence of the inlier selection to be accurate. Prior probability matrix requires normalization:

$$p_{mn} := \frac{p_{mn}}{\sum_{k=1}^{N}p_{mk}} \tag{15}$$

By the equation (15), the $M \times N$ posterior probability matrix is obtained, which is used as the fuzzy correspondence matrix $\mathbf{P}$ between I_s and I_r. Then, the corresponding target point set is obtained by

$$\mathbf{A}_\psi = \mathbf{PA} \tag{16}$$

Though the target coordinate $\mathbf{A}_\psi$ is estimated by GMM, the method will inescapable produce mismatching.

- Transformation Updating. Firstly, a positive definite kernel (e.g., Gaussian kernel) is chosen; and a reproducing kernel hilbert space (RKHS) [38], [39] $\mathcal{H}$ is defined. Then, we employ the Gaussian Radial Basis Function (GRBF), which is in the form $\mathcal{G}(b_i, b_j) = \exp(-\frac{|b_i - b_j|^2}{\beta^2})$, where β is a constant to control the spatial smoothness and $\mathcal{G}$ is of size $m \times m$. According to the representation theorem, a displacement function $\nu(\mathbf{B})$ takes the form

$$\tau(\mathbf{B}) = \sum_{m=1}^{M}\mathcal{G}(b, b_m)\psi \tag{17}$$

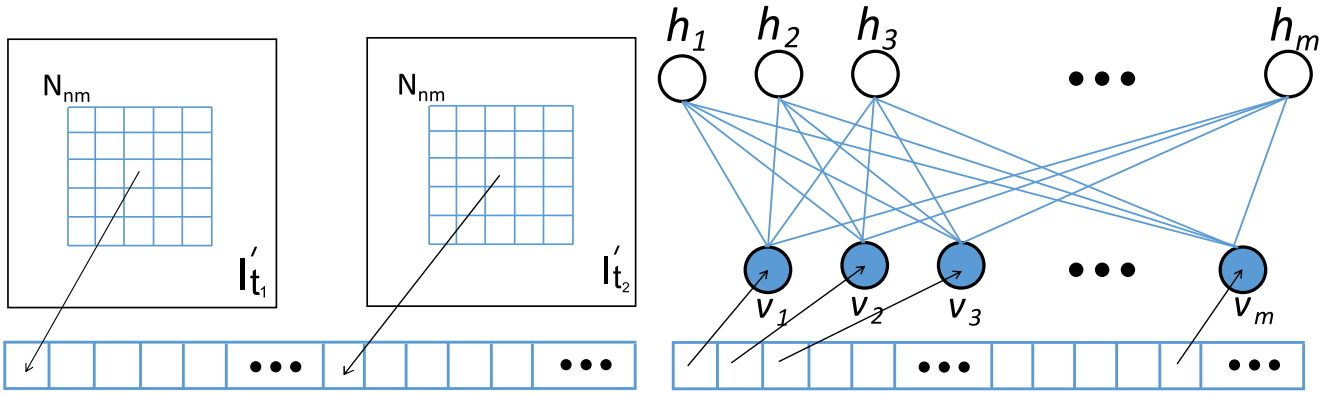

FIGURE 5. Left: Vectorization of neighborhood features to be fed into the network. Right: The structure of an RBM, consisting of two layers, one visible (v) and one hidden (h), with no connections within a layer. Hidden nodes are indicated by blue filled circles and the visible nodes indicated by unfilled circles.

where $\psi = (\psi_1, \psi_2, ..., \psi_m)^T$ is a $D \times 1$ coefficient matrix. Therefore, the minimization over energy equation in $\mathcal{H}$ boils down to finding a finite coefficients matrix ψ. transformation function $\nu(\mathbf{B})$ is equivalent to the initial position plus a displacement function $\tau(\mathbf{B})$, i.e., $\nu(\mathbf{B}) = \mathbf{B} + \tau(\mathbf{B})$.

Though the reliable target coordinate $\mathbf{A}_\psi$ is estimated by GMM, the method will inescapable produce mismatching. Therefore, our next concern draws on formulating a function, by which a reasonable position $\tau(b_m)$ of b_m is determined. This position in turn improves the accuracy of the correspondence estimation as subsequent iterations interlock. Since the error of L_2-minimizing estimator is less than the error of maximum likelihood estimation (MLE), L_2 Euclidean an distance is widely used in multiple applications, and many registration methods, especially, the problem of point set registration can be well formulated by minimizing the L_2 Euclidean distance between two point sets. Therefore, we employ the L_2E [39] based energy function to estimate the transformation function τ, which is written as

$$E(\psi, \sigma^2) = \frac{1}{2^D(\pi\sigma)^{\frac{D}{2}}} - \bar{p} + \lambda \parallel \tau \parallel^2_{\mathbf{G}} \qquad (18)$$

where $\bar{p} = \frac{2}{m}\sum_{m=1}^{M} \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} \exp\left(-\frac{\|\mathbf{A}_\psi - \mathbf{U}_{m,\cdot}\psi\|^2}{2\sigma^2}\right)$, $\mathbf{U}_{ij} = \mathcal{G}(b_i, b_j)$, $\mathbf{U}_{m,\cdot}$ denotes the m^{th} row of matrix $\mathbf{U}$, ψ_i denotes the i^{th} row of the coefficient matrix $\psi_{h\times D}$. Next, we can directly take the partial derivatives of equation (18) with respect to coefficients matrix ψ, By setting them to zero, and solve the resulting linear system of equations. As follows:

$$\frac{\partial E}{\partial \psi} = \mathbf{U}^T\left(\frac{2\Phi \odot (\mathbf{H} \otimes \mathbf{1})}{n\sigma^2(2\pi\sigma^2)^{\frac{D}{2}}}\right) + 2\lambda\mathcal{G}\psi \qquad (19)$$

where $\Phi = \mathbf{U}\psi - \mathbf{A}_\psi$, $\mathbf{H} = \exp\{diag(\Phi\Phi^T)/2\sigma^2\}$ is a $M \times 1$ vector, $diag(\cdot)$ denotes the diagonal of a matrix, $\mathbf{1}$ is a $1 \times D$ row vector of all ones. Symbols $\odot$ and $\otimes$ denote the Hadamard product and Kronecker product, respectively.

After updating the coordinates of the source point set by $\mathbf{B} = \mathbf{B} + \mathbf{U}\psi$, we anneal the covariances of the GMM by $\sigma^2 = \rho\sigma^2$, then return to correspondence estimation and continue the feature point sets registration process until the maximum iteration number is reached. Note the transformed source point set $\mathbf{B}^*$ is obtained in the final iteration. Next, we employ the backward approach [40] to establish a thin-plate spline (TPS) [41] transformation model, then the transformed image I_t can be calculated using the model. I'_{t_1} and I'_{t_2} is obtained by equal split of I_t and I_{t_2} according to a certain ratio. (see Figure 1)

2) ESTABLISHING AND TRAINING THE DEEP NEURAL NETWORKS FOR CHANGE DETECTION

Although the difference image method is well researched, change detection is a comprehensive procedure that requires careful consideration of many factors such as the nature of change detection problems, image preprocessing, selection of suitable variables and algorithms. DNN has brought in profound and revolutionary changes to the realm of artificial intelligence, and achieved great improvements in many domains such as computer vision, speech recognition and natural language processing, etc. Therefore, we employ DNN to train the pre-classification results and create a change detection map from pre-processed image pair directly without generating difference images. After pre-classification, the neighborhood features of each pixel and its corresponding pixel in another image are converted into a vector as inputs to a neural network.

The Restricted Boltzmann Machine (RBM). RBM is a stochastic neural network, which consists of two layers of binary units: a visible layer $\mathbf{v}$ with n visible units and hidden layer $\mathbf{h}$ with m hidden units. An example of this structure is in Figure 5 with the hidden nodes indicated by blue circles and the visible nodes indicated by white circles. A common use for RBMs is to create features for use in classification. The energy function of the RBM model for visible and hidden units can be represented by the following:

$$E(\mathbf{v}, \mathbf{h}) = -\eta^T v - \varsigma^T h - h^T \mathbf{W} v \qquad (20)$$

where η and ς are biases of the visible units and hidden units, respectively. The matrix $\mathbf{W}$ denotes weights of the connection between visible and hidden layer units, where each matrix element is the conditional probability of the next layer neuron conditioned on the previous layer neuron. The joint probability distribution of visible units v and hidden units h of the RBM is interpreted by

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} e^{E(v,h)} \tag{21}$$

where $Z = \sum_{v'} \sum_{h'} e^{E(v',h')}$ is the partition function of the system. The conditional distributions are:

$$P(h_j = 1|\mathbf{v}) = \sigma(\varsigma_j + \mathbf{v}^T \mathbf{W}_{(:,j)}) \tag{22}$$

$$P(v_i = 1|\mathbf{h}) = \sigma(\eta_i + \mathbf{W}_{(i,:)}\mathbf{h}) \tag{23}$$

Previous studies [42], [43] that the updating rules for $\mathbf{W}$, η and ς during the training process with a learning rate γ are the following:

$$\Delta \mathbf{W}_{ij} = \gamma(\langle v_i h_j \rangle_d - \langle v_i h_j \rangle_m) \tag{24}$$

$$\Delta \eta_i = \gamma(\langle v_i \rangle_d - \langle v_i \rangle_m) \tag{25}$$

$$\Delta \varsigma_j = \gamma(\langle h_j \rangle_d - \langle h_j \rangle_m) \tag{26}$$

where $\langle \cdot \rangle_d$ and $\langle \cdot \rangle_m$ are the expectations under the distribution specified by the training input data and the theoretical RBM model. Although computing $\langle v_i h_j \rangle_d$ is straightforward, $\langle v_i h_j \rangle_m$ is intractable due to the large number of possible joint (v, h) configurations. Contrastive Divergence (CD) algorithm [42] is a learning procedure being used to approximate $\langle v_i h_j \rangle_m$. For every input, it starts a Markov Chain by assigning an input vector to the states of the visible units and attempts a small number of full Gibbs sampling steps. Resulting reconstructed visible units are applied to approximate the expectation of the model distribution.

A deep neural network (DNN) [44] pre-trained via stacking restricted Boltzmann machines (RBMs) demonstrates high performance. Therefore, we utilize DNN to train networks using the features of images for detecting land cover change. The process mainly contains the following three steps: (1) neighborhood features of each pixel at the same location on the image pair are fitted to the DNN; (2) RBMs are then unrolled to create a deep neural network for training. Note that CD training algorithm is used to pre-train each RBM in the stack of RBMs via training data; (3) DNN is fine-tuned by the backpropagation of error derivatives.

After the training, the deep neural network is established. Next, I'_{t_1} and I'_{t_2} can be inputted to the network and a robust and contractive change map S_{map} will be built.

E. OUR ALGORITHM AND PARAMETER SETTINGS

Our method is summarized in Algorithm 1. There are five groups of parameters in our method: (1) in the feature pre-matching stage, treshold θ_0 matically determined by selecting the most reliable 128 pairs of feature points . Similarly, $\tilde{\theta}$ is determined by selecting the most reliable 64 pairs of feature points; (2) in the inlier selection stage, the step-length ι is found by $\delta = (\tilde{\theta} - \theta_0)/10$, the covariance parameters δ^2; (3) outlier balancing weight is initialized as 0.5; (4) in the Gaussian radial basis function (GRBF), β is used to control the spatial smoothness. Since we normalize the spatial coordinates of the sensed image feature points to [-1.5, 1.5], β set to 2; (5) in equation (15) and energy equation (18), δ^2 are initialized to 1 and 0.05 respectively.

Algorithm 1 Land Cover Change Detection Using Multi-Temporal and Multi-Sensor Remote Sensing Images in Mountainous Terrain

Input: The source point set **A** and the target point set **B**
Output: The transformed image I_t
Initialize θ_0, $\tilde{\theta}$, ι, k, β, ω, δ^2, W and λ;
Image Registration.
while *not reach the maximum iteration number* **do**
  Correspondence Estimation:
    Compute C^{θ}_{cfd}, C_{sc} and C_{mfd} by equation (6), (7) and (8), respectively;
    Compute the posterior probability matrix **P** by equation (15);
    Compute the corresponding target point set $\mathbf{A}_{\psi}$ by equation (16);
  Transformation Updating:
    Construct the kernel matrix $\mathcal{G}$ and **U**;
    Compute ψ by using equation (18);
    Update $\sigma^2 = \rho\sigma^2$;
    Update the sensed image's feature point set by $\mathbf{B} = \mathbf{B} + \mathbf{U}\psi$;
end
The transformed source point set $\mathbf{B}^*$ is obtained in the final iteration;
The transformed image I_t can be calculated using a thin-plate spline (TPS) [41] transformation model that is established by the backward approach [40].
Change Detection.
  Pre-Classification:
    Compute a similarity matrix s'_{ij} by equation (9).
    Set a global threshold value of similarity T .
  Pre-Establishing and Training the Deep Neural Networks:
    The neighborhood features of each pixel and its corresponding pixel in another image are converted into a vector as inputs to DNN.
    training.
  Test:
    I'_{t_1} and I'_{t_2} can be inputted to the network.
    A change detection map S_{map} will be obtained.

III. EXPERIMENTS AND RESULTS

A. STUDY AREA AND DATA SOURCE

The study was mainly carried out in the ten key land conservation regions of Sichuan, Guizhou and Hunan China (see Figure 6). The regions locate in the warm temperate zone

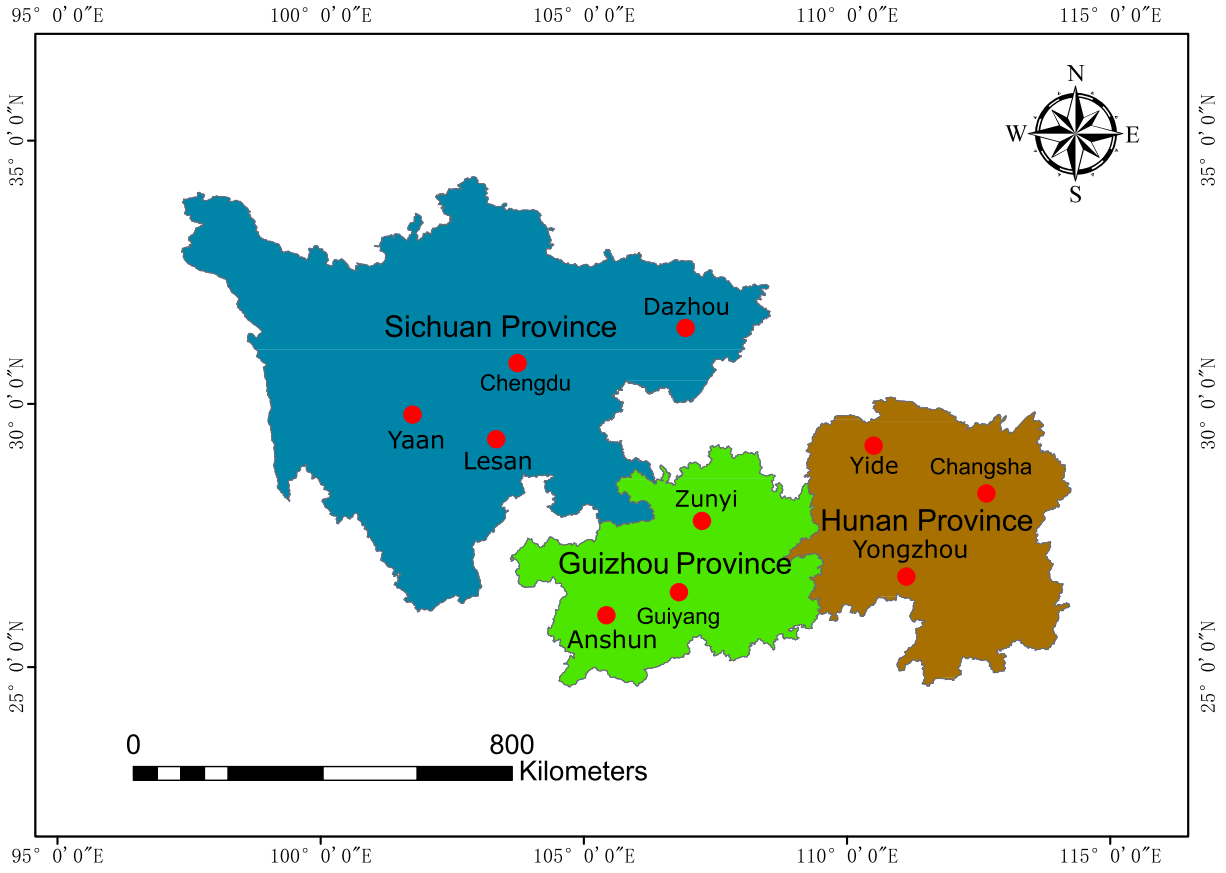

FIGURE 6. Location of study area in mountainous terrain of southern China. Red dots represent ten key land conservation regions regions of Sichuan, Guizhou and Hunan China. Note that Sichuan Province, China (Longitude range: 97°21′E to 108°33′E; Latitude range: 26°03′N to 34°19′N). Guizhou Province, China (Longitude range: 103°36′E to 109°35′E; Latitude range: 24°37′N to 29°13′N). Hunan Province, China (Longitude range: 111°53′E to 114°15′E; Latitude range: 27°51′N to 28°41′N).

TABLE 1. The experimental dataset (I) and (II).

Attributes	Dataset	
	(I)	(II)
Sensors	Landsat7 and Landsat8	GF1 and GF2
Type	Same satellite	Different satellites
Number of image pairs	2000	2000
Images size (mm)	$200 \times 450 \sim 1200 \times 800$	$200 \times 450 \sim 1200 \times 800$
Number of feature points	$100 \sim 400$	$100 \sim 400$

and have four distinct seasons because of the continental monsoon. These areas have a variety of land cover types including cropland, building-up, forest, etc. Among these land cover types, the most dominant one is cropland, which can be easily affected by pseudo changes of phenological differences. In addition, we also obtained some satellite remote sensing data from other foreign mountainous terrain to verify the applicability of the method.

We evaluate the performance of the proposal framework on an available data set. The data set contains a total of 6000 image pairs. To facilitate a fair comparison with other methods, we divided this dataset into three pars, 3000 for training, 1000 for validation and the remaining 2000 image pairs for testing. In order to achieve better training effect, date set is formed by two categories of remote sensing image pairs:

(1) 4000 image pairs are acquired by different type of multi-sensor and multi-temporal satellites including Chinese GF, Landsat. The details of dataset (I) and (II) are summarized in Table 1. In this dataset, a same satellite generally follow the same orbital paths with the same viewing angles and passed over a certain spot on earth at the same local time due to orbital mechanics. Therefore, image pairs acquired by

the same satellite cannot contains lager viewpoint change. However, image pairs acquired by different sensors suffer serious scale change.

TABLE 2. The experimental dataset (III).

Attributes	Values
Sensors	DJI Phantom 4 Pro
Number of image pairs	2000
Images size (mm)	$200 \times 450 \sim 1200 \times 800$
Depression angle (°)	$30 \sim 60$
Horizontal rotation angle (°)	$-90 \sim 90$
Number of feature points	$70 \sim 250$

(2) 2000 image pairs are acquired by a small-sized UAV, the DJI Phantom 4 Pro (DJI, Shenzhen, China) with a CMOS camera, basically maintained the same flight altitude (around $50 \sim 150\ m$) for collecting these images of the same location at different times. The details of dataset (III) are shown in Table 2. In this dataset, it was not easy to navigate the aircraft along the planned lines since operation of small UAV is always limited by air traffic constraints and monitoring object, e.g., mountainous landforms are the rugged terrain area, and in most seasons these areas often are overcast and foggy. Thus, image pairs of the same scene have to be captured from different viewpoints through multiple flight routes so that a full coverage of the object surface can be obtained. In addition, small UAVs cannot avoid the influence of flight attitude (pitch, roll, yaw) in flight practices due to its flight high, speed, airflow and other factors, which will cause the acquired images to be squeezed, twisted, stretched and offset relative to the target position of ground. Therefore, these image pairs often contain large rotation angles.

B. EXPERIMENTAL DESIGN

To qualitatively evaluate the proposal framework results, two kinds of experiments are conducted: image registration and change detection. The former uses some state-of-the-art methods, such as SIFT [45], SURF [46], CPD [21], GLMDTPS [24] and ZGL_CATE [28]. The latter adopts PCA_Kmeans [14], SSFA [15], LEGS [4] and Semi_FCM [5] . In addition, we adopt two standards and widely used evaluation metrics, precision-recall curve (PRC) and root of mean square error (RMSE). These experiments are performed on a PC with 2.5GHz Intel Core CPU, 8GB memory.

1) IMAGE REGISTRATION ACCURACY TEST

The root of mean square error (RMSE) is used to quantify the image registration accuracy. We manually construct at least 15 pairs of corresponding points in each image pair as landmarks. Note that all the landmarks are well-distributed and selected the interest areas in which the surface features are distinct, easily distinguished places or where the colour contrast is large with nearby surface features. The related formulations are as follows:

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(b_i^t - a_i^t)} \tag{27}$$

The *RMSE* can well reflect the spatial deviation of corresponding landmarks in the sensed image and the reference image, respectively. Where M is the total number of the selected landmarks, and b_i^t is the landmark that corresponds to a_i^t;

2) CHANGE DETECTION PRECISION TEST

A ground truth is compared to the change detection map to measure the accuracy of the performance of change detection. In precision-recall curve, the precision metric measures the fraction of detections that are true positives and the Recall metric measures the fraction of positives that are correctly identified. Precision and Recall can be defined as:

$$Precision = \frac{TP}{TP + FP} \tag{28}$$

$$Recall = \frac{TP}{TP + FN} \tag{29}$$

where TP denotes true positives in which changed pixels are detected correctly, FP denotes the false positives in which are detected as changed when compared with the ground truth, and FN denotes false negatives in which changed pixels are detected as unchanged when compared with the ground truth, respectively.

C. RESULTS AND DISCUSSION

1) RESULTS AND DISCUSSION OF IMAGE REGISTRATION ACCURACY TEST

In this experiment, Table 3 shows quantitative comparisons on image registration measured using the mean RMSE, where it from left to right has a decreasing tendency. In addition, Figures 7, 8 and 9 show results of image registration on six representative image pairs. The results show that our method reached the best performance in most cases, especially when the appearance difference in the image pair is challenging. Therefore, our method can be used widely since an efficient change rule should have robust image registration. Moreover, ZGL_CATE can yield a better performance. However its drawback originates from the extracted feature points that are not sensitive enough to multi-temporal images. CPD performs unsatisfying in some cases. In contrast, GLMDTPS can achieve better performance since it employs mixture feature descriptor. However, GLMDTPS emphasizes one-to-one correspondence relationship which is vulnerable under the presence of outliers.

2) RESULTS AND DISCUSSION OF CHANGE DETECTION PRECISION TEST

In this experiment, the effectiveness of the proposed framework is evaluate by comparing the different change detection methods for PCA_Kmeans, SSFA, LEGS, Semi_FCM.

TABLE 3. Experimental results on image registration. Quantitative comparisons on image registration measured using the mean RMSE are carried out.

Dataset	RMSE (%)					
	SIFT	SURF	CPD	GLMDTPS	ZGL_CATE	Ours
(I)	12.1102	11.7867	7.8326	5.0431	2.9253	**1.0431**
(II)	13.4207	12.0134	9.0390	7.5341	5.8648	**1.2681**
(III)	15.4336	13.5132	10.2496	9.8300	6.9648	**1.9234**

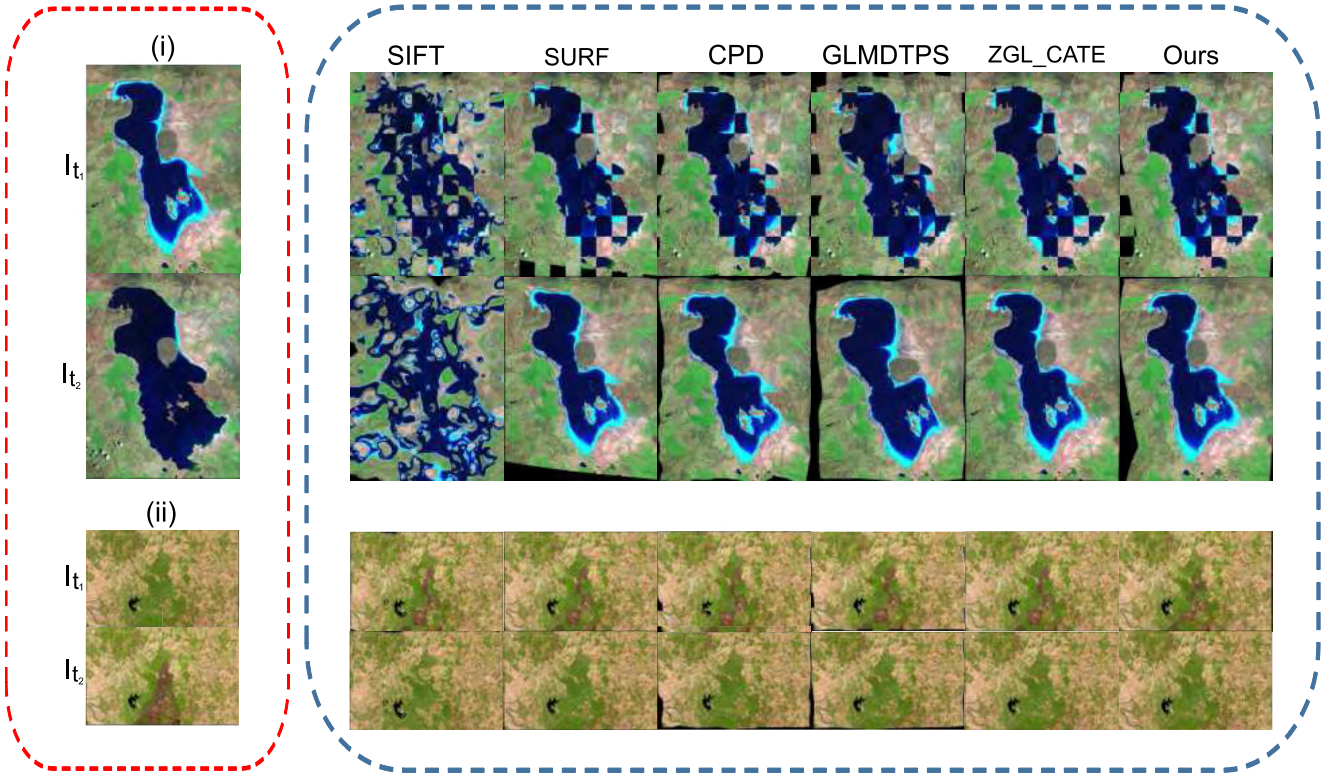

FIGURE 7. Registration examples on two typical image pairs from dataset (I). (i) LakeOroumeih; (ii) Bastrop. Left: Image pair I_{t_1} and I_{t_2} acquired over the same geographical area at two different time t_1 and t_2 by Landsat 8. Right: The first column until the end show the registration results of SIFT, SURF, CPD, GLMDTPS, ZGL_CATE and Ours. For each method, the first row shows 5 × 5 checkboard and the second row shows the transformed image I_t.

TABLE 4. Experimental results on change detection. Quantitative comparisons on change detection measured using the PRC are carried out.

Dataset	(Precision (%), Recall (%))				
	PCA_Kmeans	SSFA	LEGS	Semi_FCM	Ours
(I)	(78.3, 77.4)	(92.4, 91.8)	(93.3, 93.1)	(94.3, 93.8)	**(98.3, 97.5)**
(II)	(73.2, 72.9)	(90.6, 90.9)	(91.7, 91.2)	(92.7, 91.9)	**(97.9, 96.3)**
(III)	(76.8, 77.9)	(91.5, 91.4)	(91.9, 92.6)	(92.2, 92.4)	**(98.4, 96.8)**

The comparison results are depicted Figures 10, 11 and 12 and Table 4. As shown in Table 4, the average precision of our method on dataset (I), (II) and (III) have reached to (98.3%, 97.5%), (97.9%, 96.3%), (98.4%, 96.8%) . However, the average precision of PCA_Kmeans only reach (78.3%, 77.4%), (73.2%, 72.9%), (76.8%, 77.9%). This is mainly because our method adopts DNNs to directly create a change detection map from pre-processed image pair by bypassing the steps of filtering or generating a difference image (DI). In contrast, PCA_Kmeans performs unsatisfying in some cases since it often have noisy result of not considering the spatial relationship among image pixels.

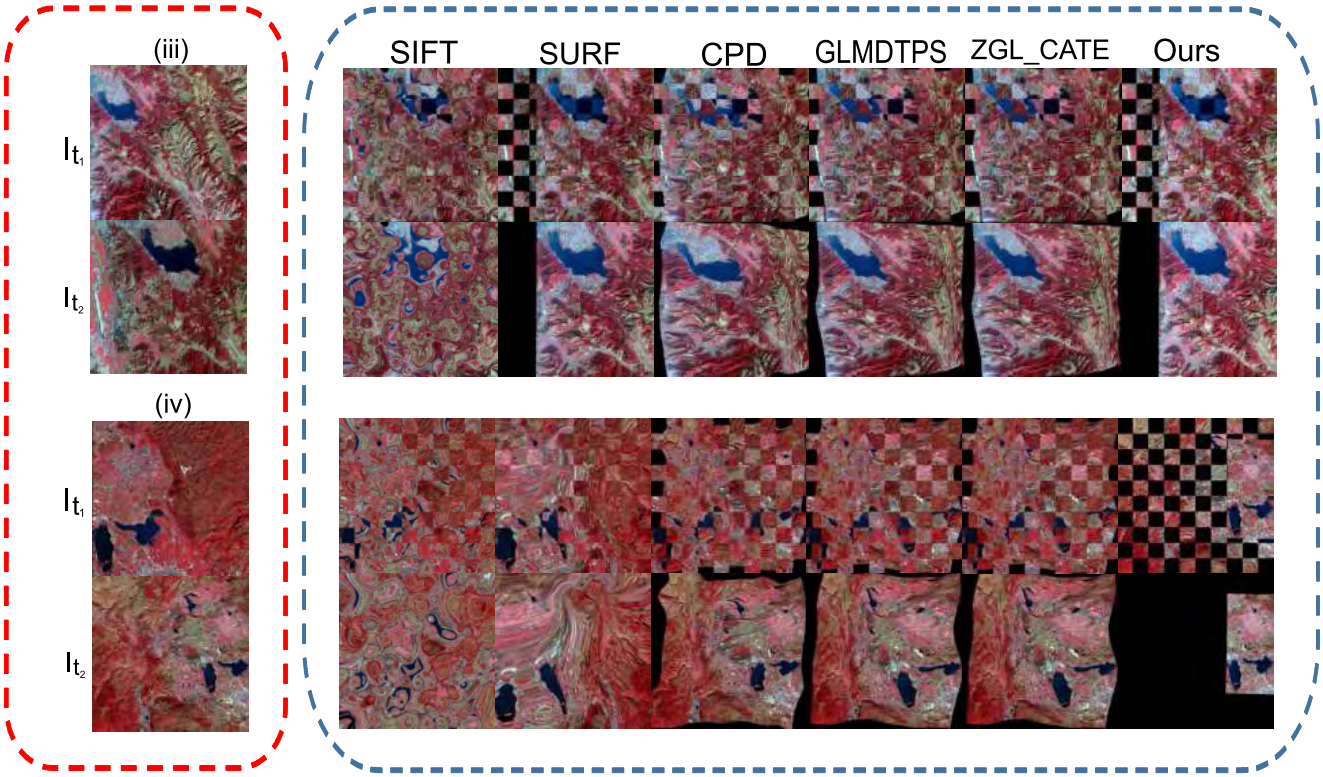

FIGURE 8. Registration examples on two typical image pairs from dataset (II). (iii) Guizou; (iv) Hunan. Left: Image pair I_{t_1} and I_{t_2} acquired over the same geographical area at two different time t_1 and t_2 by Chinese GF1 and Chinese GF2 respecively respectively. Right: The first column until the end show the registration results of SIFT, SURF, CPD, GLMDTPS, ZGL_CATE and Ours. For each method, the first row shows 5 × 5 checkboard and the second row shows the transformed image I_t.

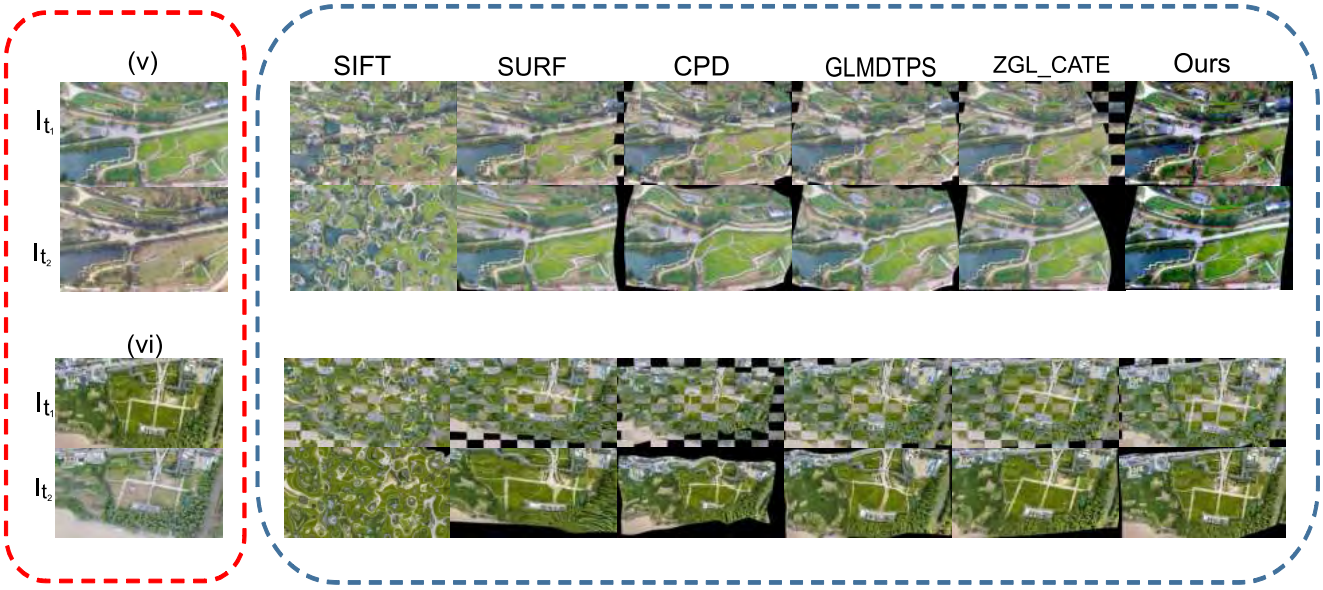

FIGURE 9. Registration examples on two typical image pairs from dataset (III). (v) Sichuan; (vi) GuiZhou. Left: Image pair I_{t_1} and I_{t_2} acquired over the same geographical area at two different time t_1 and t_2 by small UAV. Right: The first column until the end show the registration results of SIFT, SURF, CPD, GLMDTPS, ZGL_CATE and Ours. For each method, the first row shows 5 × 5 checkboard and the second row shows the transformed image I_t.

Moreover, SSFA and Semi_FCM can achieve better performance. Since SSFA employs the slow feature analysis (SFA) algorithm to extract the most temporally invariant component from the multi-temporal images and transform the data into a new feature space, DI can be better generated. The compared methods in terms of PCA_Kmeans and SSFA, Semi_FCM

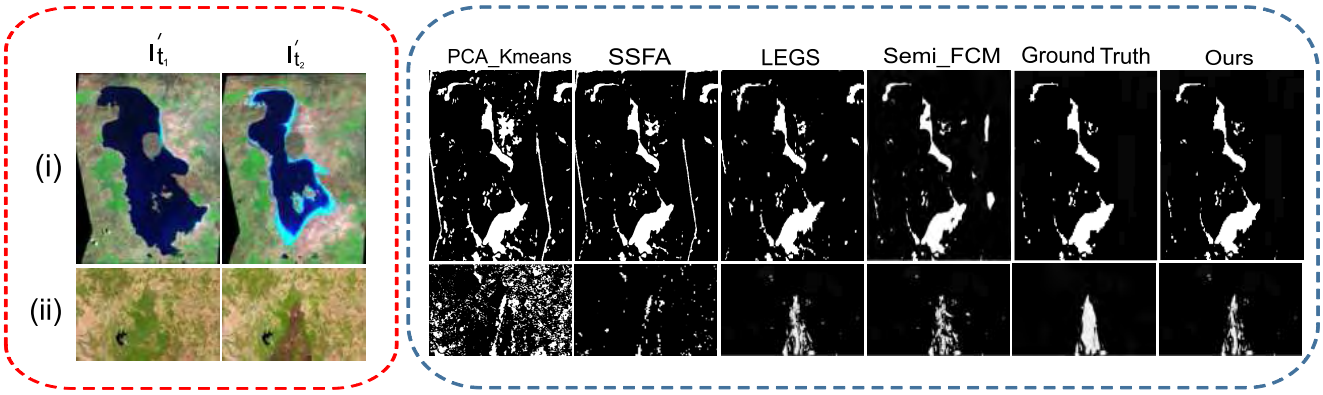

FIGURE 10. Change detection examples on two typical image pairs from dataset (I). (i) LakeOroumeih; (ii) Bastrop. (i) Yanan, Sichuan Province; (ii) Ansun, Guizhou Province. Left: I'_{t_1} and I'_{t_2} is the division of I_t and I_{t_2} according to a certain ratio. Right: The first column until the end show the change detection results of PCK_Kmeans, SSFA, Ground Turth and Ours. (i) PCK_Kmeans (TP:23; FP:8; FN:7; Precision:76.7%; Recall: 75.2%), SSFA (TP:25; FP:5; FN:5; Precision:83.3%; Recall: 83.3%), LEGS (TP:25; FP:3; FN:5; Precision:89.2%; Recall:83.3%), Semi_FCM (TP:24; FP:5; FN:6; Precision:82.7%; Recall: 80.0%), Ours (TP:28; FP:0; FN:2; Precision:93.3%; Recall: 100%). (ii) PCK_Kmeans (TP:22; FP:7; FN:8; Precision:73.3%; Recall: 75.9%), SSFA (TP:27; FP:4; FN:3; Precision: 90.0%; Recall: 87.1%), LEGS (TP:26; FP:3; FN:4; Precision:89.6%; Recall:86.7%), Semi_FCM (TP:28; FP:4; FN:2; Precision:87.5%; Recall: 93.3%), Ours (TP:29; FP:2; FN:1; Precision:96.7%; Recall: 93.4%).

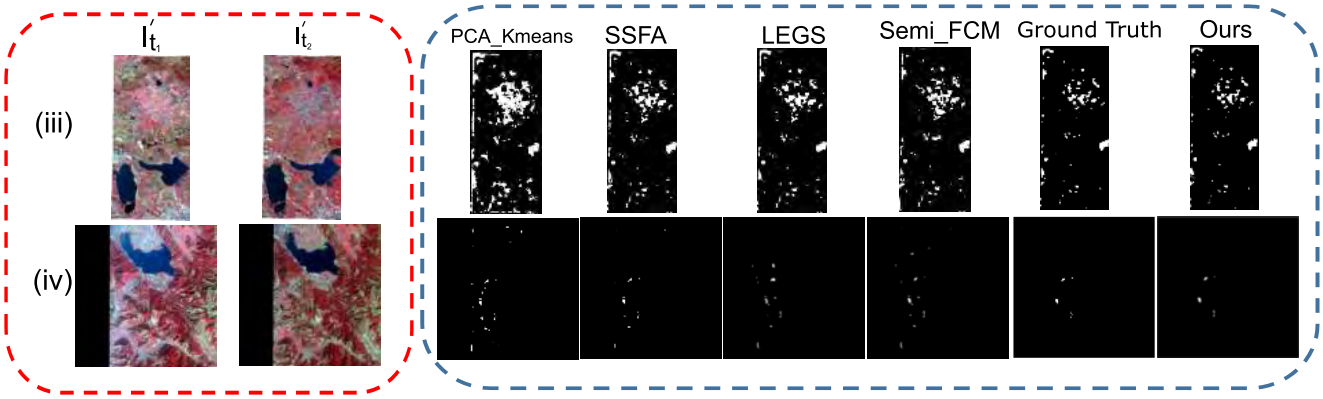

FIGURE 11. Change detection examples on two typical image pairs from dataset (II). (iii) Guizou; (iv) Hunan. Left: I'_{t_1} and I'_{t_2} is the division of I_t and I_{t_2} according to a certain ratio. Right: The first column until the end show the change detection results of PCK_Kmeans, SSFA, Ground Turth and Ours. (iii) PCK_Kmeans (TP:21; FP:5; FN:9; Precision:70.0%; Recall: 80.8%), SSFA (TP:24; FP:2; FN:6; Precision:80.0%; Recall: 92.3%), LEGS (TP:24; FP:2; FN:6; Precision:80.0%; Recall: 92.3%), Semi_FCM (TP:26; FP:4; FN:4; Precision:86.7%; Recall: 86.7%), Ours (TP:29; FP:1; FN:1; Precision:96.7%; Recall: 96.7%). (iv) PCK_Kmeans (TP:23; FP:3; FN:7; Precision:76.7%; Recall: 88.4%), SSFA (TP:25; FP:6; FN:5; Precision: 83.3%; Recall: 80.6%), LEGS (TP:26; FP:3; FN:4; Precision:89.6%; Recall:86.7%), Semi_FCM (TP:26; FP:4; FN:4; Precision:86.7%; Recall: 86.7%), Ours (TP:29; FP:2; FN:1; Precision:96.7%; Recall: 93.3%).

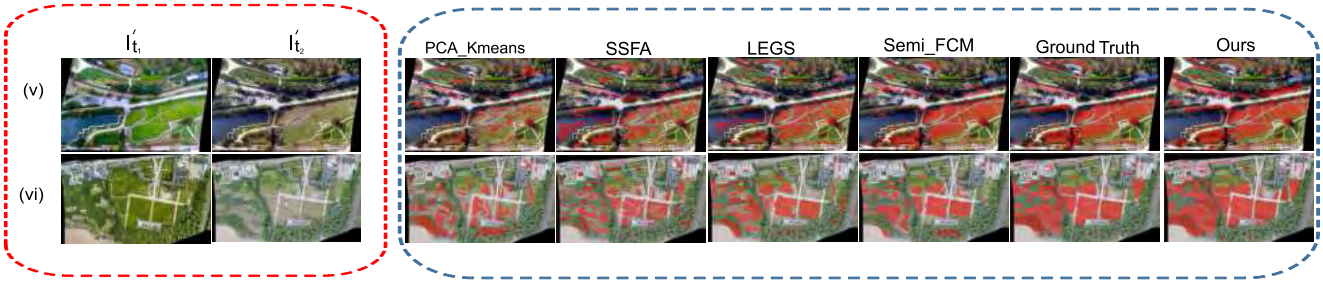

FIGURE 12. Change detection examples on two typical image pairs from dataset (III). (v) Sichuan; (vi)GuiZhou. Left: I'_{t_1} and I'_{t_2} is the division of I_t and I_{t_2} according to a certain ratio. Right: The first column until the end show the change detection results of PCK_Kmeans, SSFA, Ground Turth and Ours. (v) PCK_Kmeans (TP:19; FP:5; FN:11; Precision:63.3%; Recall: 79.1%), SSFA (TP:23; FP:2; FN:7; Precision:76.7%; Recall: 92.0%), LEGS (TP:27; FP:6; FN:3; Precision:81.8%; Recall: 90.0%), Semi_FCM (TP:27; FP:6; FN:3; Precision:81.8%; Recall: 90.0%), Ours (TP:29; FP:1; FN:1; Precision:96.7%; Recall: 96.7%). (vi) PCK_Kmeans (TP:20; FP:3; FN:10; Precision:66.7%; Recall: 86.9%), SSFA (TP:24; FP:5; FN:6; Precision: 80.0%; Recall: 77.4%), LEGS (TP:27; FP:6; FN:3; Precision:81.8%; Recall: 90.0%), Semi_FCM (TP:28; FP:6; FN:2; Precision:82.3%; Recall: 93.3%), Ours (TP:28; FP:2; FN:2; Precision:93.3%; Recall: 93.3%).

use semi-supervised fuzzy C-means filter the pseudolabels from the difference image. Since LEGS can effectively capture local contrast, texture and shape information for saliency detection, the complex relationship between different global saliency cues by local estimation and global search. Therefore, LEGS also achieve better performance.

IV. CONCLUSION

In this paper, a robust set of change detection framework for land cover change in mountainous terrain is proposed, which can detect multi-temporal remote sensing image pairs acquired by different type of sensors. The superiority of our framework can be summarized through three main contributions as follows: 1) a multi-scale feature description is generated using layers formed via a pretrained VGG network; 2) a gradually increasing selection of inliers is realized to estimate correspondence and transformations; 3) fuzzy C-means classifier is adopted to generate a similarity matrix between image pair of geometric correction, deep neural networks (DNNs) are applied to directly create a change detection map from pre-processed image pair by bypassing the steps of filtering or generating a difference image (DI). The proposed framework can provide a stable change rule for monitoring land cover change from multi-temporal data. For the purpose of experimental evaluation, dataset was mainly collected in the ten key land conservation regions of Sichuan, Guizhou and Hunan, China. Compared with five state-of-the-art registration methods and four state-of-the-art change detection methods, our method shows better performances in most cases.

Future studies will be conducted in two directions: (i) thematic applications of land cover changes, such as cultivated land changes; (ii) different sourcing images, such as image pair of combination of UAV image and satellite remote sensing image. Indeed, combining different sourcing images will identify more regions of change in many other typical regions with various land cover types.

ACKNOWLEDGMENT

We are grateful to David G. Lowe, Herbert Bay, Andriy Myronenko, Turgay Celik, Chen Wu, Lijun Wang and Pan Shao for providing their implementation source codes and test data sets. This greatly facilitated the comparison experiments. *(Fei Song and Zhuoqian Yang contributed equally to this work.)*

REFERENCES

[1] Y. Wang, F. Zhao, L. Cheng, and K. Yang, "Framework for monitoring the conversion of cultivated land to construction land using SAR image time series," *Remote Sens. Lett.*, vol. 6, no. 10, pp. 794–803, 2015.

[2] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[3] Y. Wu, S. Li, and S. Yu, "Monitoring urban expansion and its effects on land use and land cover changes in Guangzhou city, China," *Environ. Monitor. Assessment*, vol. 188, no. 1, p. 54, 2016.

[4] L. Wang, H. Lu, R. Xiang, and M.-H. Yang, "Deep networks for saliency detection via local estimation and global search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3183–3192.

[5] P. Shao, W. Shi, P. He, M. Hao, and X. Zhang, "Novel approach to unsupervised change detection based on a robust semi-supervised FCM clustering algorithm," *Remote Sens.*, vol. 8, no. 3, p. 264, 2016.

[6] S. A. Azzouzi, A. Vidal-Pantaleoni, and H. A. Bentounes, "Desertification monitoring in Biskra, Algeria, with Landsat imagery by means of supervised classification and change detection methods," *IEEE Access*, vol. 5, pp. 9065–9072, 2017.

[7] K. Yang, Z. Yu, Y. Luo, Y. Yang, L. Zhao, and X. Zhou, "Spatial and temporal variations in the relationship between lake water surface temperatures and water quality—A case study of Dianchi Lake," *Sci. Total Environ.*, vol. 624, pp. 859–871, May 2018.

[8] J. D. T. De Alban, G. M. Connette, P. Oswald, and E. L. Webb, "Combined Landsat and L-band SAR data improves land cover classification and change detection in dynamic tropical landscapes," *Remote Sens.*, vol. 10, no. 2, p. 306, 2018.

[9] Z. Wei *et al.*, "A small UAV based multi-temporal image registration for dynamic agricultural terrace monitoring," *Remote Sens.*, vol. 9, no. 9, p. 904, 2017.

[10] K. Yang, A. Pan, Y. Yang, S. Zhang, S. H. Ong, and H. Tang, "Remote sensing image registration using multiple image features," *Remote Sens.*, vol. 9, no. 6, p. 581, 2017.

[11] A. S. Milas, K. Arend, C. Mayer, M. A. Simonson, and S. Mackey, "Different colours of shadows: Classification of UAV images," *Int. J. Remote Sens.*, vol. 38, nos. 8–10, pp. 3084–3100, 2017.

[12] G. Li and Y. Yu, "Visual saliency detection based on multiscale deep CNN features," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5012–5024, Nov. 2016.

[13] Z. Lv, W. Shi, X. Zhou, and J. A. Benediktsson, "Semi-automatic system for land cover change detection using bi-temporal remote sensing images," *Remote Sens.*, vol. 9, no. 11, p. 1112, 2017.

[14] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and *k*-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.

[15] 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.

[16] H. Lyu, H. Lu, and L. Mou, "Learning a transferable change rule from a recurrent neural network for land cover change detection," *Remote Sens.*, vol. 8, no. 6, p. 506, 2016.

[17] H. Zhang, M. Gong, P. Zhang, L. Su, and J. Shi, "Feature-level change detection using deep representation and feature change analysis for multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1666–1670, Nov. 2016.

[18] E. M. de Oliveira Silveira, J. M. de Mello, F. W. Acerbi, Jr., and L. M. T. de Carvalho, "Object-based land-cover change detection applied to Brazilian seasonal savannahs using geostatistical features," *Int. J. Remote Sens.*, vol. 39, no. 8, pp. 2597–2619, 2018.

[19] R. Xiao, R. Cui, M. Lin, L. Chen, Y. Ni, and X. Lin, "SOMDNCD: Image change detection based on self-organizing maps and deep neural networks," *IEEE Access*, vol. 6, pp. 35915–35925, 2018.

[20] B. Uamkasem, H. L. Chao, and B. Jiantao, "Regional land use dynamic monitoring using Chinese GF high resolution satellite data," in *Proc. Int. Conf. Appl. Syst. Innov.*, 2017, pp. 838–841.

[21] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.

[22] L. Greengard and J. Strain, "The fast gauss transform," *SIAM J. Sci. Stat. Comput.*, vol. 12, no. 1, pp. 79–94, 2006.

[23] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, no. 4, pp. 891–909, Apr. 2008.

[24] Y. Yang, S. H. Ong, and K. W. C. Foong, "A robust global and local mixture distance based non-rigid point set registration," *Pattern Recognit.*, vol. 48, no. 1, pp. 156–173, 2015.

[25] J. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 53–64, Jan. 2016.

[26] K. Yang *et al.*, "Quake warning funds on shaky ground," *Science*, vol. 358, no. 6368, p. 1263, 2017.

[27] S. Zhang, Y. Yang, K. Yang, Y. Luo, and S. H. Ong, "Point set registration with global-local correspondence and transformation estimation," in *Proc. Int. Conf. Comput. Vis.*, 2017, pp. 2688–2696.

[28] S. Zhang, K. Yang, Y. Yang, and Y. Luo, "Nonrigid image registration for low-altitude SUAV images with large viewpoint changes," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 592–596, Apr. 2018.

[29] S. Zhang, K. Yang, Y. Yang, Y. Luo, and Z. Wei, "Non-rigid point set registration using dual-feature finite mixture model and global-local structural preservation," *Pattern Recognit.*, vol. 80, pp. 183–195, Aug. 2018.

[30] F. Song, M. Li, Y. Yang, K. Yang, X. Gao, and T. Dan, "Small UAV based multi-viewpoint image registration for monitoring cultivated land changes in mountainous terrain," *Int. J. Remote Sens.*, vol. 39, no. 21, pp. 7201–7224, 2018.

[31] T, Dan *et al.*, "Multifeature energy optimization framework and parameter adjustment-based nonrigid point set registration," *J. Appl. Remote Sens.*, vol. 12, no. 3, pp. 12–27, 2018.
[32] P. A. Permatasari, A. Fatikhunnada, Liyantono, Y. Setiawan, Syartinilia, and A. Nurdiana, "Analysis of agricultural land use changes in Jombang Regency, East Java, Indonesia using BFAST method," *Procedia Environ. Sci.*, vol. 33, pp. 27–35, Apr. 2016.
[33] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.
[34] J. Bohg and D. Kragic, "Learning grasping points with shape context," *Robot. Auto. Syst.*, vol. 58, no. 4, pp. 362–377, 2010.
[35] Y. Gu, K. Ren, P. Wang, and G. Gu, "Polynomial fitting-based shape matching algorithm for multi-sensors remote sensing images," *Infr. Phys. Technol.*, vol. 76, pp. 386–392, May 2016.
[36] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, no. 4, pp. 325–340, Nov. 1987.
[37] A. L. Yuille and N. M. Grzywacz, "A mathematical analysis of the motion coherence theory," *Int. J. Comput. Vis.*, vol. 3, no. 2, pp. 155–175, 1989.
[38] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.
[39] J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized Gaussian fields criterion," *Pattern Recognit.*, vol. 48, no. 3, pp. 772–784, 2015.
[40] S. Ji and S. Peng, "Terminal perturbation method for the backward approach to continuous time mean–variance portfolio selection," *Stochastic Process. Appl.*, vol. 118, no. 6, pp. 952–967, 2008.
[41] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, Jun. 1989.
[42] G. E. Hinton, *Training Products of Experts by Minimizing Contrastive Divergence*. Cambridge, MA, USA: MIT Press, 2002.
[43] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," *Momentum*, vol. 9, no. 1, pp. 599–619, 2010.
[44] A. Lozano-Diez, R. Zazo, D. T. Toledano, and J. Gonzalez-Rodriguez, "An analysis of the influence of deep neural network (DNN) topology in bottleneck feature based language recognition," *PLoS ONE*, vol. 12, no. 8, p. e0182580, 2017.
[45] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[46] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

FEI SONG received the B.S. degree from Sichuan Normal University, China, in 2013. He is currently pursuing the M.S. degree with the School of Information Science and Technology, Yunnan Normal University. His current research interests include image registration, point set registration, and change detection.

ZHUOQIAN YANG is currently pursuing the B.S. degree with the College of Software, Beihang University. His research interest includes computer vision and image registration.

XUEYAN GAO received the B.S. degree from Henan Normal University, China, in 2016. She is currently pursuing the M.S. degree with the School of Information Science and Technology, Yunnan Normal University. Her current research interests include image registration, point set registration, pattern recognition, and change detection.

TINGTING DAN received the B.S. degree from China West Normal University, China, in 2016. She is currently pursuing the M.S. degree with the School of Information Science and Technology, Yunnan Normal University. Her current research interests include image registration, point set registration, pattern recognition, and change detection.

YANG YANG received the master's degree from Waseda University, Japan, in 2007, and the Ph.D. degree from the National University of Singapore, Singapore, in 2013. He is currently an Associate Professor with the School of Information Science and Technology, Yunnan Normal University. His research interest covers image registration, remote sensing image processing, medical image processing, geography information system, and human masticatory system.

WANJING ZHAO received the B.S. degree from China Yunnan Normal University, China, in 2018. She is currently pursuing the M.S. degree with the School of Information Science and Technology, Yunnan Normal University. Her current research interests include image registration, point set registration, and pattern recognition.

RUI YU received the B.S. degree from Harbin Huade University, China, in 2017. She is currently pursuing the M.S. degree with the School of Information Science and Technology, Yunnan Normal University. Her current research interests include image registration, point set registration, and pattern recognition.