SALIENCY-GUIDED CHANGE DETECTION OF REMOTELY SENSED IMAGES USING RANDOM FOREST

Wenqing Feng[1], Haigang Sui[1,2] *, Xu Chen[1]

[1] State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, P.R. China
(corresponding author, e-mail: haigang_sui@263.net)
[2] Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan, P.R. China

Commission III, ICWG III/Iva

KEY WORDS: Remote Sensing, Change Detection, Segmentation, Super-pixel, Saliency, Random Forest

ABSTRACT:

Studies based on object-based image analysis (OBIA) representing the paradigm shift in change detection (CD) have achieved remarkable progress in the last decade. Their aim has been developing more intelligent interpretation analysis methods in the future. The prediction effect and performance stability of random forest (RF), as a new kind of machine learning algorithm, are better than many single predictors and integrated forecasting method. In this paper, we present a novel CD approach for high-resolution remote sensing images, which incorporates visual saliency and RF. First, highly homogeneous and compact image super-pixels are generated using super-pixel segmentation, and the optimal segmentation result is obtained through image superimposition and principal component analysis (PCA). Second, saliency detection is used to guide the search of interest regions in the initial difference image obtained via the improved robust change vector analysis (RCVA) algorithm. The salient regions within the difference image that correspond to the binarized saliency map are extracted, and the regions are subject to the fuzzy c-means (FCM) clustering to obtain the pixel-level pre-classification result, which can be used as a prerequisite for superpixel-based analysis. Third, on the basis of the optimal segmentation and pixel-level pre-classification results, different super-pixel change possibilities are calculated. Furthermore, the changed and unchanged super-pixels that serve as the training samples are automatically selected. The spectral features and Gabor features of each super-pixel are extracted. Finally, superpixel-based CD is implemented by applying RF based on these samples. Experimental results on Ziyuan 3 (ZY3) multi-spectral images show that the proposed method outperforms the compared methods in the accuracy of CD, and also confirm the feasibility and effectiveness of the proposed approach.

1. INTRODUCTION

Change detection (CD) is an important research topic that leverages quantitative analysis of multi-temporal remotely sensed images to determine the process of land cover change, especially in the monitoring of building land, urban development and disaster assessment (Hazel 2001; Hussain et al. 2013). Along with the rapid development of remotely sensed image acquisition means and the gradual shortening of the acquisition cycle, the scope of its applications is becoming increasingly widespread and the application demand is expanding. This presents higher requirements and challenges for CD technology.

With the improvement in resolution, the internal spectral difference of the same pairs of similar features increases gradually. The automatic CD technology based on pixel spectral statistics is not able to meet the requirement of the extraction of change information and becomes the main obstacle for the widespread application of high resolution remotely sensed images. The emergence of object-oriented technology for high-resolution remote sensing image analysis provides a new way of thinking, and the basic unit of CD has also transformed from pixel to object (Hazel 2001). Since the object-based change detection (OBCD) approach has more advantages than the pixel-based change detection (PBCD) approach, it has received extensive attention and been developed in recent years (Wang, Zhao and Zhu 2007; Emary et al. 2010; Wang and Xu et al. 2013; Hao and Shi et al. 2016; Xiao and Zhang et al. 2016;

Xiao and Yuan et al. 2017). The object is defined as a single homogeneous region with shape and spectral properties. Each object has features such as spectrum, shape, texture and context. Therefore, in the process of CD, we can take full advantage of spectral features and combine other features to improve the CD accuracy. In this field, the most commonly used methods are object-based change vector analysis (OCVA), object-based correlation coefficient (OCC), object-based chi square transformation (OCST) (Wang, Yan and Wang 2014) etc. These methods take advantage of the various features of the object, and incorporate them into analyses in later stage. They compared the methods using only a single feature, can significantly improve the accuracy of CD. The performances of these methods, however, heavily rely on the quality of feature selection, the allocation of feature weights and the determination of the change threshold. Moreover, due to the difficulty to determine the segmentation scale, it is likely to introduce uncertainty in the CD process, and reduce the reliability of the detection results. In order to obtain a better result, the segmentation scale, feature extraction, the change of threshold and many other factors need to be taken into consideration. The future trend of CD is the automation and intellectualization of the analysis process. Although a large number of CD methods and theoretical models are proposed from the object-based or pixel-based perspective, or for different application purposes, there are still many uncertainties. The combination of object-based and pixel-based CD approaches helps to reduce the uncertainty (Aguirre et al. 2011; Lu and Li et al. 2015; Xiao and Zhang et al. 2016; Feng and Sui

---

* Corresponding author

et al. 2017). The purpose of combining two image analysis approaches is to obtain better image analysis results. There are two kinds of usage of the combined approach, one is that the two methods are executed in parallel, and then integrated; the other is that the two methods are performed consecutively, which allows one method to obtain results as the premise for the execution of the other method. Both of these two strategies are for the purpose of achieving better image analysis results.

This paper adopts the strategy that performs the pixel-based and object-based approaches consecutively. In the pixel-level analysis process, the robust change vector analysis (RCVA) method is used to obtain the difference image, and the visual attention mechanism is introduced to find the regions that are most likely to change. Saliency, which is closely related to human visual perception, helps people understand the image (Zheng and Jiao et al. 2016; Wang and Yang et al. 2016; Hou and Wang et al. 2016). Saliency detection has significant effects on improving the efficiency of computer processing of image information and obtaining better processing results, which broadly benefits e.g. image segmentation, object recognition, and detection task (Zheng and Jiao et al. 2016; Wang and Yang et al. 2016; Li and Xu et al. 2017). Motivated by these advantages, we explore the saliency cue for CD from remotely sensed images, based on the assumption that changed regions have higher saliency than unchanged regions nearby in the local context. The strong visual contrast of local areas makes saliency suitable to guide the sample selection for the object-level CD. This paper makes full use of the advantages of pixel-based and object-based analysis method and combines it with RF model to analyse the influence of saliency, the influence of sample selection and feature extraction on the performance of the final classifier. The entire workflow of the proposed method is shown in Figure 1.
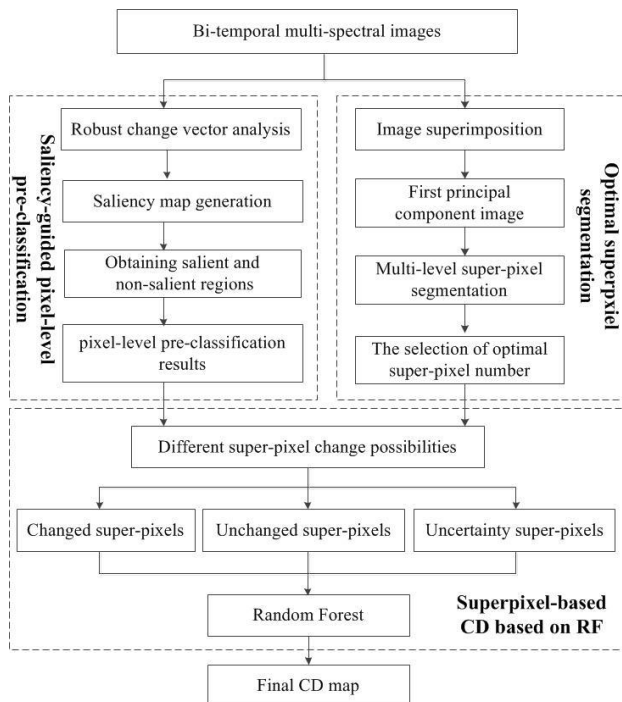

Figure 1. Flowchart of the proposed approach

The rest of this paper is organized as follows. Section 2 describes the proposed method. Section 3 presents the experimental results and discussion. Finally, we conclude this paper in Section 4.

2. METHODOLOGY

2.1 Optimal Super-pixel Segmentation

We exploit the entropy-rate segmentation algorithm (Liu, Tuzel, Ramalingam and Chellappa 2011) to segment the image into many super-pixel regions for subsequent information extraction. Further details of the entropy-rate segmentation algorithm are given by Liu et al (2011). The purpose of super-pixel segmentation is to segment the same type of features in order to obtain a series of homogeneous regions with compact and strong regional consistency. However, in the process of segmentation, the selection of the number of super-pixels plays an important role in image segmentation, and it is also the key to improve the image segmentation quality.

For multi-level and super-pixel segmentation in object-oriented remote sensing information extraction methods, the definition of optimal number of super-pixel could be expressed by one or several super-pixels. This requires that the super-pixel size should be close to the target feature, and its geometry (polygon) should not be too broken; the boundaries of the super-pixel should be clear; the heterogeneity within the super-pixel should be as small as possible; the heterogeneity among the different types of super-pixels should be as large as possible; and the super-pixel could express the basic characteristics of a certain object. The homogeneity of super-pixels guarantees the purity, while the heterogeneity between super-pixels ensures the separability. In our study, the weighted variance of the super-pixels is used to express the internal homogeneity, and Moran's I index (Espindola, Camara, Reis and Bins 2006) is used to represent the heterogeneity between the super-pixels. The formula is as follows:

$$\mathbf{H} = \frac{\sum\limits_{k=1}^{n} a_k v_k}{\sum\limits_{k=1}^{n} a_k} \tag{1}$$

$$\mathbf{MI} = \frac{n \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} w_{ij}(y_i - \overline{y})(y_j - \overline{y})}{(\sum\limits_{i=1}^{n}(y_i - \overline{y})^2)(\sum\limits_{i \neq j} \sum w_{ij})} \tag{2}$$

In equation (1), $\mathbf{H}$ is the homogeneity index, a_k is the area size of the super-pixel k which is expressed as the number of pixels inside the super-pixel, v_k is the standard deviation of the super-pixel k, n is the total number of super-pixels in the image segmentation. The equation represents the process of assigning weights to areas; the weights can reduce the small super-pixel-induced instability. The higher the value of $\mathbf{H}$, the higher homogeneity of the super-pixels. In equation (2), $\mathbf{MI}$ is the heterogeneity index, w_{ij} indicates whether the super-pixel i and the super-pixel j are adjacent; if $w_{ij} = 1$, they are adjacent; if $w_{ij} = 0$, they are not adjacent. The value of y_i represents the average grey value of the super-pixel i; the value of y_j represents the average grey value of the super-pixel j; and the

value of $\bar{y}$ represents the average grey value of the entire image. The smaller the value of **MI**, the lower the correlation between the super-pixels, and the more clearly of the boundary between the super-pixels.

This paper adopts formula (3) to express the evaluation index of the optimal number of super-pixel (Espindola, Camara, Reis and Bins 2006). It uses the homogeneity index and heterogeneity index of super-pixel to construct a function to measure the quality of the segmentation results. The formula is as follows:

$$F(\mathbf{H}, \mathbf{MI}) = (1 - \rho)F(\mathbf{H}) + \rho F(\mathbf{MI}) \qquad (3)$$

In equation (3), $F(\mathbf{H})$ denotes the homogeneity evaluation index, $F(\mathbf{MI})$ denotes the heterogeneity evaluation index, ρ denotes the heterogeneity weight, $\rho \in [0,1]$, and the homogeneity weight is $1 - \rho$. In our study, the value of ρ is 0.5. The homogeneity index and heterogeneity index should be normalized before being used for evaluating the optimal number of super-pixels. The formula is as follows:

$$F(\mathbf{H}) = \frac{\mathbf{H}_{\max} - \mathbf{H}}{\mathbf{H}_{\max} - \mathbf{H}_{\min}} \qquad (4)$$

$$F(\mathbf{MI}) = \frac{\mathbf{MI}_{\max} - \mathbf{MI}}{\mathbf{MI}_{\max} - \mathbf{MI}_{\min}} \qquad (5)$$

On this basis, an optimal model of the number of super-pixels selection could be obtained by cubic spline function interpolation:

$$s_3(x) = \alpha_0 + \alpha_1 x + \frac{\alpha_2}{2!} x^2 + \frac{\alpha_3}{3!} x^3 + \sum_{j=1}^{n-1} \frac{\beta_j}{3!} (x - x_j)_+^3 \qquad (6)$$

In equation (6), when the function $s_3(x)$ takes the maximum value in the super-pixel interval $[x_{\min}, x_{\max}]$, the corresponding super-pixel number x is the optimal super-pixel number.

2.2 Saliency-guided Pixel-level Pre-classification

2.2.1 Initial Difference Image Generation: The traditional robust change vector analysis (RCVA) algorithm, which was proposed by Thonfeld et al. (2016), only considers the spectral difference in a single band and computes the least difference of each pixel. It does not incorporate the whole spectral difference of the corresponding pixels. To solve the aforementioned problems, we proposed an improved RCVA algorithm to reduce the influence of registration errors.

The improved RCVA algorithm is used for the analysis – not only on the pixels $x_1(i, j)$ in T1 phase image and $x_2(i, j)$ in T2 phase image, but also on the pixels in the adjacent neighborhood including diagonal neighbors. It is based on the assumption that a pixel $x_2(i \pm w, j \pm w)$ showing the least spectral variance to $x_1(i, j)$ is the pixel containing most of the corresponding ground information of $x_1(i, j)$. That is, if the bi-temporal images have geometric registration error, and

providing that the difference between a certain pixel in T1 phase image and the other pixel in T2 phase image is the smallest, then the two pixels are called corresponding image points. So the influence of the registration error can be effectively reduced. The improved RCVA algorithm uses a moving window with the size of $2w+1$ to incorporate the spectral difference of adjacent pixels, where w refers to the neighborhood expressed as number of pixels from the pixels of concern. In our study, we use the value of $w=2$ resulting in a 5×5 moving window. The calculation process is divided into two steps. In the first step, differencing images are calculated considering pixel neighborhood for subtracting T1 from T2 and vice versa. The specific formula is shown in equations (7) and (8).

$$x_{\text{diff}_a}(i, j) = \min_{(p \in [i-w, i+w], q \in [j-w, j+w])} \left\{ \sqrt{\sum_{k=1}^{n} (x_2^k(i, j) - x_1^k(p, q))^2} \right\} \quad (7)$$

$$x_{\text{diff}_b}(i, j) = \min_{(p \in [i-w, i+w], q \in [j-w, j+w])} \left\{ \sqrt{\sum_{k=1}^{n} (x_1^k(i, j) - x_2^k(p, q))^2} \right\} \quad (8)$$

The second step is to obtain the change intensity image incorporating the neighborhood information by the equation (9). After the acquisition of change intensity image, it is important to distinguish the real change regions and pseudo regions accurately.

$$m(i, j) = \begin{cases} x_{\text{diff}_b}(i, j), & x_{\text{diff}_a}(i, j) \geq x_{\text{diff}_b}(i, j) \\ x_{\text{diff}_a}(i, j), & x_{\text{diff}_a}(i, j) < x_{\text{diff}_b}(i, j) \end{cases} \qquad (9)$$

2.2.2 From Saliency to Pixel-level Pre-classification: For the difference image, the user is only interested in some parts of it. The salient region in difference image which attracts the users could also best deliver image content. Saliency is used to extract regions that are distinct from local and global regions. After acquiring the change intensity image, the problem of CD can be seen as finding a region with rigor distinctions with other regions. In the change intensity image, the changed region has a larger pixel value, but the pixel value of unchanged region is close to zero. Furthermore, the area of the changed regions is usually smaller than the unchanged region. Therefore, the changed region can be further highlighted by the graph-based visual saliency (GBVS) (Schölkopf, Platt and Hofmann 2007) model as a target with high contrast. From this point of view, CD and saliency detection problems are coincided in nature. And regarding visual performance, the changed areas also correspond to the salient regions of the change intensity image.

By using a thresholding method, the pixels can be preserved in the extracted areas when their values are larger than a given threshold T_{saliency}, otherwise the pixels are neglected. The thresholding function is as follows:

$$\mathbf{S}_{\text{Map}} = \begin{cases} 1, & \mathbf{S} \geq T_{\text{saliency}} \\ 0, & \text{otherwise} \end{cases} \qquad (10)$$

Let $\mathbf{S}_{\text{Map}}$ be a thresholding map, where "1" indicated that the corresponding pixel is preserved in the extracted areas and "0" stands for neglection. With the usage of a thresholding method on the saliency map S, the interest areas with discriminative information are well preserved and the false-changes generated

via the noise are greatly neglected. In our study, the regions within the change intensity image that correspond to the thresholding map are extracted, and the regions are subject to the fuzzy c-means (FCM) clustering to obtain the pixel-level pre-classification result.

2.3 Superpixel-based CD based on RF

2.3.1 Different Super-pixel Change Possibilities: The sample selection is conducted based on the results from optimal segmentation and pixel-level pre-classification. For the *i-th* super-pixel R_i, its uncertainty index T is calculated based on the pixel-level pre-classification result, as shown in the following equation (11).

$$T = \begin{cases} \dfrac{n_c}{n} & n_c \geq n_u \\ -\dfrac{n_u}{n} & n_c < n_u \end{cases} \quad (11)$$

Where n_c, n_u, and n are the changed, unchanged and total numbers of pixels in the super-pixel R_i respectively. After setting the threshold T_m, we then use the following equation (12) to determine the properties of the super-pixel R_i:

$$l_i = \begin{cases} 1 & T < -T_m \\ 2 & -T_m \leq T \leq T_m \\ 3 & T > T_m \end{cases} \quad (12)$$

Where $l_i = 1, 2, 3$ indicate that the attributes of the super-pixel R_i are non-change, uncertain and change respectively, and the range of the threshold is $T_m \in [0.5, 1)$. The changed and unchanged super-pixels are selected as training samples for RF, and the uncertain super-pixels will be further classified. With regard to the selection of the threshold T_m, we calculate the average accuracy after randomly selecting samples repeated 20 times in the interval [0.5, 1), with the step size of 0.05. When the CD accuracy is the best, the corresponding threshold T_m is the best and selected.

2.3.2 Classification based on RF: The key step after sample selection is the change feature extraction and analysis. In our study, after segmentation, the spectral and Gabor features of the super-pixels at the same position are extracted from the T1 and T2 phase image respectively. Among them, the spectral features (SF) mainly include the mean value, standard deviation, ratio, maximum value and minimum value, as shown in equations (13) and (14).

$$SF_{T1} = (Mean_{T1}, Std_{T1}, Ratio_{T1}, Minvalue_{T1}, Maxvalue_{T1}) \quad (13)$$
$$SF_{T2} = (Mean_{T2}, Std_{T2}, Ratio_{T2}, Minvalue_{T2}, Maxvalue_{T2}) \quad (14)$$

The Gabor wavelet function is obtained by Fourier transform of the Gaussian function, which can extract the image correlation feature in different scales and directions (Daugman 1988; Li, Shi, Zhang and Hao 2017). In our study, Gabor wavelet is employed to extract the texture features. The two-dimensional

Gabor function $F_\varphi(x, y)$ is expressed as:

$$F_\varphi(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j\omega x \right] \quad (15)$$

In equation (15), ω is the copy frequency of the Gaussian function and $\omega = 8$, $j = \sqrt{-1}$ in our study. σ_x and σ_y are the variance of the Gabor wavelet basis function in x- and y- axes respectively, and $\sigma_x = 1$ and $\sigma_y = 2$. In case a two-dimensional convolution is operated towards the experimental images $\mathbf{I}(x, y)$ and $F_\varphi(x, y)$, the real part of the result is calculated by equation (16).

$$\mathbf{G}_\varphi(x, y) = \mathrm{Re}\left\{ I(x, y) * F_\varphi(x, y) \right\} \quad (16)$$

Where $\mathbf{G}_\varphi(x, y)$ is the extracted feature image from the original image $\mathbf{I}(x, y)$ after Gabor filtering, and $*$ is the two-dimensional convolution operation. The mean value and standard deviation at different bands of each super-pixel in the Gabor feature image are extracted as the texture features.

The spectral features and Gabor features can reflect different object type information from different angles, and often complement each other. After extracting the above two types of features, we need to combine them as the characteristic input data for the RF model training. The category of uncertain super-pixels will be further predicted by the well trained RF model, and the voting result of all the decision trees is taken as the final classification result (Breiman 2001; Wessels and Bergh et al. 2016).

3. EXPERIMENTAL ANALYSES AND DISCUSSION

3.1 Dataset Description and Evaluation Metrics

The experimental dataset is the ZY3 multi-spectral images which are acquired in year 2014 and 2015 covering the city of Wuhan, China. ZY3 is China's first civilian high-resolution stereo mapping satellite, which was launched on January 2012. The orbit height of ZY-3 satellite is 505.984 km, and the orbit inclination is 97.421 °. The bi-temporal images are produced by national census geography of China and have been orthorectified. The pan and multi-spectral images are fused by Pan-sharp algorithm (Amro et al. 2011) and the spatial resolution of the fusion image is 2.1m. The bi-temporal images in the experiments mainly include three bands of R, G and B and the image size is 1564×1424. The main land cover types are vegetation, water, road, building, and bare land, etc. We can find that the change of the main features is the transformation between water and bare land, building and bare land, etc.
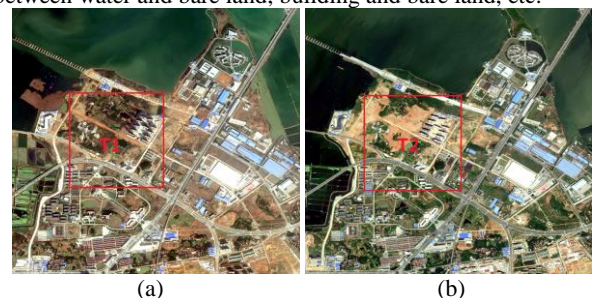

(a)   (b)

(c)   (d)

Figure2. The experimental dataset. (a) Image acquired on year 2014; (b) Image acquired on year 2015; (c) T1 zoomed image; (d) T2 zoomed image.

Accuracy assessment is important for understanding CD results and final decision making. Four indexes are used to evaluate the accuracy of the final results (Hou, Wang and Liu, 2016). The criteria are set as follows: the false alarms (FA) which is the number of unchanged pixels that are incorrectly detected as changed ones, i.e., N_{FA} and the false alarm rate in percentage is $R_{FA} = N_{FA} / N_0 \times 100\%$, where N_0 is the total number of unchanged pixels; the missed alarms (MA) which is the number of changed pixels that are incorrectly detected as unchanged ones, i.e., N_{MA} and the missed alarm rate in percentage is $R_{MA} = N_{MA} / N_1 \times 100\%$, where N_1 is the total number of changed pixels; the overall error (OE) which is the total number caused by FA and MA and the overall alarm rate in percentage is $R_{OE} = (N_{FA} + N_{MA}) / (N_0 + N_1) \times 100\%$; the kappa index that is a statistical measurement of accuracy or agreement and it reflects the consistency between experimental results and the ground truth. The Kappa index is expressed as $Kappa = (p_0 - p_c) / (1 - p_c)$, where p_0 indicates the real consistency and p_c indicates the theoretical consistency.

3.2 Results and Discussion

3.2.1 Parameter Setting: The ideal result of remote sensing image segmentation is that the interior of the obtained object has significant homogeneity, whereas the adjacent object has significant heterogeneity. To obtain the best super-pixel segmentation results, this paper uses the evaluation index to determine the final optimal number of super-pixel. In the segmentation experiment, we utilize the entropy rate segmentation algorithm to segment the first principal component image by changing the number of super-pixels to obtain the multi-level super-pixel regions with different sizes. The selected number range of super-pixels is increased from 2000 to 6000 and the step size is 200. The value of evaluation index F(**H,MI**) at each scale after segmentation is calculated, and the cubic spline interpolation is performed. The result of Figure 3 substantiates that F(**H,MI**) is the highest when the number of super-pixel is 3745. That is, at this time the inside of each super-pixel is homogeneous and different super-pixels are heterogeneous. This paper selects the number of 3745 as the optimal value of super-pixel segmentation.
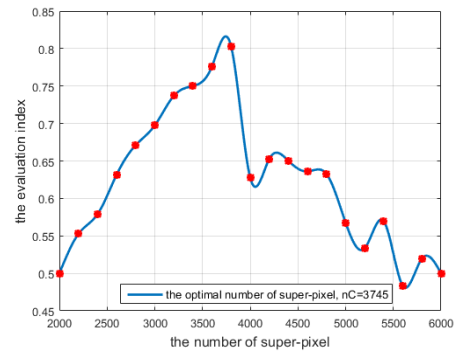

Figure 3. The evaluation index of the optimal super-pixel number

Since the saliency map highlights the regions with strong local contrast in vision, these regions draw people's attention mostly. There is a theoretical and visual commonality between saliency and CD. By setting thresholds for the saliency map, the regions within the change intensity image that correspond to the binarized saliency map are extracted. Then, the regions are subject to the FCM algorithm to obtain the pixel-level pre-classification result, which can be used as a prerequisite for object-oriented CD analysis. The pixel-level CD accuracy result is shown in Figure 4. When the threshold $T_{saliency} = 90$, the Kappa index is the largest. In the interval (45, 90], both the false alarm rate and the overall alarm rate are decreasing, while the missed alarm rate is increasing slowly. When the threshold is more than 90, the trends of the false alarm rate and the overall alarm rate have become flat, whereas the missed alarm rate is still rising. Therefore, the best threshold value is 90. Figure 5 shows the change intensity images and saliency maps of the experimental dataset, as well as the salient and non-salient regions extracted at the optimal saliency threshold.
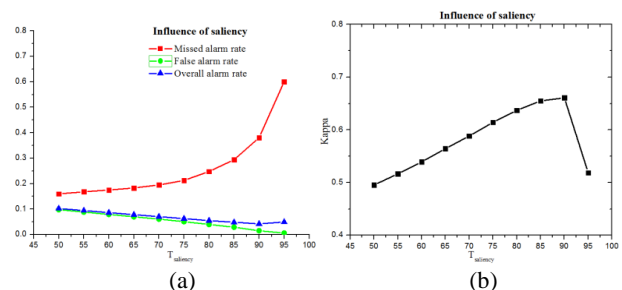

(a)   (b)

Figure 4. The influence of saliency on the experimental dataset. Influence of saliency (a) on FA, MA and OE; (b) on kappa index.

On the basis of the optimal segmentation and pixel-level pre-classification results, the changed and unchanged super-pixels that serve as the training samples are automatically selected. And the super-pixels' spectral and Gabor features are extracted as the characteristic input data for the RF model training. As shown in Figure 6, in the experimental dataset, when the threshold value $T_m = 0.75$, the Kappa index is the largest. When the threshold is in the interval [0.5,0.75], the missed alarm rate increases slowly while the false alarm rate and the overall alarm rate are declining; when the threshold T_m has exceeded 0.75, the missed alarm rate grows dramatically, while the false alarm rate and the overall alarm rate become flat, thus the best threshold is 0.75. Therefore, based on the above discussions, for the experimental dataset, $T_{saliency} = 90$ and $T_m = 0.75$ are adopted.
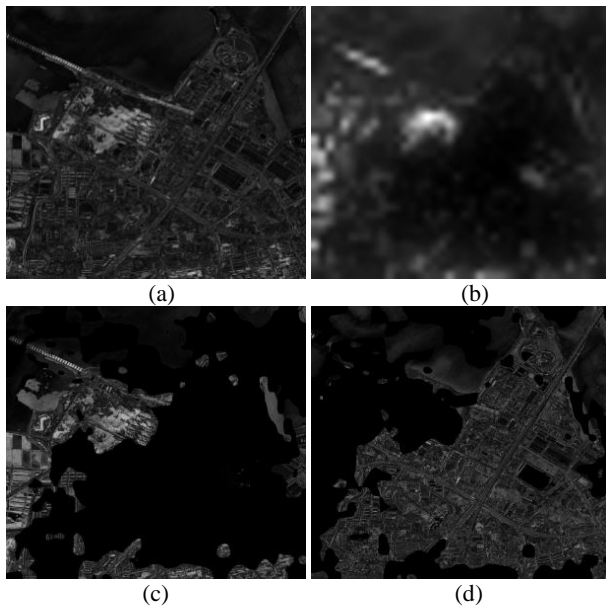
Figure 5. Intermediate results of saliency. (a) The change intensity image; (b) the saliency map; (c) the salient region; (d) the non-salient region.
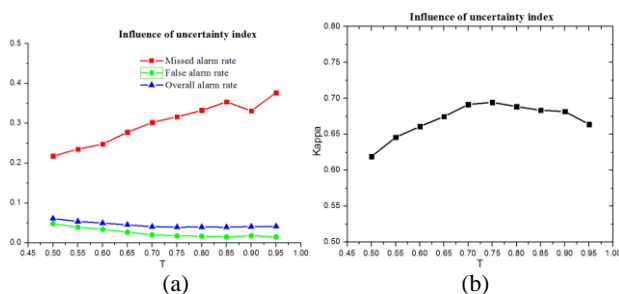

Figure 6. The influence of uncertainty index on the experimental dataset. Influence of uncertainty index (a) on FA, MA and OE; (b) on kappa index.

3.2.2 Results and Analysis: In order to verify the feasibility and effectiveness of the proposed approach, we challenge the high-resolution remote sensing image CD problem on ZY3 multi-spectral images. Typical correction, such as co-registration and relative radiometric correction are done for the three datasets before applying the proposed approach. And also, three pixel-based CD methods, such as Iterative Condition model (MRF-ICM), PCA-kMeans (Celik 2009), edge-based distance regularized level set evolution (DRLSE) (Li and Xu et al. 2010; Lei, Shi and Wu 2017), and object-based change vector analysis (OCVA) are selected as the comparison methods. These methods are performed to obtain best CD maps and are applied to demonstrate the superiority of the proposed approach based on the combination of visual saliency and RF. In MRF-ICM, the k-means algorithm is used to obtain the initial clustering value and the class number is $K = 2$, $\beta = 0.5$, and the maximum number of iterations is 30. In PCA-kMeans, $H = 5$ and $S = 5$ are adopted. H is the block size, and S is the dimensionality of the feature vectors. We set $\triangle t = 1.5$, $\mu = 0.04$, $\lambda = 5$ and $\alpha = -3.0$ for the DRLSE model. And we select the optimal super-pixel segmentation result to carry out the OCVA method. The OSTU algorithm (Otsu, 1979) is adopted for automatic threshold segmentation.

The results are demonstrated in two ways: the final CD results in figure form and the criteria in tabular form. The reference images are manually generated according to a detailed visual interpretation. The black areas indicate the unchanged regions, and the white areas indicate the changed regions. Figure 7 shows the experimental results carried out upon the experimental dataset. Table 1 lists the values for evaluation metrics. In the table, the results of the proposed method are written in bold.
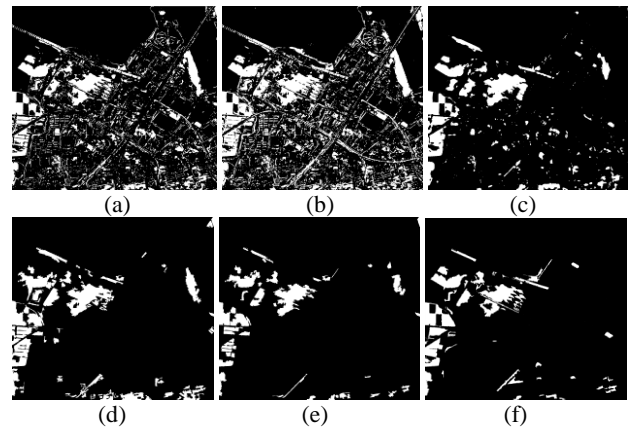

Figure 7. The results of different methods. (a) MRF-ICM; (b) PCA-kMeans; (c) DRLSE; (d) OCVA; (e) Proposed; (f) Reference image.

The CD accuracy of traditional pixel-level methods such as MRF-ICM, PCA-kMeans are comparatively low as shown in Table 1. This is mainly because only the spectral features of bi-temporal images are used in these algorithms. With regard to the time efficiency, the iterative optimization is needed in the MRF-ICM, thus it has lower time efficiency than that of PCA-kMeans. The DRLSE algorithm can involve neighborhood information between pixels in the difference image. It can be seen that compared to MRF-ICM, PCA-kMeans methods, the DRLSE can further improve the accuracy of CD result. However, it takes the longest time. This is because the adjacency restriction relation between the regions is taken into account in the evolutionary segmentation process. The result from using OCVA is not prominently better than the traditional pixel-level CD methods. This is caused by over dependence on the selection of threshold value and the grey mean information of the super-pixel while it fails to take advantage of grey distribution information, and it results the poor CD result. The proposed approach in our study obtains the best result and gains higher accuracy with respect to the comparison methods.

Overall, in the application of CD, we can combine the pixel-based and object-based image analysis approaches according to different purposes by using visual saliency and RF, to get the final object-level CD result. The changed objects are more regular, and are corresponding to actual geographical features. Therefore, the combination can not only absorb the advantages from both pixel-based and object-based approaches, but also can obtain the best accuracy.

Accuracy	MRF-ICM	PCA-kMeans	DRLSE	OCVA	Proposed
False Alarms	0.1225	0.1708	0.0398	0.0485	**0.0175**
Missed Alarms	0.1566	0.1083	0.1856	0.2204	**0.3180**
Overall Alarms	0.1250	0.1662	0.0505	0.0610	**0.0395**
Kappa index	0.4390	0.3706	0.6756	0.6192	**0.6955**

Table 1. Performance comparisons against different approaches.

4. CONCLUSION

A novel CD approach based on visual saliency and RF from multi-temporal high resolution remotely sensed images is proposed in our paper. Saliency detection is used to guide the search of interest regions in the initial difference image obtained via the improved RCVA algorithm, where the effect of noise can be reduced to some extent. On the basis of the optimal segmentation and pixel-level pre-classification results, the object-level CD is implemented by applying RF model. The bi-temporal ZY3 multi-spectral images are used to verify the effectiveness and show the superiority of the proposed approach. There still exist some works to further improve the performance of the proposed method. In view of the problem of high-resolution remote sensing image CD, because of the phenomenon of different objects which have the same spectrum and the same objects have different spectrum is more serious, by only using spectral feature and Gabor feature is still inadequate. In the future research, we can further introduce the edge features, geometric features, and elevation so as to further improve the accuracy of CD.

REFERENCES

Aguirre-Gutiérrez, J., Seijmonsbergen, A. C., and Duivenvoorden, J. F. 2012. Optimizing land cover classification accuracy for change detection, a combined pixel-based and object-based approach in a mountainous area in Mexico. *Applied Geography* 34(1), pp. 29-37.

Amro, I., Mateos, J., Vega, M., Molina, R., and Katsaggelos, A. K. 2011. A survey of classical methods and new trends in pansharpening of multispectral images. *Eurasip Journal on Advances in Signal Processing*, 2011(1), pp.79.

Breiman, L. 2001. Random forest. *Machine Learning* 45, pp. 5-32.

Celik, T. 2009. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE Geoscience & Remote Sensing Letters* 6(4): 772-776.

Daugman, J. G. 1988. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics Speech & Signal Processing* 36(7), pp. 1169-1179.

Emary, E., Mostafa, K., and Onsi, H. 2010. A proposed multi-scale approach with automatic scale selection for image change detection. *IEEE International Conference on Image Processing* , pp. 3149-3152.

Feng, W., Sui, H., Tu, J., Huang, W., and Sun. K. 2017. Remote Sensing Image Change Detection Based on the Combination of Pixel-level and Object-level Analysis. *Acta Geodaetica et Cartographica Sinica* 46(9), pp. 1147-1155.

G. M. Espindola, G. Camara, I. A. Reis, L. S. Bins, and A. M. Monteiro. 2006. Parameter selection for region‐growing image segmentation algorithms using spatial autocorrelation. *International Journal of Remote Sensing* 27(14), pp. 3035-3040.

Hussain, M., Chen, D., Cheng, A., Wei, H., and Stanley, D. 2013. Change detection from remotely sensed images: from pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry & Remote Sensing* 80(2), pp. 91-106.

Hazel, G. G. 2001. Object-level change detection in spectral imagery. *IEEE Transactions on Geoscience & Remote Sensing* 39(3), pp. 553-561.

Hao, M., Shi, W., Zhang, H., Wang, Q., and Deng, K. 2016. A scale-driven change detection method incorporating uncertainty analysis for remote sensing images. *Remote Sensing* 8(9), pp. 745.

Hou, B., Wang, Y., and Liu, Q. 2016. A saliency guided semi-supervised building change detection method for high resolution remote sensing images. *Sensors* 16(9), pp. 1377.

Li, C., Xu, C., Gui, C., and Fox, M. D. 2010. Distance regularized level set evolution and its application to image segmentation. *IEEE Transactions on Image Processing* 19(12), pp. 3243-3254.

Lu, J., Li, J., Chen, G., Zhao, L., Xiong, B., and Kuang, G. 2015. Improving pixel-based change detection accuracy using an object-based approach in multitemporal SAR flood images. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 8(7), pp. 3486-3496.

Li, E., Xu, S., Meng, W., and Zhang, X. 2017. Building extraction from remotely sensed images by integrating saliency cue. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 10(3), pp. 906-919.

Liu, M. Y., Tuzel, O., Ramalingam, S., and Chellappa, R. 2011. Entropy rate superpixel segmentation. *IEEE Computer Vision and Pattern Recognition*, pp. 2097-2104.

Li, Z., Shi, W., Zhang, H., and Hao, M. 2017. Change detection based on gabor wavelet features for very high resolution remote sensing images. *IEEE Geoscience & Remote Sensing Letters*, 99, pp. 1-5.

Lei, Y., Shi, J., and Wu, J. 2017. Region-driven distance regularized level set evolution for change detection in remote sensing images. *Multimedia Tools & Applications* , pp. 1-16.

Otsu, N. 1979. Threshold selection method from grey-level histograms. *IEEE Transactions on Systems Man and Cybernetics* 9(1), pp. 62-66.

Schölkopf, B., Platt, J., and Hofmann, T. 2007. Graph-based visual saliency. *Advances in Neural Information Processing Systems*, 19, pp. 545-552.

Thonfeld, F., Feilhauer, H., Braun, M., and Menz, G. 2016. Robust change vector analysis (RCVA) for multi-sensor very high resolution optical satellite data. *International Journal of Applied Earth Observation & Geo-information* , pp. *131-140*.

Wang, L., Yan, L. I., and Wang, Y. 2014. Research on land use change detection based on an object-oriented change vector analysis method. *Geographical Research* 27(2), pp. 74-80.

Wang, W. J., Zhao, Z. M., and Zhu, H. Q. 2009. Object-oriented change detection method based on multi-scale and multi-feature fusion. *Urban Remote Sensing Joint Event*, pp. 1-5.

Wang, X. 2013. Object-oriented change detection approach for high-resolution remote sensing images based on multi-scale fusion. *Journal of Applied Remote Sensing* 7(1), pp. 073696.

Wang, S., Yang, S., and Jiao, L. 2016. Saliency-guided change detection for SAR imagery using a semi-supervised laplacian SVM." *Remote Sensing Letters* 7(11), pp.1043-1052.

Wessels, K. J., Bergh, F. V. D., Roy, D. P., Salmon, B. P., Steenkamp, K. C., and Macalister, B. 2016. Remote sensing rapid land cover map updates using change detection and robust random forest classifiers. *Remote Sensing* 8(11).

Xiao, P., Zhang, X., Wang, D., Yuan, M., Feng, X., and Kelly, M. 2016. Change detection of built-up land: a framework of combining pixel-based detection and object-based recognition. *ISPRS Journal of Photogrammetry & Remote Sensing*, 119, pp. 402-414.

Xiao, P., Yuan, M., Zhang, X., Feng, X., and Guo, Y. 2017. Co-segmentation for object-based building change detection from high-resolution remotely sensed images. *IEEE Transactions on Geoscience & Remote Sensing*, 99, pp. 1-17.

Zheng, Y., Jiao, L., Liu, H., Zhang, X., Hou, B., and Wang, S. 2016. Unsupervised saliency-guided SAR image change detection. *Pattern Recognition* , 61, pp. 309-326.