Change Detection Based on Deep Features and Low Rank

Bin Hou, Yunhong Wang, *Senior Member, IEEE*, and Qingjie Liu, *Member, IEEE*

***Abstract*— In this letter, we address the problem of change detection for remote sensing images from the perspective of visual saliency computation. The proposed method incorporates low-rank-based saliency computation and deep feature representation. First, multilevel convolutional neural network (CNN) features are extracted for superpixels generated using SLIC, in which a fixed-size CNN feature can be formed to represent each superpixel. Then, low-rank decomposition is applied to the change features of the two input images to generate saliency maps that indicate change probabilities of each pixel. Finally, binarized change map can be obtained with a simple threshold. To deal with scale variations, a multiscale fusion strategy is employed to produce more reliable detection results. Extensive experiments on Google Earth and GF-2 images demonstrate the feasibility and effectiveness of the proposed method.**

***Index Terms*— Change detection (CD), convolutional neural network (CNN), low rank, remote sensing (RS), visual saliency.**

I. Introduction

CHANGE detection of remote sensing (RS) images plays an important role in earth observation, and it has been widely used in many practical applications such as disaster management, environment monitoring, and urban planning. With the rapid development of sensor technologies, the spatial resolution of RS images has been improved significantly, which provides fresh opportunities and great challenges for change detection (CD). Traditional CD methods mainly focus on pixel-wise multitemporal analysis by extracting spectral or textual features. They are commonly used for low- and medium-resolution imageries; however, they are hardly applicable for high-resolution imageries [1]. The object-based CD (OBCD) considers a local region instead of a pixel as basic analysis unit that can get richer information including texture, shape, and spatial relationship with neighboring objects.

OBCD is gaining increasing interest and a number of methods have been published in recent years [2]–[8]. The success of object-based analysis is that it offers an adequate framework to model spatial structure and contextual information instead of that the limited pixel-based features have, of which an effective feature representation is very important for constructing high-performance CD systems. The very first step of OBCD is to segment an image into disjoint homogeneous regions with a similar spectral characteristic using image segmentation methods such as simple linear iterative clustering (SLIC) [9]. Then low-level features [e.g., local binary patter (LBP), histograms of oriented gradients (HOG), and scale invariant feature transformation (SIFT)] are extracted for comparison or classification.

Recently, deep neural networks, known as deep learning, have achieved astonishing improvements over a variety of computer vision tasks, including image classification [10] and object detection [11]. In RS community, there are also great efforts attempting to apply deep learning techniques to RS imagery analysis, such as scene classification [12], building detection [13], CD [14], and so on. Of all the deep neural networks, convolutional neural network (CNN) is probably the most widely used one [15], [16]. CNN can learn hierarchical image representations from the input data by successively abstracting higher level features from the previous low-level ones. Despite its success in feature extraction and various vision tasks, a large volume of data is desired to train a usable network, which is intractable when only limited data are available. One way to solve this problem is transferring a model pretrained on related tasks with a very large data set to the current task. As has been evaluated in [17] and [18], CNNs trained on large natural scene data set (e.g., ImageNet) can generalize well to RS domains. In this letter, we also use pretrained CNNs to extract features from multitemporal RS images.

After feature extraction, most of existing approaches regard CD as a classification problem, either using supervised [5], [14], [19] or unsupervised [8], [20], [21] methods to obtain final change maps (CMs). Inspired by the recent progresses in visual saliency computation [22], [23] and the fact that changed objects or regions are salient objects/regions that attract human attention when displaying them alternately to an observer, we propose to use saliency detection framework to address CD problem. In this letter, we assume that only a small part of an image is changed regions, which will leave us large part of unchanged regions. Thus, CD can be achieved through matrix low-rank decomposition, in which the low-rank part will be considered as the unchanged areas and the sparse part will be the changed areas. Experiments on multiple data sets demonstrate the feasibility and effectiveness of the proposed method.

The remainder of this letter is organized as follows. In Section II, we elaborate the details of the proposed method. Section III presents the qualitative and quantitative experimental results. Finally, the conclusions are given in Section IV.

II. Change Detection Based on CNN and Low Rank

In this section, we will give the details of our deep features and low-rank decomposition-based CD method. The proposed method can be divided into four parts, as described in Fig. 1.

Manuscript received June 13, 2017; revised September 11, 2017 and October 20, 2017; accepted October 22, 2017. Date of publication November 10, 2017; date of current version December 4, 2017. This work was supported by the Natural Science Foundation of China under Grant 61601011 and Grant 61421003. *(Corresponding author: Qingjie Liu.)*

The authors are with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China (e-mail: qingje.liu@buaa.edu.cn).

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

Digital Object Identifier 10.1109/LGRS.2017.2766840

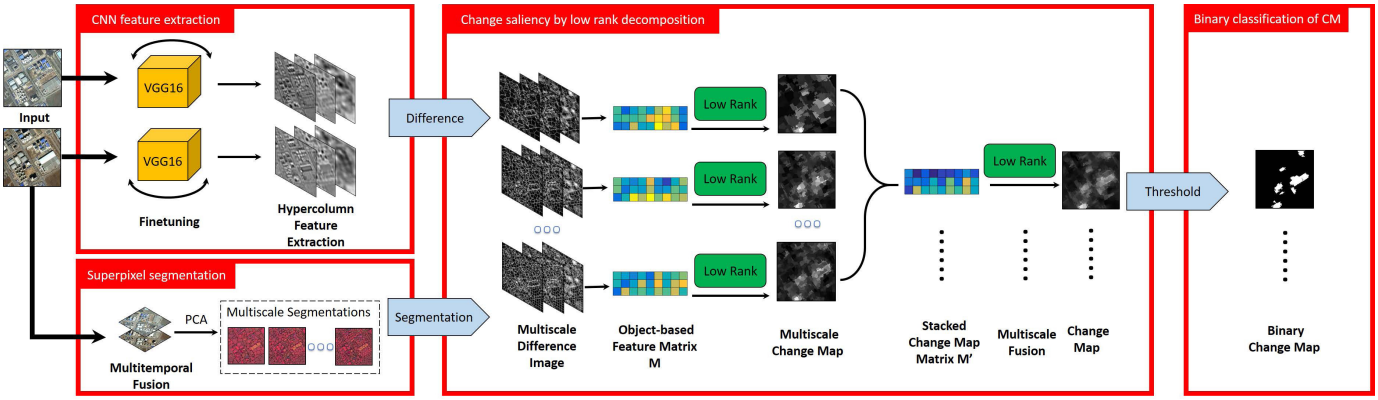


Fig. 1. Flowchart of the proposed method.

First, the input bitemporal images are partitioned into spectrum homogeneous and spatial adjacent regions. In this letter, we use SLIC [9] to perform segmentation. CNN features are employed to represent each segment of the two images using a strategy similar to [24]. After a simple absolute difference between the corresponding segments, a matrix representing difference features of the two images can be obtained. Subsequently, low-rank decomposition is applied to decompose the matrix into two parts, in which the low-rank part explains the unchanged regions and the sparse noises indicate the changed ones. Since there may exist changed objects in various sizes, to obtain better results, a multiscale segmentation and fusion strategy similar to [23] is adopted. The output of low-rank decomposition and fusion is a saliency map indicating probability of each pixel being changed. Finally, binarized CM can be obtained by simply thresholding the saliency map.

A. CNN Feature Extraction

Although CNNs have shown great capability of learning high-level features of images, the utilization of them for processing remotely sensed imagery is relatively recent. There is still a lack of investigations on using CNNs for CD. In this letter, we explore to utilize CNN as feature extractor to represent RS data.

An ideal way to harness capacity of CNN is designing end-to-end architectures that accept raw images as input and output the desired results. However, training such networks is difficult and expensive. Amount of labeled images and basic hardware resource (e.g., a GPU with CUDA support) should be provided. To alleviate this problem, we transfer CNNs pretrained on large-scale natural image data set, e.g., ImageNet, to RS domains.

One typical approach to transfer CNN is utilizing CNN features from the last fully connected layers that represent global feature of the input image and display more semantic abstraction properties. In CD, the changed parts are local regions, and thus local features are more appropriate choice for it. Inspired by [24] that uses *hypercolumns* of pixels to perform object segmentation and localization, we employ a similar idea that uses activations of CNN units above one pixel as its descriptor to conduct CD. The concept of hypercolumn is shown in Fig. 2.

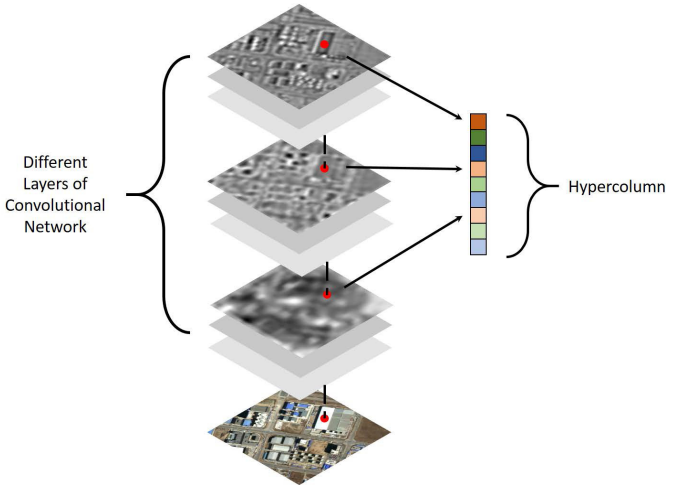


Fig. 2. Illustration of hypercolumn representation. This is a modified version of [24].

In this letter, we adopt VGG-16 network [25] to extract CNN features from the input RS images. VGG-16 contains 13 convolutional layers and three fully connected layers. There are, respectively, one pooling layer after the 2nd, 4th, 7th, 10th, and the 13th convolutional layers [25]. To get a better result, we domain-specifically fine-tune the VGG-16 to adapt to our optical RS images on an aerial image dataset (AID) [26] by replacing the original softmax with 30-way outputs.

Considering the subsampling and pooling operations in CNN, we resize each feature map to the same size of the input images by bilinear interpolation. Then for each pixel, its hypercolumn feature is formed by concatenating activations from feature maps at the same location of the pixel. Let $f_i (i = 1, \ldots, n)$ be the hypercolumn representation extracted from image I

$$f_i = [x_{i1}, x_{i2}, \ldots, x_{il}] \tag{1}$$

where x_{il} indicates the activation taken from the lth layer of the resized feature map at the ith pixel.

B. Superpixel Segmentation

Considering the deficiencies of traditional per-pixel analysis, object-based strategies are adopted in our method. To achieve this, SLIC [9] is applied to segment image

into meaningful geographic objects. First, we stack all bands of the two input images and acquire the first few principle components (three for our experiments) by principal component analysis (PCA). Then the blended image is decomposed into small homogeneous regions. For each segment $\{R_s\}_{s=1,\ldots,S}$ (S denotes the total number of segments), the mean of feature vectors within this segment is treated as the feature of this segment

$$F_j = \frac{\sum_{i=1}^{N} f_i'}{N} \tag{2}$$

where N denotes the number of pixels in the jth region R_j, f_i' is the subtraction of two images, i.e., $f_i' = |f_i^1 - f_i^2|$, and F_j is hypercolumn representation of segment R_j. Then we stack them to form the matrix representation of this image $M = [F_1, F_2, \ldots, F_s]$.

C. Change Saliency by Low-Rank Decomposition

In this letter, we address the problem of CD from a new perspective of visual saliency computation. Among numerous saliency computation models, low-rank based is one of the most successful ones. A common assumption of them is that a data matrix can be represented as a highly redundant information part (e.g., background regions) and a noisy part (e.g., foreground objects). In this letter, we make the same assumption. For the multitemporal images, the unchanged background structures are of the same regular shapes and keep no change as the time goes, only leaving a few changed regions. The redundant unchanged part usually lies in low-dimensional feature subspace, which can be approximated as a low-rank feature matrix, while the changed part can be viewed as a sparse noisy matrix.

Following this assumption, the CD can be formulated as:

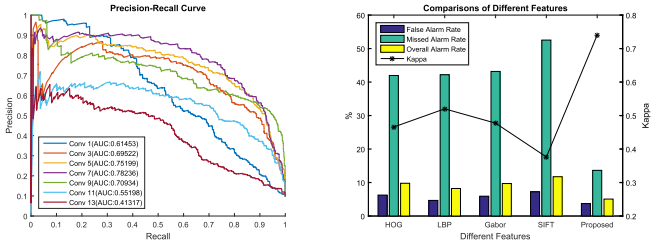
$$\begin{aligned} (\hat{L}, \hat{S}) &= \arg\min_{L,S} \operatorname{rank}(L) + \lambda \|S\|_0 \\ &\text{s.t. } M = L + S \end{aligned} \tag{3}$$

where M is the feature matrix obtained in Section II-B, L is the recovered low-rank matrix modeling background, and S is the sparse matrix modeling sparsity noisy. Since the above problem is NP-hard, the alternative convex surrogate to solve that is

$$\begin{aligned} (\hat{L}, \hat{S}) &= \arg\min_{L,S} \|L\|_* + \lambda \|S\|_1 \\ &\text{s.t. } M = L + S \end{aligned} \tag{4}$$

where $\|\cdot\|_*$ denotes the nuclear norm, which is a convex relaxation of the matrix rank function, $\|\cdot\|_1$ indicates l_1-norm that promotes sparsity, and the parameter $\lambda > 0$ is a tradeoff between the two items. L and S can be perfectly recovered by performing low-rank matrix recovery in (4) using robust PCA [27].

After obtaining S, the l_1-norm of each column S_j in S is used to measure the changing degree in the corresponding region. Larger changing degree indicates more remarkable changes, and vice versa. Finally, the changing degree of each region can be combined to form a CM that can be thought as the change probability of each pixel. Since we use a single descriptor to represent each segment, pixels within one segment share the same probability.



(a) (b)

Fig. 3. (a) Precision–recall curves of our method on the three data sets with different convolutional layers. (b) Comparisons with handcrafted features, including HOG, LBP, Gabor, and SIFT.

Detecting changes using an object-based strategy in a single scale may fail in capturing the intrinsic change objects. To overcome the above disadvantage, we employ a multiscale fusion strategy similar to [23]. First, a multiscale superpixel pyramid with different segmentation scales is built in order to fit the various sizes of objects. For each scale, a CM can be obtained by applying low-rank decomposition. Then all CMs can be fused to generate the final CM using the method in [23]. The rationale of fusion method is that CD results from different scales should be consistent. If we stack all the CMs in one matrix $\tilde{M}$ (each row is a CM), then $\tilde{M}$ must satisfy the characteristics of low rank. It can be represented as $\tilde{M} = \tilde{L} + \tilde{S}$. Each row $\tilde{S}_i$ of $\tilde{S}$ represents the disparity of the corresponding scale. The larger the $\tilde{S}_i$ is, the more inconsistent this scale is with others, so the CM_i should be endowed with a relatively small weight

$$\omega_i = \frac{\exp\left(-\|\tilde{S_i}\|_1^2\right)}{\sum \exp\left(-\|\tilde{S_i}\|_1^2\right)}. \tag{5}$$

Ultimately, the fused change probability can be obtained by $CM = \sum \omega_i \cdot CM_i$.

D. Binary Classification of CM

With the generation of the CM, the final binary CD result, namely, BCM, can be obtained by segmenting the CM using a threshold defined as

$$T = \alpha \cdot \operatorname{mean}(CM) \tag{6}$$

where mean$(\cdot)$ is the mean value of all pixels. The pixels in CM with probabilities higher than T are identified as changes; otherwise, they are identified as unchanges.

III. Experiments and Discussion

A. Data Set

To assess the effectiveness of the proposed method, three different sets of HR RS images are used. The first two pairs are cropped from Google Earth and with the same size of 500×500 pixels, three bands, and 1-m/pixel spatial resolution. The two sites are acquired on September 30, 2012, and March 4, 2013, respectively, at Beijing, as shown in the first two rows of Fig. 4(a) and (b). The last pair is derived from GF-2 satellite and acquired on February 17, 2015, and September 12, 2015, at Beijing with a spatial resolution of 1 m, 500×500 pixels, and four bands (R, G, B, and NIR) as shown in the last row of Fig. 4(a) and (b). The ground

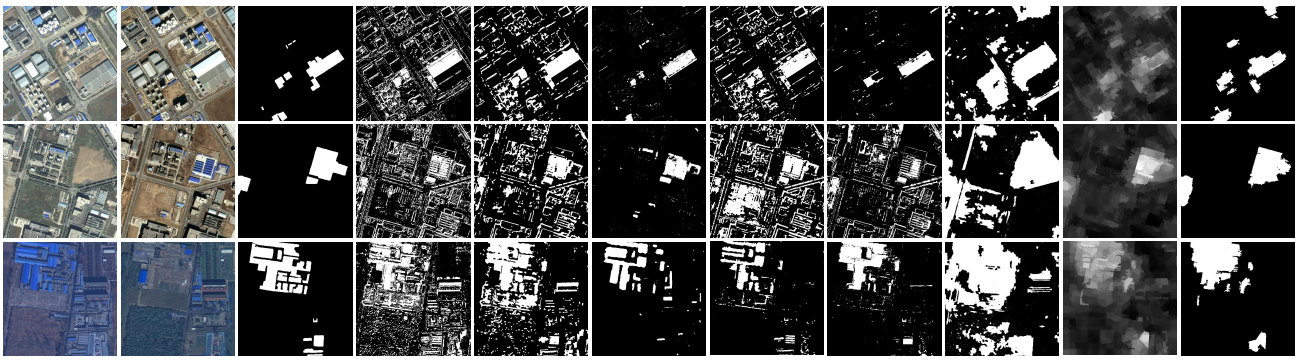


Fig. 4. Comparisons of different methods on the three image pairs. (a) and (b) Images from two different times. (c) Ground truth. (d) EM-based method [20]. (e) MRF-based method [20]. (f) IR-MAD-based method [28]. (g) PCA-based method [21]. (h) Parcel-based method [3]. (i) SHC-based method [19]. (j) CM of the proposed method. (k) BCM of the proposed method.

TABLE I

PERFORMANCE COMPARISONS OF DIFFERENT APPROACHES ON THE THREE DATA SETS

Methods	Dataset 1				Dataset 2				Dataset 3			
	False Alarms	Missed Alarms	Overall Alarms	Kappa	False Alarms	Missed Alarms	Overall Alarms	Kappa	False Alarms	Missed Alarms	Overall Alarms	Kappa
EM-based [20]	43098(0.1836)	3257(0.2137)	46355(0.1854)	0.2712	54452(0.2363)	8304(0.4242)	62756(0.2510)	0.1633	48366(0.2317)	13371(0.3544)	61737(0.2504)	0.3000
MRF-based [20]	30986(0.1320)	2552(0.1675)	33538(0.1342)	0.3741	40386(0.1753)	6916(0.3533)	47302(0.1892)	0.2645	32655(0.1564)	10520(0.2788)	43175(0.1751)	0.4553
IR-MAD-based [28]	2840(0.0121)	5579(0.3661)	**8419(0.0337)**	0.6788	2357(0.0102)	9434(0.4819)	11791(0.0472)	0.6085	5591(0.0268)	18671(0.4948)	24262(0.0984)	0.5576
PCA-based [21]	31301(0.1333)	3202(0.2101)	34503(0.1380)	0.3526	60915(0.2644)	7389(0.3774)	68304(0.2732)	0.1591	4105(0.0197)	18673(0.4949)	22778(0.0924)	0.5766
Parcel-based [3]	2511(0.0107)	6527(0.4283)	9038(0.0362)	0.6398	36837(0.1599)	11553(0.5901)	48390(0.1936)	0.1571	6842(0.0328)	14675(0.3889)	21517(0.0873)	0.6320
SHC-based [19]	49822(0.2122)	796(0.0522)	50618(0.2025)	0.2937	57698(0.2504)	25(0.0013)	57723(0.2309)	0.3187	62824(0.3009)	2401(0.0636)	65225(0.2646)	0.3837
Proposed	8297(0.0353)	1440(0.0945)	9737(0.0389)	**0.7189**	6121(0.0266)	555(0.0283)	**6676(0.0267)**	**0.8363**	10436(0.0500)	10807(0.2864)	**21243(0.0862)**	**0.6663**

truth of the above data sets is shown in Fig. 4(c), which is manually created based on visual interpretation. The main land-cover types include soil, road, building, grass, and tree. As mentioned above, we fine-tune VGG-16 on AID data set [26]. Three scales 100, 250, 400, and the ratio between clustering appearance and spatial regularization 60 are selected for SLIC segmentation. The threshold T for segmenting CM into binary mask is set in such a way that $\alpha = 1.7$ for Google Earth images and $\alpha = 1.5$ for GF-2 images.

B. Comparisons of Different Convolutional Layers

Considering that the unit in CD is either a pixel or a segment, we follow Hariharan's idea [24] that employed activations of CNN layers to form a feature vector for each pixel. Instead of using all CNN layers to form representation for each pixel, which is highly redundant and memory cost, we selectively choose convolutional layers to form descriptors. The features derived from different layers ranging from the 1st to 13th are extensively evaluated on CD task. The results are reported in Fig. 3(a). It can be seen that the first and the last two convolutional layers obtain the worst results, and layer 7 achieves the best result. This is consistent with the common understanding of CNN that the first layer of CNN acts like a Gabor filter bank extracting low-level features, while the highest layers with receptive field of entire image represent high-level semantic information of the image; however, they have low discriminability for pixels. Thus, in this letter, we use a combination of intermediate layers (i.e., layers 5, 7, and 9) as features.

C. Comparisons With Handcrafted Features

We further compare CNN features with the most widely used low-level handcrafted features, including HOG, LBP, Gabor, and SIFT. We use false alarm rate (FAR), missed alarm rate (MAR), overall alarm rate (OAR), and *kappa* coefficient to evaluate the performance of these features. In experiments, except for different features, other experimental settings are the same. The performances are shown in Fig. 3(b). It can be seen that CNN features significantly outperform handcrafted features in terms of the lowest FAR, MAR, OAR, and the best *kappa* value, which reveals the substantial superiority in the representative ability of CNN features.

D. Comparisons With Other Methods

In this section, we make comparisons between the proposed and traditional algorithms, that is, EM-based [20], Markov random field (MRF)-based [20], iteratively reweighted multivariate alteration detection (IR-MAD)-based [28], PCA-based [21], parcel-based [3], and sparse hierarchical clustering (SHC)-based methods [19]. The involved parameters are set as that in the original paper. The results on the three data sets are shown in Fig. 4. Four indicators FAR, MAR, OAR, and *kappa* are adopted for quantitative comparisons, as listed in Table I.

It can be observed from Fig. 4 that the results of EM- and MRF-based methods are filled with lots of white spots noise. MRF- and PCA-based methods both utilize contextual information to make final decision, and thus perform better than EM-based method, as shown in Fig. 4(d), (e), and (g). IR-MAD-based method detects changes by calculating the canonical variates to obtain the orthogonal differences that contain maximum information on joint change and gets better results as shown in Fig. 4(f). Parcel-based method models complex objects by utilizing multilevel context information, resulting in better performance with accurate change region boundary, as shown in Fig. 4(h).

SHC-based method generates CMs with clean backgrounds; however, it brings in many false alarms. The proposed method works better at highlighting changed pixels and suppressing unchanged ones. There are only a few scattered and wrongly identified changed pixels in the results. One deficiency of our method is that the borders of the changed regions are not satisfactory caused by inconsistent segmentation as shown in Fig. 4(k). In addition, for the last data set, the unchanged building is wrongly detected as a changed one as shown in Fig. 4(d), (e), (h), and (i). From the CM in Fig. 4(j), we can see that the intensities of changed areas are higher than that of unchanged, which reflects the superiority of our method.

From Table I, we can find that the proposed method can yield lower OAR and higher *kappa* than other methods. It is worth noting that the OAR value of our method is slightly lower than that of IR-MAD-based method and parcel-based method on data set 1. It can be seen that the MRF-, PCA-, and parcel-based methods can obtain a higher *kappa* compared with other methods on data sets 1 and 3. However, the last two ones are not robust to speckle noise, leading to a lower *kappa* on data set 2. Parcel-based method outperforms other methods on data sets 1 and 3; however, it fails on data set 2. The IR-MAD-based method performs well on all three data sets especially when most methods fail. Even with the lowest missed alarms, the SHC-based method cannot give satisfied results with the highest false alarms. Our method, benefiting from CNN features and low-rank model, can greatly improve the detection accuracies. Furthermore, it works well for different types and scales of changes on all data sets. It is clear that the proposed method outperforms other methods regardless of the qualitative or quantitative analysis.

IV. Conclusion

In this letter, a novel CD technique based on CNN features and low-rank decomposition is proposed, in which the CD is achieved by detecting the salient regions of the difference feature images using low-rank decomposition. Instead of using traditional handcrafted features, the CNN features from multiple layers are employed to form representations for each superpixel. After applying low-rank decomposition, a saliency map indicating change probabilities of pixels will be generated. Binarized CMs can be obtained using a simple threshold. We test our method on multiple image pairs retrieved from Google Earth and GF-2 images, and compare it with other six methods. The results demonstrate the effectiveness of the proposed method.

Acknowledgment

The authors would like to thank K. Ding for providing the related comparison experiment source codes and the reviewers for the suggestions to improve this letter.

References

[1] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 91–106, Jun. 2013.

[2] J. Im, J. R. Jensen, and J. A. Tullis, "Object-based change detection using correlation image analysis and image segmentation," *Int. J. Remote Sens.*, vol. 29, no. 2, pp. 399–423, 2008.

[3] F. Bovolo, "A multilevel parcel-based approach to change detection in very high resolution multitemporal images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 33–37, Jan. 2009.

[4] G. Chen, G. J. Hay, L. M. T. Carvalho, and M. A. Wulder, "Object-based change detection," *Int. J. Remote Sens.*, vol. 33, no. 14, pp. 4434–4457, 2012.

[5] B. Hou, Y. Wang, and Q. Liu, "A saliency guided semi-supervised building change detection method for high resolution remote sensing images," *Sensors*, vol. 16, no. 9, p. 1377, 2016.

[6] N. Falco, M. Dalla Mura, F. Bovolo, J. A. Benediktsson, and L. Bruzzone, "Change detection in VHR images based on morphological attribute profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 636–640, May 2013.

[7] B. Wang, S. Choi, Y. Byun, S. Lee, and J. Choi, "Object-based change detection of very high resolution satellite imagery using the cross-sharpening of multitemporal data," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1151–1155, May 2015.

[8] L. Ma *et al.*, "Object-based change detection in urban areas: The effects of segmentation strategy, scale, and feature space on unsupervised methods," *Remote Sens.*, vol. 8, no. 9, p. 761, 2016.

[9] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2015.

[12] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep learning earth observation classification using ImageNet pretrained networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 105–109, Jan. 2016.

[13] Q. Zhang, Y. Wang, Q. Liu, X. Liu, and W. Wang, "CNN based suburban building detection using monocular high resolution Google Earth images," in *Proc. IEEE IGARSS*, Jul. 2016, pp. 661–664.

[14] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[16] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[17] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, 2015.

[18] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proc. CVPR Workshops*, 2015, pp. 44–51.

[19] K. Ding, C. Huo, Y. Xu, Z. Zhong, and C. Pan, "Sparse hierarchical clustering for VHR image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 577–581, Mar. 2015.

[20] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[21] 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.

[22] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. CVPR*, 2012, pp. 853–860.

[23] R. Huang, W. Feng, and J. Sun, "Saliency and co-saliency detection by low-rank multiscale fusion," in *Proc. ICME*, 2015, pp. 1–6.

[24] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," in *Proc. CVPR*, 2015, pp. 447–456.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015.

[26] G.-S. Xia *et al.*, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.

[27] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. NIPS*, 2009, pp. 2080–2088.

[28] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 463–478, Feb. 2007.