VHR IMAGE CHANGE DETECTION BASED ON DISCRIMINATIVE DICTIONARY LEARNING

Kun Ding, Chunlei Huo, Yuan Xu, Chunhong Pan

NLPR, Institute of Automation, Chinese Academy of Sciences
kding@nlpr.ia.ac.cn

ABSTRACT

The difficulty of Very High Resolution (VHR) image change detection is mainly due to the low separability between the changed and unchanged class. The traditional approaches usually address the problem by solving the feature extraction and classification separately, which cannot ensure that the classification algorithm makes the best use of the features. Considering this, we propose a novel approach that combines the feature extraction and the classification task by utilizing the sparse representation algorithm with discriminative dictionary. Experiments on real data sets show that our method achieves effective results.

Index Terms— Change detection, VHR remote sensing image, sparse representation, discriminative dictionary

1. INTRODUCTION

Change detection problem aims at detecting the changed regions of the co-registered images acquired over the same scene at different times. With the development of VHR sensors, change detection of VHR images has a great potential applications such as disaster management, environmental monitoring, urban management, etc. However, the nature of VHR image changes the prospectives of change detection, and the traditional change detection methods for low-to-moderate resolution images cannot be applied for VHR images directly. The existing techniques are far from the practical requirements with respect to accuracy, speed or degree of automation.

In general, the difficulty of VHR image change detection lies in the low separability between the changed and unchanged class, which is attributed to the complex imaging procedure of VHR images. For the traditional low-to-moderate resolution images, the changes are available related to the spectral difference, and most of the traditional change detection approaches are based on the pixel-wise comparison. In contrast, the problem of mixed pixels is alleviated for VHR images, but the interclass variance is not improved simultaneously with the spatial resolution increase (the discriminability between different land-cover classes is determined simultaneously by the spatial resolution and the spectral resolution, and there is a tradeoff between the spatial resolution and the spectral resolution). In consequence, the change features are difficult to be classified with high accuracy even by a "good" classifier. In addition, due to the impacts caused by the image registration error, view-angle variation, meteorological or seasonal changes, the unchanged class is more difficult to be separated from the changed class.

Many approaches are presented in the literature to address the above difficulties. For instance, Huo proposed to extract discriminative local features [1] (e.g., Scale Invariant Feature Transform (SIFT)) to represent complex urban objects and utilize robust distance metric to improve the separability between the changed and unchanged class. Mura [2] proposed to use morphological filters to preserve the geometrical structures and filter the homogeneous areas simultaneously. However, one limitation of the above approaches lies in the separation of feature extraction and classification. In our opinion, the discriminativeness of the features is closely related to the classifier. High separability between the changed and unchanged class cannot be guaranteed if the feature extraction and classification are implemented step by step. For the similar reason, [3] concludes that sparsity cannot help improve classification by the traditional manner. Sparse representation and dictionary learning have been widely used in signal de-noising [4], face recognition [5], etc. But they are new for remote sensing image processing, and only few change detection approaches [6, 7] are related to sparse representation and dictionary learning; Nevertheless, all of them are designed for low resolution images, and the above two key steps (e.g., change feature extraction and change feature classification) are implemented separately. To address the difficulties of VHR images, a novel approach is proposed in this paper. The rationale of the proposed approach is to improve the low separability between the changed class and the unchanged class by considering change feature extraction and change feature classification simultaneously, which is implemented by sparse representation and discriminative dictionary learning. To our best knowledge, there is no such papers in the literature.

The remainder of this paper is organized as follows: Section 2 describes the related work. Section 3 elaborates the proposed approach step by step. Section 4 reports the

experimental results on real QuickBird images. Finally, Section 5 draws the conclusions.

2. RELATED WORK

2.1. Sparse Coding

Given a signal $\mathbf{y} \in \mathbf{R}^n$ and a dictionary $\mathbf{D} \in \mathbf{R}^{n \times K}$, $\mathbf{y}$ can be represented by the sparse coding $\mathbf{x} \in \mathbf{R}^K$ as follows

$$\mathbf{x} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2, \ s.t. \ \|\mathbf{x}\|_0 \leq T, \quad (1)$$

where T controls the sparsity of $\mathbf{x}$. There are many algorithms for solving this problem. For instance, Matching Pursuit (MP) [8] and Orthogonal Matching Pursuit (OMP) [9] solve l_0 norm problem directly, and other methods such as Gradient Projection (GP) [10], Homotopy [11] solve the following relaxed problem

$$\mathbf{x} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2, \ s.t. \ \|\mathbf{x}\|_1 \leq T. \quad (2)$$

2.2. Dictionary Learning

Given signals $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N] \in \mathbf{R}^{n \times N}$, the dictionary learning problem can be formulated as follows

$$< \mathbf{D}, \mathbf{X} >= \arg\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2, \ s.t. \forall i, \ \|\mathbf{x}_i\|_0 \leq T, \quad (3)$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbf{R}^{K \times N}$ are sparse representation coefficients. The problem (3) can be solved by K-SVD [12], ODL [13], etc.

A recent trend is adding the discriminativeness to the learned dictionary. For example, Zhang proposed a discriminative K-SVD method [14], which is a relaxed version of supervised dictionary learning [15]. Other discriminative dictionary learning methods can be found in [16, 17].

3. THE PROPOSED APPROACH

Let us consider two multi-temporal co-registered VHR images $\mathbf{Z}_1$ and $\mathbf{Z}_2$ of size $I \times J$ acquired in the same geographical area at different times t_1 and t_2. The objective of change detection is estimating a binary change map $\mathbf{B}$, where the value of pixel (i, j) will be 1 if there is a significant change and 0 otherwise.

The main idea of the proposed approach is to capture the complex objects by local features, encode and classify the discriminative change features by sparse coding and discriminative dictionary learning. As illustrated by Fig. 1, the proposed change detection method consists of the following four steps: feature extraction, training sample selection, dictionary and classifier learning and change map generation.

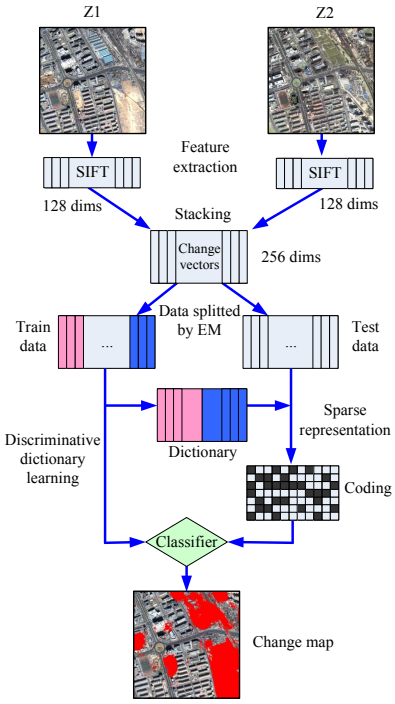

Fig. 1. Flowchart of the proposed change detection method.

3.1. Feature Extraction

Due to the intricacies of VHR images, it is difficult for the pure usage of the spectral features to capture the complex structures of urban objects. For this reason, SIFT [18], a local descriptor, is used to encode the local structure and texture information. The invariance to the linear changes of spectral features can help reduce the false alarms caused by impacts such as seasonal and lighting variation. In addition, the rotation invariance makes it robust to the misalignment. In this paper, dense SIFT features $\mathbf{z}_1$ and $\mathbf{z}_2$ are extracted at each pixel of the two images $\mathbf{Z}_1$ and $\mathbf{Z}_2$ individually. Then the change vector is formed by stacking $\mathbf{z}_1$ and $\mathbf{z}_2$, i.e., $\mathbf{y} = [\mathbf{z}_1^T \ \mathbf{z}_2^T]^T$. An alternative way is differencing $\mathbf{z}_1$ and $\mathbf{z}_2$, but it will lose information, so it is not used this paper.

3.2. Training Sample Selection

For classification-based change detection method, training samples need to be selected. The traditional manual labeling is very tedious, especially in the cases where a large amount of data need be processed. In consequence, automatic training sample selecting is preferred. Inspired by the EM-based change detection method for low resolution remote sensing images in [19], we adopt double thresholds to select the reliable training samples. Different from the method in [19], the proposed approach is based on the amplitude of SIFT feature difference, instead of the spectral difference. This amplitude is computed by the Euclidean distance between two corresponding SIFT features, i.e. $z_d = \|\mathbf{z}_1 - \mathbf{z}_2\|_2$. Without loss of generality, supposing all the z_d are independent with

each other and obey Gaussian Mixture Distribution:

$$p(z_d) = p(z_d|\omega_n)P(\omega_n) + p(z_d|\omega_c)P(\omega_c), \quad (4)$$

where $P(\omega_c)$ and $P(\omega_n)$ are the prior probability of the changed and unchanged class respectively, $p(z_d|\omega_c)$ and $p(z_d|\omega_n)$ can be modeled by Gaussian distributions, and the parameters can be estimated by EM algorithm. The optimum threshold thr for dividing the changed and unchanged class can be obtained by solving the following equation

$$p(z_d|\omega_n)P(\omega_n) = p(z_d|\omega_c)P(\omega_c). \quad (5)$$

Based on the estimated parameters: the mean values of the changed and unchanged class m_c, m_n, and the optimum threshold thr, we can define the double thresholds

$$thr_u = m_u + \theta(thr - m_u), \quad (6)$$

$$thr_c = m_c + \theta(thr - m_c), \quad (7)$$

where θ controls the number of positive (changed) and negative (unchanged) training samples. The smaller the θ is, the less the training data will be. In other words, we want to select the most reliable training samples, but the distribution may not be typical. After determining the double thresholds, the candidate training samples are chosen as follows: if the sample z_d satisfies $z_d \geq thr_c$, it will be added into the positive training set $\mathbf{Y}_c$; if $z_d \leq thr_u$, it will be added into the negative training set $\mathbf{Y}_u$. For convenience, the whole training set is denoted as $\mathbf{Y} = [\mathbf{Y}_c\ \mathbf{Y}_u]$, and the corresponding label matrix $\mathbf{H} = [\mathbf{H}_c\ \mathbf{H}_u]$. To make the proposed method much faster and keep good performance, only $r\%$ (e.g. 4%) of the candidate training samples are selected at random for the latter dictionary learning.

3.3. Dictionary and Classifier Learning

Given the matrix $\mathbf{Y}$ and $\mathbf{H}$, the discriminative dictionary $\mathbf{D}$ and the classifier $\mathbf{W}$ are expected to achieve simultaneously. To this aim, we adopt the following discriminative dictionary learning model

$$\begin{aligned} <\mathbf{D},\mathbf{W},\mathbf{X}> = \arg\min_{\mathbf{D},\mathbf{W},\mathbf{X}} & \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2 \\ & +\beta\|\mathbf{H}-\mathbf{W}\mathbf{X}\|_F^2 + \alpha\|\mathbf{W}\|_F^2 \\ & s.t.\forall i,\ \|\mathbf{x}_i\|_0 \leq T, \end{aligned} \quad (8)$$

where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N] \in \mathbf{R}^{C\times N}$ is the label matrix of training samples, $\mathbf{h}_i$ is the label vector of the i-th sample (for 2 classes case, $\mathbf{h}_i$ is $[0\ 1]^T$ or $[1\ 0]^T$), C is the number of classes ($C = 2$ in change detection task), $\mathbf{W}$ is the classifier coefficient matrix. The first term represents reconstruction errors, the second term stands for classifying losses, the last term is used for preventing over-fitting. The above problem can be solved by K-SVD effectively. By the above model, the dictionary and classifier are learned simultaneously, which provides the potential to improve the low separability between the changed and unchanged class by coding the change features as mentioned above.

3.4. Change Map Generation

After achieving discriminative dictionary $\mathbf{D}$ and the classifier $\mathbf{W}$, the task left is to encode the change vector $\mathbf{y}$ and label it based on the classifier $\mathbf{W}$. For a change vector $\mathbf{y}$, the Sparse Change Vector (SCV) $\mathbf{x}$ is obtained by solving Eq. (1). Then, the classifier score vector is $\mathbf{h} = \mathbf{W}\mathbf{x}$. The final label of change vector $\mathbf{y}$ is computed as follows

$$label = \arg\max_i h_i, \quad (9)$$

where h_i is the i-th element of $\mathbf{h}$. By this way, the change map can be acquired by classifying all the SCVs.

It is worth noting that SIFT features are computed within the local patch of the predefined size, and the size is closely related to the scale (or the size) at which the changes are observed. As we know, change detection accuracy is dependent on the scale (the image resolution or the observation window), and it is limited for the traditional approach to detect the complex changes reliably if only one scale information is considered. So multi-level method can be used to improve the performance. In this paper, we merge the results generated at different levels based on a simple majority voting method (as shown in Fig. 2), where the window size (equal to $4 \times binSize$) of SIFT descriptors varies at different levels.

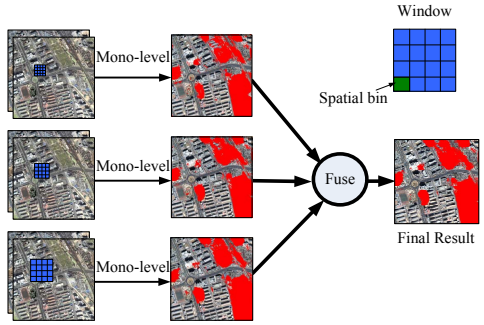

Fig. 2. Multi-level change detection by decision fusion.

4. EXPERIMENTAL RESULTS

For space limitation, experiments on two data sets are discussed in this paper. The images are acquired over Beijing (China) by QuickBird in 2002 and 2003 respectively, the image size of the first data set is 1024×1024 pixels(as shown in Figs. 3(a)(b)), and the second data set is 1120×1120 pixels (as shown in Figs. 3(d)(e)), the spatial resolutions are both 0.7m/pixel. The reference change maps for the two data sets are shown in the last column of Fig. 3.

One important difference of the proposed approach with others lies in the sparse coding and discriminative dictionary learning. To demonstrate the effectiveness of the proposed approach, we compare the proposed approach with EM-based approach (After getting the threshold thr, EM-based approach achieves the change map by thresholding z_d.) with

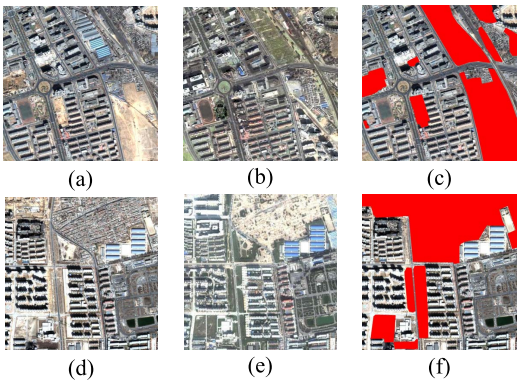

Fig. 3. Data sets used in our experiments. (a) and (d) are images taken in 2002. (b) and (e) are images taken in 2003. (c) and (f) are the reference change maps.

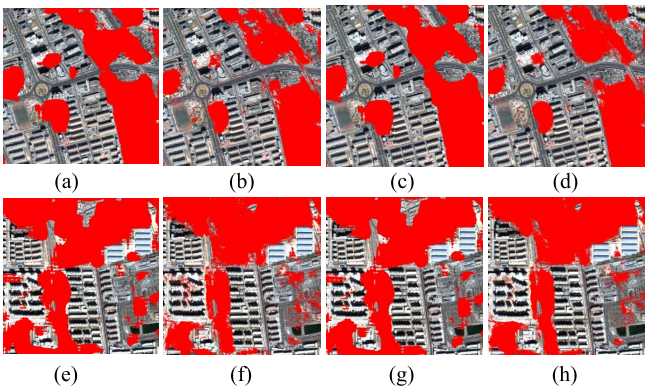

Fig. 4. Result comparison. (a) and (e) are obtained by EM threshold using single level context. (b) and (f) are obtained by our method using single level context. (c) and (g) correspond to three-level EM threshold method. (d) and (h) correspond to our method using three levels of context.

respect to False Alarm (FA), Missed Alarm (MA) and Total Error (TE). The above two approaches are conducted both in mono-level and multi-level fashion. In all cases, we keep the parameters θ, β, T, K, r fixed, i.e. $\theta = 0.1, \beta = 25, T = 10, K = 500, r = 4$. In mono-level case, parameter $binSize$ is set to be 24 pixels; in multi-level case, we change parameter $binSize$ from 20 to 28 pixels with a step size 4, i.e. we use three levels of context.

The obtained change maps for the first data set are shown in first row of Fig. 4 and the corresponding FAs, MAs and TEs are listed in the first six rows of Table.1. As can be seen from Table.1, our methods (S-DDL short for mono-level method based on discriminative dictionary learning and M-DDL is the corresponding multi-level version) get lower FAs than EM-based methods (S-EM and M-EM) in both mono-level and multi-level cases. But, the MAs increase a little, meaning that our methods make a trade-off between FA and MA. The possible reason lies in the following fact: the sampling rate r is so small that can't find an appropriate change vector (even if there is) to reflect the small regions of change, or the value of θ is too small to make the training samples representative enough.

The change maps for the second data set are shown in second row of Fig. 4 and the corresponding FAs, MAs and TEs are listed in the last six rows of Table.1. By comparison, it can be concluded that our method obtains better performances in both FA and MA than that by EM-based method.

The changes may happen in different scales, so multi-level context can help improve the performance. The above conclusion holds for the proposed approach. But for the EM-based method, the improvements benefited from multi-level context are insignificant, this is due to the fact that the changed areas detected using larger $binSize$ are usually covered by the regions detected using smaller $binSize$ for EM-based method.

Table 1. False Alarms (FAs), Missed Alarms (MAs), and Total Errors (TEs) resulted from EM threshold and our method for mono-level (S-) and multi-level (M-) cases. (DS1 stands for data set1 for short, DS2 is similar.)

Method			S-EM	S-DDL	M-EM	M-DDL
DS1	FA	pixels	88245	51826	82673	46674
		rate(%)	13.64	**8.01**	13.10	**7.39**
	MA	pixels	14593	37087	14324	33455
		rate(%)	**5.63**	14.30	**5.68**	13.26
	TE	pixels	102838	88913	96997	80129
		rate(%)	11.35	**9.81**	10.98	**9.07**
DS2	FA	pixels	146345	102897	137127	86853
		rate(%)	20.22	**14.22**	19.38	**12.27**
	MA	pixels	38107	24837	37321	20078
		rate(%)	10.17	**6.63**	10.21	**5.49**
	TE	pixels	184452	127734	174448	106931
		rate(%)	16.79	**11.63**	16.25	**9.96**

5. CONCLUSIONS

In this paper, a novel change detection method is proposed based on sparse representation and discriminative dictionary learning. By learning a discriminative dictionary, the sparse representation coefficients are easier to be separated, which makes our method enjoy a higher accuracy than the simple EM-based method. The experiments in both mono-level and multi-level context for different data sets demonstrate the effectiveness of our proposed method. Future work will concentrate on more superior multi-level context methods.

6. ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China under Grants 61005013, 61005036, 60723005, and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA06030300).

7. REFERENCES

[1] C. Huo, B. Fan, C. Pan, and Z. Zhou, "Combining local features and progressive support vector machine for urban change detection of vhr images," in *ISPRS Annals*, 2012, vol. 1, pp. 221–226.

[2] M. Dalla Mura, J.A. Benediktsson, F. Bovolo, and L. Bruzzone, "An unsupervised technique based on morphological filters for change detection in very high resolution images," in *Geoscience and Remote Sensing Letters, IEEE*, 2008, vol. 5, pp. 433–437.

[3] Roberto Rigamonti, Matthew A. Brown, and Vincent Lepetit, "Are sparse representations really relevant for image classification?," in *CVPR*, 2011, pp. 1545–1552.

[4] Julien Mairal, Michael Elad, and Guillermo Sapiro, "Sparse representation for color image restoration," 2008, vol. 17, pp. 53–69.

[5] John Wright, Allen Y. Yang, Arvind Ganesh, Shankar S. Sastry, and Yi Ma, "Robust face recognition via sparse representation," 2009, vol. 31, pp. 210–227.

[6] Leyuan Fang, Shutao Li, and Jianwen Hu, "Multitemporal image change detection with compressed sparse representation," in *ICIP*, 2011.

[7] L.H. Nguyen and T.D. Tran, "A sparsity-driven joint image registration and change detection technique for sar imagery," in *ICASSP*, 2010, pp. 2798–2801.

[8] Stphane Mallat and Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," 1993, vol. 41, pp. 3397–3415.

[9] Y.C. Pati, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, 1993, vol. 1, pp. 40–44.

[10] M.A.T. Figueiredo, R.D. Nowak, and S.J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," 2007, vol. 1, pp. 586– 597.

[11] M. R. Osborne, Brett Presnell, and B.A. Turlach, "A new approach to variable selection in least squares problems," 1999.

[12] M. Aharon, M. Elad, and A. M. Bruckstein, "The k-svd: An algorithm for designing of overcomplete dictionaries for sparse representations," 2006, vol. 54, pp. 4311–4322.

[13] Julien Mairal, Francis Bach, and Jean Ponce, "Online dictionary learning for sparse coding," 2009, pp. 689–696.

[14] Qiang Zhang and Baoxin Li, "Discriminative k-svd for dictionary learning in face recognition," in *CVPR*, 2010, pp. 2691–2698.

[15] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman, "Supervised dictionary learning," in *NIPS*, 2008, pp. 1033–1040.

[16] Zhuolin Jiang, Zhe Lin, and Larry S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent k-svd," in *CVPR*, 2011, pp. 1697–1704.

[17] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman, "Discriminative learned dictionaries for local image analysis," in *CVPR*, 2008, pp. 1–8.

[18] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, pp. 91–110, 2004.

[19] F. Bovolo, "Automatic analysis of the difference image for unsupervised change detection," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 38, pp. 1171–1182, 2000.