ROBUST VHR IMAGE CHANGE DETECTION BASED ON LOCAL FEATURES AND MULTI-SCALE FUSION

Yuan Xu, Chunlei Huo, Shiming Xiang, Chunhong Pan

NLPR, Institute of Automation, Chinese Academy of Sciences
{yxu,clhuo,smxiang,chpan}@nlpr.ia.ac.cn

ABSTRACT

Urban change detection of Very High Resolution (VHR) remote sensing images is challenging, due to the ill-posed nature of change detection problem, the inherent nature of VHR image, the complex morphology of urban scenes, etc. To address the above difficulties, a robust approach is proposed, which is based on discriminative local features, robust distance metric and novel multi-scale fusion strategy. By integrating these components synergistically, the proposed approach is superior to the traditional approaches in capturing semantic changes and removing the false changes. Comparative experiments demonstrate the effectiveness and advantages of the proposed approach.

Index Terms— Change detection, VHR image, local features, belief propagation, multi-scale fusion.

1. INTRODUCTION

The goal of image change detection is to determine the changed areas in the co-registered multi-temporal images. With the development of VHR satellites such as QuickBird and WorldView, the details of changed areas can be observed and they are very important for practical applications. For this reason, VHR image change detection receives widespread attention. However, there are some difficulties due to the nature of VHR image and change detection technique.

Firstly, the inherent nature of VHR image, such as high intra-class and low interclass variabilities [1], leads to the reduction of the statistical separability of different landcover classes. And the statistical separability between the changed and unchanged classes is not improved simultaneously with the increase of the spatial resolution. For this reason, traditional change detection approaches [2] [3] [4] are difficult to be applied to VHR image without considering its complexities.

Secondly, the definition of 'change' is ambiguous. Actually, changes can be defined differently by users in various work settings related to specific tasks. It is difficult for computers to detect the interested changes for different users automatically. Moreover, as the increase of spatial resolution, trivial changes are mixed up with the interested ones and they are difficult to be removed.

Last but not least, changes are related to scales. Specifically, subtle changes can be detected at a high resolution level, but the detection result is susceptible to the interference of registration error and view-angle variation, etc. Those interference can be reduced at a low resolution level, while subtle changes will be missed.

The key techniques in change detection of VHR images are change feature extraction, classification and multi-scale fusion. Many novel approaches are proposed in the literatures to address the above problems. For instance, Volpi et al. [5] proposed to concatenate spectral, textural and morphological features to structure a change feature. Huo et al. [6] proposed to classify object-specific change features by progressive transductive SVM, where initial training samples are selected automatically based on the magnitude of change features. Celik [7] proposed to fuse the multi-scale results from coarse to fine in AND logic function and the final result is optimized scale by scale.

To overcome the aforementioned difficulties, a novel change detection approach is proposed in this paper for VHR images of urban scenes by taking the above three important components into account simultaneously. Compared with the related techniques, performance is promising in capturing subtle changes and removing false changes.

2. PROPOSED CHANGE DETECTION APPROACH

As our focus is on urban images, the complex constitution of urban area should be considered. In this paper, the complex change features of urban area are encoded by discriminative local feature in a robust distance metric. The trivial change classes and the interested classes are distinguished by a semi-supervised classification trained by both unlabeled samples and samples labeled by user. To cope with the scale dependence, the final result is captured by a novel multi-scale fusion strategy with context and scale information. The flowchart of the proposed approach is shown in Fig. 1, which consists of three parts, multi-scale decomposition, change detection and multi-scale fusion. We will elaborate each component step by step.

2.1. Multi-scale Decomposition

Changes between multi-temporal images are highly dependent on scales. This paper aims at reducing the dependence of changes on scales by fusion strategy. Multi-scale decomposition is thus required to generate hierarchical images of different resolutions for each VHR image. This can be implemented by direct sub-sampling or pyramid decomposition. By experiments, we found that similar results are achieved by different multi-scale decomposition approaches. For efficiency, we use direct sub-sampling in this paper. For the co-registered multi-temporal images $I_i\,(i = 1, 2)$, $I_i^{(n)}$ is obtained by sub-sampling $I_i^{(n-1)}$ in a factor 2 based on the bicubic interpolation, where $I_i^{(0)} = I_i$. By this way, we get two image pyramids

This work was supported in part by the National Basic Research Program of China under Grant 2012CB316304, and the National Natural Science Foundation of China under Grants 61005013, 61175025, and 60723005.

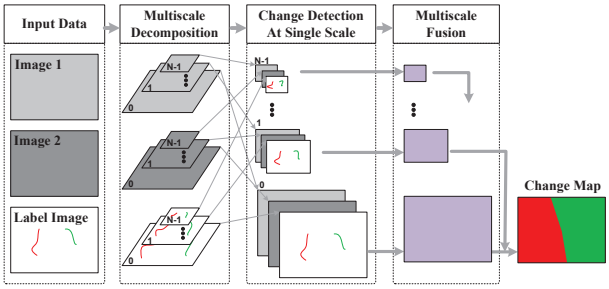


Fig. 1. The flowchart of the proposed approach.

$\left\{I_i^{(0)}, I_i^{(1)}, \cdots I_i^{(N-1)}\right\}(i=1,2)$. The sub-sampling level N is chosen as follows:

$$N = floor(\log_2(M/K)), \tag{1}$$

where $floor(\cdot)$ is the floor function, M is the minimum of the width and height of images, K is the user-defined minimized size of the image after sub-sampling.

2.2. Change Detection At Single Scale

Before multi-scale fusion, changes are required to be detected at each scale individually. The performance of the final result is closely related to the detection performance at each scale. So it is necessary to design an effective change detection approach for each scale. Considering the difficulties of VHR images, a novel change detection is designed in this paper, which consists of two steps: change feature extraction and change feature classification.

2.2.1. Change Feature Extraction

To improve the statistical separability between the changed and unchanged classes of VHR images, we propose to use discriminative local features and robust distance metric.

The pure usage of spectral difference is limited to represent the complex structures on man-made objects. To encode the structure feature of VHR image, SIFT [8] descriptor is extracted at each pixel to characterize local image structures and encode the contextual information. Compared to the raw spectral features, SIFT descriptor is a higher level feature and more powerful in capturing the salient structures of man-made objects.

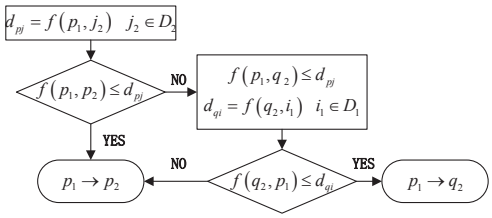


Fig. 2. Illustration of searching the true corresponding pixel.

Even if the object is encoded effectively by the discriminative local features, the difference between multi-temporal images is usually disturbed by view angle variation and registration error, etc. However, human can easily avoid these errors. The underlying reason is that the appearances of the objects under different view angles are regarded to be very similar by human vision system as long as the local displacement is within certain ranges. Motivated by this observation, a robust distance measure is proposed as illustrated by Fig. 2, which is based on local search and Euclidean distance between SIFT descriptors. For illustration, p_1 and q_1 in image I_1 are coordinate corresponding pixel of p_2 and q_2 in image I_2 respectively. D_2 and D_1 are sets of neighborhood pixels j_2 and j_1 of p_2 and q_1. Function $f(\cdot)$ outputs the Euclidean distance d_{pj} between SIFT features of pixel p_1 and j_2. $p_1 \to q_2$ denotes that q_2 is the true corresponding pixel of p_1.

2.2.2. Change Feature Classification

Once the above robust change features are computed, changes can be detected by classifying the change features. Noting the ambiguous definition of 'change' and the low statistical separability between the changed and unchanged classes, training samples selection and classification strategy need being considered carefully.

For change detection, training samples selection is to select the pixels in the changed and unchanged regions, based on the change features and the corresponding labels, hyperplane can be achieved which can separate the changed and unchanged features optimally in the sense of structural risk. Without doubt, training samples can be selected automatically. As mentioned above, the definition of 'change' is user-specific, it is difficult for the automatic training samples selection manner to capture the user's preference. For this reason, the training samples are selected by the user interaction in this paper.

Training samples selection is important for the classification accuracy. To reduce the dependence of classification accuracy on training samples selection, semi-supervised transductive SVM [9] is utilized. The role of transductive SVM is to consider the relation between the training samples and the test samples being considered. By this way, the user's preference is added and the classification accuracy is improved. Compared with the traditional SVM [10], the other advantage of semi-supervised transductive SVM is to tune the training samples and the hyperplane simultaneously, then the less representative samples are removed progressively.

Due to the limited discriminability of change features at one scale, the labels of the pixels whose classification results $w^T f$ are closed to zero are unreliable, where w and f represents the support vector and the change feature. In order to preserve the detail classification information, the change possibility maps $\{C^{(0)}, C^{(1)}, \cdots, C^{(N-1)}\}$ are created. The pixel $c_p^{(s)}$ at scale $C^{(s)}$ denotes the change probability of pixel $p^{(s)}$ and is obtained as follows:

$$c_p^{(s)} = \frac{1}{1 + \exp\left(A^{(s)} w^T f_p^{(s)} + B^{(s)}\right)} \tag{2}$$

where $f_p^{(s)}$ represents the change feature of pixel $c_p^{(s)}$.

$A^{(s)}$ and $B^{(s)}$ are learned by minimizing the negative log likelihood of the pixels $p^{(s)}$:

$$\min : -\sum_{p^{(s)}} t_p \log\left(c_p^{(s)}\right) + (1 - t_p)\log\left(1 - c_p^{(s)}\right) \tag{3}$$

where $t_p = \begin{cases} 1 & w^T f_p^{(s)} > 0 \\ 0 & w^T f_p^{(s)} < 0 \end{cases}$. The more detailed description of this step can be found in [11].

(a) Scale $C^{(s-1)}$ (b) Scale $C^{(s)}$

Fig. 3. Message propagation across scales.

2.3. Multi-scale Fusion

The goal of this step is to confirm the labels of all the pixels at the original scale according to the change probability pyramid $\left\{C^{(0)}, C^{(1)}, \cdots, C^{(N-1)}\right\}$. The label $l_{p^{(s)}}$ of pixel $p^{(s)}$ is selected, which minimizes the energy function $E\left(l_{p^{(s)}}\right)$ as follows:

$$\min: E\left(l_{p^{(s)}}\right) = \sum_{p^{(s)} \in I^{(s)}} D\left(l_{p^{(s)}}\right) + \sum_{\left(p^{(s)}, n^{(s)}\right)} S\left(l_{p^{(s)}}, l_{n^{(s)}}\right) \tag{4}$$

$E\left(l_{p^{(s)}}\right)$ is composed of the sum of label cost $D(\cdot)$ and the sum of the label discontinuity cost $S(\cdot)$ between four-connected pixels. In our task, function $D(\cdot)$ and $S(\cdot)$ are defined as follows:

$$\mathrm{D}\left(l_p^{(s)}\right) = \begin{cases} \left(1 - c_p^{(s)}\right)^2 & l = 1 \\ \left(c_p^{(s)}\right)^2 & l = -1 \end{cases}, \tag{5}$$

$$\mathrm{S}\left(l_p^{(s)}, l_n^{(s)}\right) = \begin{cases} 0 & l_p^{(s)} = l_n^{(s)} \\ 1 & l_p^{(s)} \neq l_n^{(s)} \end{cases}. \tag{6}$$

Eq. (4) can be solved by the loop belief propagation (BP) approach [12]. The label $l_{p^{(s)}}$ of $p^{(s)}$ is selected to minimize the following function $F(\cdot)$:

$$F\left(l_{p^{(s)}}\right) = D\left(l_{p^{(s)}}\right) + \sum_{n^{(s)} \in N_p^{(s)}} m^T_{n^{(s)} \to p^{(s)}}\left(l_{p^{(s)}}\right), \tag{7}$$

where $m^T_{n^{(s)} \to p^{(s)}}\left(l_{p^{(s)}}\right)$ is estimated by minimizing the following function:

$$S\left(l_{n^{(s)}}, l_{p^{(s)}}\right) + D\left(l_{p^{(s)}}\right) + \sum_{\substack{m^{(s)} \in N_n^{(s)}, \\ m^{(s)} \neq p^{(s)}}} m^{T-1}_{m^{(s)} \to n^{(s)}}\left(l_{n^{(s)}}\right). \tag{8}$$

$m^T_{n^{(s)} \to p^{(s)}}\left(l_{p^{(s)}}\right)$ defined in Eq. (7) is the message transfered from the pixel $n^{(s)}$ to the pixel $p^{(s)}$ at iteration T. $N_p^{(s)}$ and $N_n^{(s)}$ are pixel sets of four-connect pixels of $p^{(s)}$ and $n^{(s)}$.

From Eq. (8), message $m^T_{n^{(s)} \to p^{(s)}}\left(l_{p^{(s)}}\right)$ contains the change probability information. This change probability information at scale $C^{(s)}$ can be propagated to scale $C^{(s-1)}$. That is, the initial values $m^0_{n^{(s-1)} \to p^{(s-1)}}\left(l_{p^{(s-1)}}\right)$ shares the same value of $m^T_{n^{(s)} \to p^{(s)}}\left(l_{p^{(s)}}\right)$. As shown in Fig. 3, the messages from a to b, b to e, c to d and d to f (blue lines) in scale $C^{(s-1)}$ share the same value of the message from A to B at scale $C^{(s)}$. The change map is determined according to Eq. (7) at scale $C^{(0)}$.

Algorithm 1: Multi-scale Fusion Strategy

```
for s = N − 1 : −1 : 0 do
    if s=N-1 then
        m^0_{n(s)→p(s)}(l_{p(s)}) = 0;
    else
        Initializing m^0_{n(s)→p(s)}(l_{p(s)}) by
        m^T_{n(s+1)→p(s+1)}(l_{p(s+1)});
    end
    T = 0;
    repeat
        Updating m^T_{n(s)→p(s)}(l_{p(s)}) according to Eq. (8);
        T = T + 1;
    until E(l_{p(s)}) < σ;
end
```

3. EXPERIMENTAL RESULTS

To validate the effectiveness and reliability of the proposed approach, several experiments were carried out. For space limitation, only two datasets are discussed in this paper, the similar conclusions can be drawn from the other datasets. The images used in this paper were acquired over Beijing by QuickBird satellite in 2002 and 2003. The spatial resolution of the VHR images is about 0.7m/pixel, the sizes of the datasets are 1100×800 pixels and 1024×1024 pixels respectively.

3.1. Description of Experiments

(a) (b) (c) (d)

Fig. 4. (a) Image in 2002, (b) Image in 2003, (c) Reference change map, (d) Label image

The multi-temporal images of the first dataset are shown in Fig. 4(a) and Fig. 4(b). The reference change map is shown in Fig. 4(c), in which the white and black color represent the changed and unchanged areas respectively. As illustrated in Fig. 4(c), some areas were changed from the barely land to buildings. Except the above changed area, due to the season variation and view angle difference, the appearances of the same area of vegetation and building are very different. Since the false changes caused by the above impacts are not of interest, as shown in Fig. 4(d), they are selected as the unchanged samples and marked in green. The lands around the tall buildings at the bottom left should be taken as unchanged with respect to the spectral features, but they are undergone the structural changes simultaneously with the tall buildings nearby. For this reason, they are selected as the changed samples and marked in red.

Fig. 5(a) is the result achieved by S^3VM at the original scale.

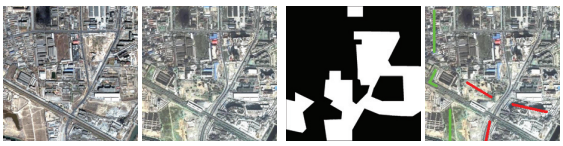
(a) (b) (c)

Fig. 5. (a) S^3VM, (b) S^3VM+BP, (c) The proposed approach.

Table 1. Performance comparison on the first dataset.

Technique	FA(pixel)	MA(pixel)	Err
S^3VM	238428	26458	30.0%
S^3VM+BP	129748	41166	19.4%
The proposed approach	22608	12644	5.0%

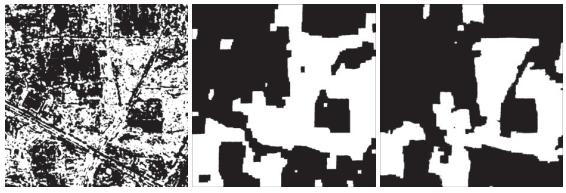
The pixel whose $w^T f$ is greater than zero is viewed as the changed area. Despite of the discrimination of local features and the robustness of change features, the missed alarm (MA) and the false alarm (FA) in Fig. 5(a) are very high. As can be illustrated by Fig. 5(b), the performance is improved significantly with the help of operating traditional multi-scale BP [12] on the result of S^3VM. This improvement is mainly due to the addition of the context information provided by BP. In spite of the advantages of S^3VM and BP in addressing the difficulties of VHR image change detection, there are still some missed alarms and false alarms. This demonstrates the insufficiency of one scale to capture the changes of different types. Fig. 5(c) is the result generated by the proposed approach. By taking the novel multi-scale fusion strategy, the result is much closer to the reference change map. As Table. 1 shows, the error rate is improved from 19.4% to 5.0%, which shows the importance of multi-scale fusion and the effectiveness of the proposed approach.



(a) (b) (c) (d)

Fig. 6. (a) Image in 2002, (b) Image in 2003, (c) Reference change map, (d) Label image.

As shown in Fig. 6(a) and Fig. 6(b), the second dataset contains more complex changed and unchanged features than the first one. Due to the seasonal variation, the color of vegetation areas is changed from yellow to green. Real changes are mixed up with some false changes (e.g., different roof appearances of the same buildings), and such false changes are very difficult to be removed. In addition, the features of objects are obscured by the dark shadows. To keep the real changes of interest and remove the disturbed false changes, in selecting training samples, two vegetation areas with different colors and the upper left area with different roof appearances are marked as the unchanged samples, and two areas about the building changes are marked as the changed samples. Fig. 7 are the results generated by S^3VM, S^3VM combining with traditional multi-scale BP and the proposed approach respectively. Although this dataset is very challenging, the error of the proposed approach reaches 9.2%, which is 6.8% and 20.8% better than the other two approaches. By the proposed approach, the complex structural changes are captured, and most of the false changes caused by view angle variation are removed correctly.



(a) (b) (c)

Fig. 7. (a) S^3VM, (b) S^3VM+BP, (c) The proposed approach.

Table 2. Performance comparison on the second dataset.

Technique	FA(pixel)	MA(pixel)	Err
S^3VM	219943	97244	30.0%
S^3VM+BP	138375	30241	16.0%
The proposed approach	36728	59755	9.2%

The above qualitative and quantitative comparisons demonstrate the effectiveness and reliability of the proposed approach in addressing the difficulties of VHR images. Of course, the effectiveness and reliability rely on the combination of the discriminative and robust change feature extraction, effective classification and multi-scale fusion strategy. In other words, the promising performance can not be achieved without either component mentioned above.

4. CONCLUSION

A novel approach is proposed for VHR image change detection. This approach integrates the discriminative local features and a robust distance metric to represent the complex change in urban VHR images. The effective classifier is utilized to remove the false changes and to capture the interested ones. Finally, the result is generated by a novel multi-scale fusion based on belief propagation to reduce the dependence of changes on the scale. Despite of the effectiveness of the proposed approach, many aspects need to be considered in the future work, including object-specific feature extraction, robust distance measure to large view angle variation and so on.

5. REFERENCES

[1] F. Bovolo and L. Bruzzone, "A multilevel parcel-based approach to change detection in very high resolution multitemporal images," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 1, pp. 33–37, 2009.

[2] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *International Journal of Remote Sensing*, vol. 10, no. 6, pp. 989–1003, 1989.

[3] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 858–867, 1997.

[4] L. Bruzzone and D. F. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 452–466, 2002.

[5] V. Michele, T. Devis, B. Francesca, K. Mikhail, and B. Lorenzo, "Supervised change detection in vhr images using contextual information and support vector machines," *International Journal of Applied Earth Observation and Geoinformation*, vol. 20, no. 0, pp. 77–85, 2013.

[6] C. Huo, Z. Zhou, H. Lu, C. Pan, and K. Chen, "Fast object-level change detection for vhr images," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 1, pp. 118–122, 2010.

[7] T. Celik, "A bayesian approach to unsupervised multiscale change detection in synthetic aperture radar images," *Signal Processing*, vol. 90, no. 5, pp. 1471–1485, 2010.

[8] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. Freeman, "Sift flow: Dense correspondence across different scenes," in *European Conference on Computer Vision*, vol. 5304, pp. 28–42. 2008.

[9] I. S. Reddy, S. Shevade, and M. N. Murty, "A fast quasi-newton method for semi-supervised svm," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2305–2313, 2011.

[10] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144–152.

[11] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in Large Margin Classifiers*, 1999, pp. 61–74.

[12] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 41–54, 2006.