SELECTING FEATURES FOR URBAN CHANGE DETECTION

Chunlei Huo a, *, Zhixin Zhou b, a, Keming Chen a, Hanqing Lu a, Jian Chenga

[a] National Laboratory of Pattern Recognition, Institute of Automation, CAS, Beijing, 100190 – (clhuo, kmchen, luhq, jcheng)@nlpr.ia.ac.cn
[b] Beijing Institute of Remote Sensing, Beijing, 100854 – zhixin.zhou@ia.ac.cn

Commission VII, WG VII /5

KEY WORDS: Urban, Change Detection, High Resolution Images, Feature Selection, Machine Learning.

ABSTRACT:

In machine learning, the preprocessing of the observations and the resulting features are one of the most important factors for the performance of the final system. In this paper, a robust approach to urban change detection for high resolution images is presented based on feature selection and machine learning. The rationale of the proposed approach is to improve the interclass variability by extracting change features of different types at different scales, to choose the informative change features by feature selection, to achieve the reliable results by machine learning. By taking advantages of feature selection and machine learning, the proposed approach is superior to the related methods in accuracy, efficiency and robustness. Experiments demonstrate the effectiveness and advantage of the proposed approach.

1. MANUSCRIPT

Urban change detection is very important in urban monitoring such as the detection of new buildings or the discovery of modifications in the existing ones. With the development of high resolution sensors, it is possible to detect changes at the smaller spatial scale. However, urban change detection of high resolution images is more difficult due to the intrinsic complexity of high resolution images and the complex nature of the urban environments. First, the improvement in spatial resolution simplifies the problem of mixed pixels encountered in the standard multispectral images, but at the same time, it increases the internal spectral variability (intraclass variability) of each land cover class and decreases the spectral variability between different classes (interclass variability) (Carleer et al., 2004). The resulting high intraclass and low interclass variability lead to the reduction in the statistical separability of the changed class from the unchanged class, which in turn brings out high change detection errors. Second, the fact that urban areas are composed of numerous materials (concrete, asphalt, metal, plastic, glass, shingles, water, grass, shrubs, trees and soil) makes it difficult to extract discriminative features to describe and separate the complex changes. In addition, the limited spectral resolution of high resolution sensors, which depends on the technical constraints, further increases the complexity of this problem. Third, most of the traditional change detection methods derive the “optimal” threshold under the assumption that both the changed class and the unchanged class are subject to Normal distributions. Unfortunately, the assumption of Normal distribution is always violated especially for changed class (Bovolo and Bruzzone, 2007). As a result, the optimal change map cannot be achieved even by the optimal threshold. Furthermore, based on threshold selection, it is difficult for the traditional methods to take into account the specific requirements peculiar to the specific end-users (Paul and Alessandro, 2000).

To address the above difficulties, in this paper, we propose a robust approach to urban change detection for high resolution images based on feature selection and machine learning. By taking advantages of feature selection, the reduced number of features not only combats the curse of dimensionality but also leads to a reduced computational complexity. Furthermore, only the representative subset is determined from a given set of features for the change detection task. As a result, the discriminability between the changed class and the unchanged class is improved. With the help of SVM, a distribution-free classifier, change features of different types at different scales are combined implicitly and optimally to the specific end-user. The paper is organized in four sections. Section 2 presents a detailed description of the proposed approach step by step. Section 3 reports the experimental results obtained on real QuickBird images. Finally, section 4 draws the conclusions of this paper.

2. CHANGE DETECTION BASED ON FEATURE SELECTION

The basic idea of the proposed approach is to improve the interclass variability by extracting change features of different types at different scales, to choose the informative change features by feature selection, to achieve the reliable results by machine learning. To this end, three modules are designed in the proposed approach: feature extraction module, feature selection module and change detection module. In the first module, change features of different types at different scales are extracted individually. In the second module, informative change features are identified based on feature selection. Reliable changes are detected in the third module based on machine learning, which provides a natural way of combining the informative features of different kinds.

* Corresponding author.

2.1 Feature Extraction

Various features and similarity measures are presented in the literature to represent the changes, among which spatial-contextual features are important since each pixel is likely to be surrounded by the pixels belonging to the same class. For the high resolution images we considered, in this paper, two simple yet effective features are used. One is the spectral difference, the other is local structural features, or texture.

2.1.1 Spectral Difference: Let $d_i(p)$ be the spectral difference between two images at pixel p computed at the ith band. In this paper, we use $d(p) = \max_{1 \le i \le K} d_i(p)$ to denote the final spectral difference at p, which combines the spectral differences at different channels, K is the number of the spectral band. To be robust to noise, the spectral difference measure is normalized into the range of $[0,1]$ by the following slope function:

$$d_s(p) = \begin{cases} 1 & : \ |d(p)| > 2T \\ |d(p)|/(2T) & : \ otherwise \end{cases} \tag{1}$$

The parameter T should be properly chosen to cope with image noise. Since the image noise can be modeled as Gaussian noise $N(0,\sigma)$, the noise in the difference image has a Gaussian distribution $N(0,\sigma_d)$ with $\sigma_d = \sqrt{2}\sigma$. However, due to the effect of illumination changes, the shifts of spectral values for unchanged regions should be compensated to compute σ_d. First, the shift of gray value at each pixel is calculated as

$$\overline{d}(p) = \frac{1}{N^2} \sum_{q \in \delta(p,N)} d(q). \tag{2}$$

Where $\delta(p,N)$ denotes the $N \times N$ spatial neighborhood centered at p. Then the difference image becomes

$$d' = d - \overline{d}. \tag{3}$$

By estimating the variance of noise to d', σ_d can then be estimated. The average shift of the spectral difference for the unchanged regions is estimated as

$$\overline{d_u} = \frac{1}{\|\Omega\|} \sum_{q \in \Omega} \overline{d}(q). \tag{4}$$

With $\Omega = \{p : |d'(p)| < 2\sigma_d$ and $|\overline{d}(p)| < T_{med}\}$, where T_{med} is the median of $\overline{d}$ and $\|\Omega\|$ is the number of pixels in the set Ω, most pixels of which come from the unchanged regions. From these estimations, the parameter T can be chosen as $|\overline{d_u}| + 3\sigma_d$.

2.1.2 Textural Difference: Texture is an important feature since it represents the spatial arrangement about the gray levels of pixels in an object. Among different texture features, local texture based on the gradient value is used in this paper since it describes how the gray level changes within a neighborhood and is less sensitive to illumination changes.

Let $f_i(p)$ be the ith image and $f_i'(p) = [f_i^x(p) \quad f_i^y(p)]$ the gradient vector at the pixel p. The texture difference between two images at p is defined based on the contextual distance:

$$d_t(p) = 1 - \frac{\sum_{q \in R} 2C_{12}(q)}{\sum_{q \in R}(C_{11}(q) + C_{22}(q))}. \tag{5}$$

Where $C_{ij}(q)$ is the cross-correlation of gradient vectors of two images at q. R is the $N \times N$ spatial neighborhood. As illustrated in (Li and Leung, 2002), such difference measure can describe the difference within the local region, both the difference in magnitude and the difference in angle. What's more, it is robust to illumination changes and noise.

2.1.3 Multiscale Change Feature Extraction: Despite of the importance of spatial-contextual features, the problem is the choice of the window size (or the scale) because an inappropriate window size will produce miss- and over-detections. One way to overcome this problem is to extract multiscale change features from a set of windows with different sizes. Cognitive-based studies suggest that human photointerpretation requires windows considerably larger than those employed in digital image classification studies (Binaghi et al., 2003). This is confirmed by the experiments conducted by (Hodgson, 1998), which show that the photointerpreters select windows of about 40 × 40 pixels (corresponding to about 60m × 60m) to identify urban patterns in high spatial resolution images. For this reason, the above contextual difference features are extracted in the concentric windows with increasing sizes. The number and size of windows depend on the specific characteristics of the target compared with the other patterns within the image. Considering the image resolution used in this paper, the change features are extracted with windows of size r ranging from 5×5 pixels to 51×51 pixels. Due to the high intraclass and low interclass variability of high resolution data, only the relative change information is not discriminative enough to represent the complex changes between images. For this reason, we analyze the original spectral signatures and the relative change information jointly. The multiscale change features can then be expressed as follows:

$$F(p) = [F_o(p), F_r(p)]. \quad (6)$$

Where

$$F_o(p) = [f_1^1(p), \ldots, f_1^K(p), f_2^1(p), \ldots, f_2^K(p)] \quad (7)$$

and

$$F_r(p) = [d_s^{r_1}(p), \ldots, d_s^{r_{24}}(p), d_t^{r_1}(p), \ldots, d_t^{r_{24}}(p)]. \quad (8)$$

Since the dimensionality of the resulting multiscale change features is high and many entries contained in F_r may contain redundant information, the next step aims to figure out which of these features are best suited for the change detection task.

2.2 Feature Selection

For supervised change detection approach, a known problem is to find ways to reduce the dimensionality of the feature space to overcome the risk of "overfitting". In the literature, different feature selection techniques are presented for the different applications. Considering the relevant and redundant properties of the multiscale change features, in this paper, we apply the feature selection procedure based on the combination of the Fisher criterion and principal feature analysis (PFA) (Lu et al., 2007), which is motivated by the following facts. First, the Fisher criterion is a supervised criterion and it can remove the features which are noisy or irrelevant, but it does not take the redundancy of features into account. For example, if two features are entirely same and both have high the Fisher values, they will be both selected with high redundancy. Second, PFA is unsupervised, it explores the correlation between features and removes the redundant ones, but it cannot distinguish the noisy features from the relevant ones.

2.2.1 Initial Feature Selection Based on Fisher Criterion: Given n_i training samples $\{x_1^i, x_2^i, \ldots, x_{n_i}^i\}$ for each class ($i = 2$, i.e., changed class and unchanged class), the class separability of a feature set can then be measured by

$$J_F = trace(S_w^{-1} S_b). \quad (9)$$

Where the within-class scatter matrix S_w and between-class scatter matrix S_b are estimated by

$$S_w = \sum_{i=1}^{2} P_i \hat{S}_i. \quad (10)$$

and

$$S_b = \sum_{i=1}^{2} P_i (\hat{\mu}_i - \hat{\mu})(\hat{\mu}_i - \hat{\mu})^T. \quad (11)$$

The priori probability of class i is estimated by $P_i = \frac{n_i}{\sum_{i=1}^{2} n_i}$, the class means μ_i are estimated by $\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j^i$, and the gross mean μ is estimated by $\hat{\mu} = \sum_{i=1}^{2} P_i \hat{\mu}_i$.

This measure serves a good criterion for feature subset selection, and has shown superior performance in many practical problems. However, its calculation for a large number of features is computationally expensive. Instead, the Fisher criterion for one single feature has been prevalently used to select the discriminant features. For the kth feature, it is calculated by

$$Fisher(k) = \frac{S_b^k}{S_w^k}. \quad (12)$$

Where S_b^k and S_w^k are the kth diagonal element of S_b and S_w respectively, and they can be calculated from the data of single feature. For initial feature selection, we calculate the Fisher criterion of each feature, order the features in the decreasing order of criterion values, and simply select the features with the maximum values, while the features with very small Fisher values are abandoned. Though the single-feature Fisher criterion does not consider the joint separability of multiple features, it is able to retain all discriminant features by only removing irrelevant and noisy features, for which the Fisher criterion is nearly zero.

2.2.2 Feature Selection Based on PFA: PFA exploits the structure of principal components to choose the principal features, which retains most of the information both in the sense of the maximum variability in low-dimensional subspace and in the sense of minimizing the reconstruction error. By PFA, the representative feature subset can be found as follows:

1) Compute the sample covariance matrix Σ for the n-dimension feature vector X.

2) Compute the principal components and eigenvalues of the covariance. i.e., $\Sigma = A \Lambda A^T$, A is the orthonormal matrix composed of the eigenvectors of Σ, and $\Lambda = diag[\lambda_1, \lambda_2, \ldots, \lambda_n]$ with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.

3) Choose the subspace dimension q and construct the matrix A_q from A. This can be chosen according to the desired variability of the data to be retained. The retained variability is the ratio between the sum of the first q eigenvalues and the sum of all eigenvalues. This ratio is empirically set to 90%.

4) Cluster the vectors $|V_1|, |V_2|, \ldots, |V_n|$ to p clusters using the $k-\text{means}$ algorithm, $|\cdot|$ denotes the vector composed of the abstract values of all elements.

5) From each cluster, find the corresponding vector $|V_i|$, which is the closest in Euclidean distance to the mean of the cluster. Choose the corresponding feature, x_i , as a principal feature. This step will yield the choice of p principal features.

2.2.3 Combined Feature Selection Based on Fisher Criterion and PFA: To remove both irrelevant and redundant features, we combine the Fisher criterion and principal feature analysis. The single-feature Fisher criterion is used as the pre-selection criterion to select the best m individual discriminant features. Then PFA is used to cluster the m pre-selected features into p groups, and one feature from each group forms a subset of principal features. In pre-selection by the Fisher criterion, to decide the number of retained features, we accumulate the Fisher values in decreasing order until the sum of values exceeds a pre-specified percentage of the total values. Since we apply the above feature selection procedure only to the multiscale “relative” change feature F_r, the final change features can be represented as:

$$F(p) = [F_o(p), F_r^{'}(p)]. \qquad (13)$$

Where $F_r^{'}$ is the subset selected from F_r by the above procedure.

3. EXPERIMENTAL RESULTS

To assess the effectiveness of the proposed approach, experiments were conducted on high resolution images of many scenes. For space limitation, only the results on one data set are shown in this paper. The data set is composed of three channels (red, green and blue) images of 401 × 701 pixels (61 cm/pixel). The data set is taken over Beijing (China) acquired by QuickBird in 2002 and 2003. The selected test sites are shown in Fig. 1.

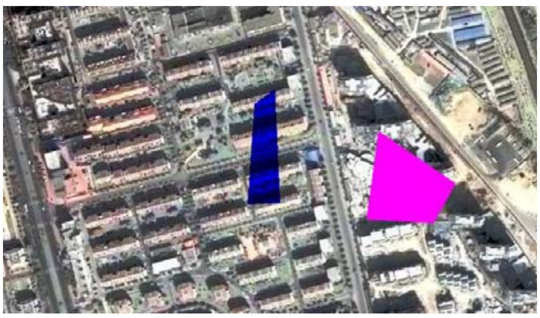
Fig. 2(a) shows the ground truth labeled manually, in which the changed class is shown in red. Training regions are shown in Fig. 2(b) (the changed regions: red, the unchanged regions: blue). Fig. 2(c) is the result by the proposed approach. To demonstrate the effectiveness of the proposed approach, the change detection results by the other three different methods are also shown. Fig.2 (d) is the result without feature selection, while Fig.2 (e) (f) show the results with feature selection based on Fisher criterion and PFA respectively. By comparing Fig.2 (c)-(f) with Fig.2 (a), most changes are detected correctly based on different methods, even without feature selection, this indicates the effectiveness of the multiscale features in improving the discriminability of the changed class from the unchanged class. To illustrate the advantages of feature selection for change detection, we use Missed Alarms, False Alarms and Overall Alarms to compare the performances of different methods. The performances by the different methods are listed in Tab. 1, from which we can conclude that Overall.

(a)

(b)

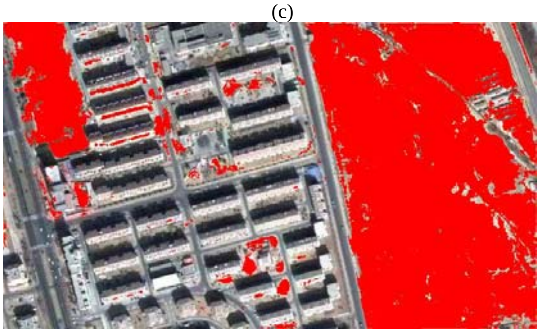
Figure 1: Images used in this paper. (a): image acquired at 2002,(b): image acquired at 2003.

(a)



(b)

(c)



(d)

(e)

(f)

Figure 2: Results comparison against different methods. (a): reference change map. (b): training samples. (c): change map by the proposed approach. (d): change map by SVM without feature selection. (e): change map by SVM and Fisher feature selection. (f): change map by SVM and PFA feature selection.

Accuracy	SVM	Fisher+PFA	Fisher	PFA
Missed alarms	14026	11306	12275	13386
False Alarms	10582	12070	11592	10832
Overall Alarms	24608	23376	23867	24218

Table 1: Accuracy comparison against different feature selection strategies

Alarms by feature selection based on the Fisher criterion and/or PFA are smaller than the method without feature selection, which shows the importance of feature selection in improving change-detection accuracy. Feature selection based on Fisher criterion can improve the performance since it can remove noisy or irrelevant features. Feature selection based on PFA can improve the performance as it can remove redundant features. Among three different feature selection strategies, Overall Alarms by the combination of Fisher criterion and PFA is the smallest. This difference implies the effectiveness of the proposed approach. Such effectiveness lies in the ability of the proposed approach in removing both irrelevant and redundant features.

Accuracy	SVM	Fisher+PFA	Fisher	PFA
Missed alarms	10649	14696	13111	12019
False Alarms	20044	10835	15104	15016
Overall Alarms	30693	25531	28215	27035

Table 2: Robustness comparison against different feature selection strategies

To evaluate the robustness of the proposed approach, we apply a conformal affine transformation (rotation: 1^{o}, translation: 1 pixel) on Fig.1 (b) to simulate the miseffects of view-angle variation and misregistration. The results by different methods are shown in Fig. 3, the corresponding performances are listed in Tab.2. In this case, Overall Alarms by different methods are larger than the first experiment. Compared with the result based on SVM without feature selection, Overall Alarms by the methods with feature selection are smaller. This comparison shows the importance of feature selection in reducing the noise caused by view-angle variation and misregistration. Among the three different methods with feature selection, Overall Alarms by Fisher+PFA is still the smallest, which indicates the robustness of the proposed approach. In detail, False Alarms by the three methods other than Fisher+PFA is larger than the first experiment since the land surface properties at wrong locations are evaluated instead of real changes at the same location between one time and another. However, the proposed approach is hardly affected by view-angle variation and misregistration, this can also be confirmed by comparing Fig. 3(a) with Fig.2(c). The underlying reason is that the complementation of Fisher and PFA is helpful in improving the robustness. As a final remark, it is worth noting that the number of features selected is not the same for the two set of experiments since the features are selected automatically based on the ratio of the first m Fisher values and/or the first p eigenvalues.

4. CONCLUSIONS

Conventional pixel-based change detection schemes are less suited for urban change detection of high resolution images due to the resulting high intraclass and low interclass variability as well as complex man-made objects. In this paper, a robust approach is proposed for urban change detection of high resolution images, the main innovation lies in the discriminative multiscale feature extraction, feature selection as well as machine learning. The experiment results reported in this paper confirm the effectiveness of the proposed approach. Despite of the promising preliminary results, many future developments need be considered to make our approach more robust and more perfect.

(a)

(b)

(c)

(d)

Figure 3: Robustness comparison against different methods. (a): change map by the proposed approach. (b): change map by SVM without feature selection. (c): change map by SVM and Fisher feature selection. (d): change map by SVM and PFA feature selection.

ACKNOWLEDGEMENTS

This work was supported by Natural Science Foundation of China (60121302, 60605004).

REFERENCES

Binaghi, E., Gallo, I. and Pepe, M., 2003. A cognitive pyramid for contextual classification of remote sensing images. IEEE Trans. Geoscience and Remote Sensing, 41(12), pp. 2906–2922.

Bovolo, F. and Bruzzone, L., 2007. A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain. IEEE Trans. Geoscience and Remote Sensing, 45(1), pp. 218–236.

Carleer, A., Debeir, O. and Wolff, E., 2004. Comparison of very high spatial resolution satellite image segmentations. SPIE Conf. Image and Signal Processing Remote Sensing IX 5238, pp. 532– 542.

Hodgson, M. E., 1998. What size window for image classification? a cognitive perspective. Photogramm. Eng Remote Sensing, 64(8), pp. 797–807.

Li, L. and Leung, M., 2002. Integrating intensity and texture differences for robust change detection. IEEE Transaction on Image Processing 11(2), pp. 105–112.

Lu, Y., Cohen, I., Zhou, X. S. and Tian, Q., 2007. Feature selection using principal feature analysis. in Proc. 15th international conference on Multimedia, pp. 301–304.

Paul, C. and Alessandro, A., 2000. Toward specification-driven change detection. IEEE Trans. Geoscience and Remote Sensing , 38(3), pp. 1484–1488.