COMBINING LOCAL FEATURES AND PROGRESSIVE SUPPORT VECTOR MACHINE FOR URBAN CHANGE DETECTION OF VHR IMAGES

Chunlei Huo[1], Bin Fan[1], Chunhong Pan[1], Zhixin Zhou[2]

1. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, 100190.
2. Beijing Institute of Remote Sensing, Beijing, China, 100854.

Commission VII/5

KEY WORDS: change detection, local features, change blindness, cognitive mechanisms, progressive transductive SVM

ABSTRACT:

The difficulties about change detection of VHR images are analyzed from different perspectives. Motivated by perception and cognition mechanism of human vision, visual change detection principles are discussed, and a unified change detection framework is proposed. To address the difficulties in change detection of VHR images, a novel approach is presented within the framework, which exploits the combination of local features and change vector displacement field to represent the complex changes of VHR images and utilizes transductive SVM(Support Vector Machine) to classify change features progressively. Experiments demonstrate the effectiveness of the proposed approach.

1 INTRODUCTION

"Everything changes but changes itself"(Kennedy). With the globalization and geographic expansion of human activities, understanding changes becomes increasingly important. Change detection is the apprehension of changes in the world around us. Remote sensing image change detection aims at detecting changes by comparing multiple images of the same scene taken at different times, and it is a powerful tool that can be used in a diverse range of applications such as disaster management, ecosystem monitoring, military surveillance, and so on. With the development of VHR (Very High Resolution) satellites, change detection receives more extensive attention since it can detect the changes at the more detailed spatial scale. However, compared with low-to-moderate resolution remote sensing images, change detection of VHR images is more challenging since the basic premise of change detection(Singh, 1989)(i.e., changes in land cover must result in changes in radiance values and changes in radiance due to land cover change must be large with respect to radiance changes caused by other factors) is broken by the complexities of such data. In detail, the difficulties lie in the following factors:

First, the difficulties lie in the intrinsic complexity of VHR image. The employment of sensors with the improved spatial resolution simplifies the problem of mixed pixels, however, the internal variability within homogenous land-cover classes increases. At the same time, the increased internal variability decreases the statistical separability between different land-cover classes in the spectral data space. The resulting high internal variability and low spectral separability lead to the reduction of the statistical separability between the changed class and the unchanged class. For this reason, traditional change detection approaches(Coppin et al., 2004, Lu et al., 2004, Radke et al., 2005) are difficult to be applied to VHR images without considering the complexities of such data. For the same reason, some key techniques such as image segmentation and image classification are not mature for VHR images, which hamper the digital change detection techniques.

Second, the difficulties lie in the incomprehensive understanding of human visual change detection mechanism. The ability to detect change is important in much of our everyday life. In spite of the pervasiveness of change detection in our lives, it has proven surprisingly difficult to study, and only recently have various approaches begun to converge in terms of what it is and how it is carried out(Rensink, 2002, Simons and Rensink, 2005). The effective change detection algorithms can be designed by imitating recognition principles of human beings, and the computer is only a tool to accelerate the computation. The incomprehensive understanding of visual change detection mechanism hinders the development of change detection techniques.

Last but not least, the difficulties lie in the "human-machine" gap. The gap is caused by the difference between 3-d real world and 2-d digital images, the difference between human eyes and satellite sensors, the difference between human brains and computers. The other cause is the ambiguous definition of "change", i.e., the definition of change is application-specific, task-specific and user-specific(Paul and Alessandro, 2000). The "human-machine" gap makes the change detection an ill-posed problem, and it is more troublesome to be solved by the computers.

Despite the importance of the existing review papers for developing new change detection approaches, it is difficult to analyze the change detection techniques within a general framework. For this reason, a unified change detection framework is presented, from which most of the existing approaches can be generated. With the help of the proposed framework, a novel approach is presented to overcome the above difficulties encountered in urban change detection of VHR images. Compared to the related work, the contributions of this paper lie in the investigation of change detection framework based on visual change detection principles, as well as the derived approach to address the above difficulties.

The paper is organized in five sections. Section 2 describes the proposed framework. Section 3 presents a detailed description of the proposed approach step by step. Section 4 reports the experimental results obtained on real QuickBird images. Finally, section 5 draws the conclusions.

2 THE UNIFIED CHANGE DETECTION FRAMEWORK

The disability to detect changes visually is called change blindness, recent research(Simons and Rensink, 2005) indicates change

blindness cannot be avoid unless the following three requirements are met simultaneously:

1) The objects being observed must be encoded, and the encoded features must be kept in mind.

2) The features encoded before and after changes occur must be compared.

3) The feature difference must be recognized by the observers.

The above change blindness principles indicate that a digital change detection algorithm should consist of at least three parts: feature space, distance space and search space. For digital change detection techniques implemented by the computer, the procedure to simulate human vision system and cognitive mechanisms to search the changes in a computable manner is very necessary. In consequence, an effective digital change detection algorithm can be divided into the following four components:

(1) Feature space, F.

The feature space determines the place where the co-registered images will be compared.

(2) Distance space, D.

The distance space provides the way how the difference between images is measured quantitatively. The combination of the feature space and the distance space is the change feature space.

(3) Search space, S and object function, f.

Given the feature space and the distance space, the search space decides the change maps by the object function. The object function provides the link between the change features and the real change detection result (which cannot usually be observed) or the approximated optimal change detection result, and it can be designed based on certain criterions such as the minimum error rate and the minimum risk.

(4) Search strategy O.

There may be many solutions that can minimize the object function, among which the optimal solution can be achieved efficiently by the search strategy.

For co-registered images I_1 and I_2, the change detection problem can then be formulated as the following unified framework:

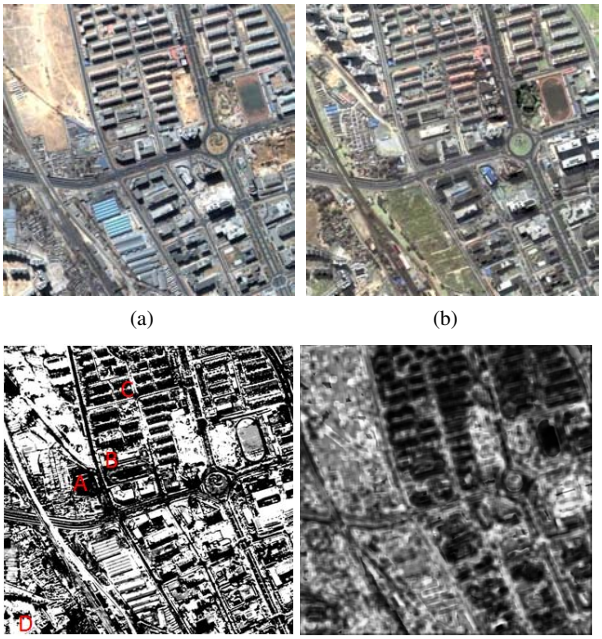
$$arg \min_{s \in S} f(D(F(I_1), F(I_2), s)). \quad (1)$$

The selection of each component is determined by the types of images to be compared. For example, spectral and structural features can be used in the feature space, and these features can be extracted in pixel- or region-based manner based on the image resolution (low-to-moderate resolution or high resolution). Based on the types of the sensors (SAR or the optical sensor), the distance can be described by difference-based or ratio-based approach.

The proposed change detection framework has the following advantages:

(1)It captures human visual change detection principles as well as the difference between visual change detection and digital change detection.

(2)It is powerful in understanding and analyzing the existing approaches. Most of the existing change detection methods(Coppin et al., 2004, Lu et al., 2004, Radke et al., 2005) can be included by this framework.



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

Figure 1: Illustration of difficulties in change detection of VHR images. (a) and (b):a multi-temporal image pair. (c)change features based on the pixel-wise difference. (d)change features based on the SIFT descriptor difference.

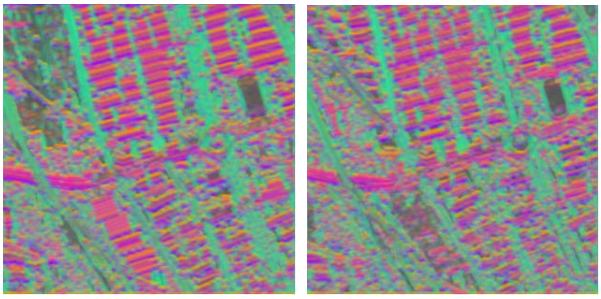
The aim of this paper is to design a new desirable change detection approach for VHR images within the above framework.

3 THE PROPOSED ALGORITHM

The focus of this paper is urban change detection of VHR images. For such data, besides the common difficulties of VHR images, the complex morphology caused by human activities must be taken into consideration. A pair of VHR images is shown in Fig.1, some significant changes can be detected with ease by human visual comparison, but some subtle changes are difficult to be detected even by the experienced experts after repeated comparisons. This indicates that the ability of human visual change detection had been overestimated. As shown by the region B, C and D in Fig.1(c), many false changes are caused by Sun angle variation, shadow and seasonal change, human vision system is robust to such changes, however, it is very hard for the computers. The rationale of the proposed approach is to reduce the missed alarms based on change blindness principles and reduce false alarms by simulating human vision system. In detail, within the proposed framework, we aim to tackle the difficulties of VHR image change detection from the viewpoint of cognitive psychology, i.e., to extracting discriminative local features to represent the complex objects, to measure the difference between objects by the robust metric which takes local nonlinear displacement into consideration, to classify the change features in a progressive fashion. Below we elaborate each component step by step.

3.1 Feature space

As stated above, the pure usage of spectral features is too simple for complex urban areas. Take the region A in Fig.1(c) as an example, the structural changes can be detected by human vision, but the compute is ”bind” to such changes. In this paper,



(a) (b)

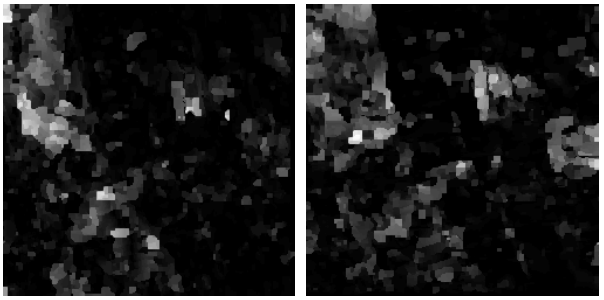
Figure 2: SIFT feature space for VHR images. (a)SIFT feature for Fig. 1(a). (b) SIFT feature for Fig. 1(b).

SIFT descriptor(Lowe, 2004), i.e., HOG(Histogram of Oriented Gradients), is used as our feature space. In other words, SIFT descriptor is extracted at each pixel to characterize local image structures and encode the contextual information. SIFT is a local descriptor to characterize local gradient information. In (Lowe, 2004), SIFT descriptor is a sparse feature representation that consists of both feature detection and description. In this paper, however, we only use the feature description component. For every pixel in an image, we divide its neighborhood (e.g. 16×16) into a 4×4 cell array, quantize the orientation into 8 bins in each cell, and obtain a 128-d vector as the SIFT representation for a pixel. We call this per-pixel SIFT descriptor SIFT image. 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, this is can be induced from Fig.2 (a) and (b), where the top 3 components of SIFT descriptor after PCA transformation are shown visually. Despite the discriminability of SIFT descriptor, as shown in Fig.1 (d), the algebra difference between SIFT descriptors is too simple to represent the complex changes in VHR images, so the next task is to design a robust distance metric to measure the difference between discriminative local features.

3.2 Distance space

Recently, SIFT has been widely used in image matching and object recognition, and most of the literatures compare SIFT descriptors in Euclidean distance space(e.g., the ratio of the nearest neighbors). However, such pixel-wise comparison is not suitable for VHR images. Even if the multi-temporal images are co-registered, local small displacements caused by Sun angle variation or the registration error are inevitable and difficult to remove. In consequence, a robust distance metric is needed which considers the local displacement adaptively to each pixel.

For the same objects taken under different Sun angles, human beings can make the correct decision that no changes happen. The underlying reason is that the appearances of the objects under different Sun angles are regarded to be very similar by human vision system as long as the local displacement is within certain ranges. Based on this observation, for a pixel p, the local displacement can be computed by searching the smallest distance within a neighborhood that makes the similarities of SIFT descriptors between the regions around it and the shifted ones in the other images maximized. Given the local displacement specific to each pixel, the distance can be re-computed more accurately by their SIFT descriptors with known displacement. The alternative choice is to represent the difference between two SIFT descriptors by the displacement directly. For a co-registered image pair, if there is no changes happened at the pixel p and its neighbors, their SIFT descriptors should be very close, and the local displacements computed by the above principle should be small. And the reverse is also true, large local displacement means changes of high probability. In fact, the latter choice is simpler and is adopted in this paper.



(a) (b)

Figure 3: Illustration of the distance space presented in this paper. (a)change vector field in horizontal direction. (b)change vector field in vertical direction.

Now, we describe the proposed distance metric in detail. Let $w(p) = (u(p), v(p))$ be the local displacement at p, $s_i(p)$ the SIFT descriptor extracted form I_i at p, the problem of comparing $s_1(p)$ and $s_2(p)$ is converted to find $w(p)$ which makes the following equation minimized:

$$E(w) = \sum_p (\|s_1(p) - s_2(p + w(p))\|_1, t) + \quad (2)$$

$$\sum_p \phi(|u(p)| + |v(p)|) + \quad (3)$$

$$\sum_{(p,q)\in\mathcal{N}} (\alpha(|u(p) - u(q)|, d) + (\alpha(|v(p) - v(q)|, d))) \quad (4)$$

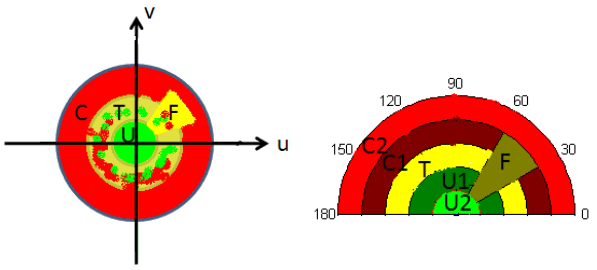
The above equation contains a data term, small displacement term and smoothness term (i.e., spatial regularization). The data term in Eqn. (2) constrains the SIFT descriptors to be compared along with the displacement w(p). The small displacement term in Eqn. (3) constrains the displacement vectors to be as small as possible when no other information is available. The smoothness term in Eqn. (4) constrains the displacement vectors of adjacent pixels to be similar. t and d are the thresholds. In order to reduce the complexity by distance transform function(Felzenszwalb and Huttenlocher, 2006) and achieve better convergence by sequential belief propagation (BP-S) (Szeliski et al., 2008), in the objective function, truncated L1 norms are used in both the data term and the smoothness term.

Fig.3(a) and Fig.3(b) show the horizontal and vertical displacements visually. Compared to Fig.1(c) and Fig.1(d), the change information contained in Fig.3(a) and Fig.3(b) is more discriminative. After achieving $w(p)$, the probability of changes at p can be represented by the magnitude and orientation of w. The next step is to classify the above 2-d change feature to distinguish the changed class from the unchanged class.

3.3 Search space

To design a desirable search space, it is necessary to analyze the cause and characteristic of the local nonlinear displacement. The local displacement is mainly caused by the real changes or the false changes caused by the impacts such as Sun angle variation and the registration error. In detail, the magnitudes of displacement vector corresponding to real changes are large, and the orientations of displacement vector are out of order. In contrast, the magnitudes of displacement vector corresponding to the unchanged class are small, the orientations of displacement vector

caused by Sun angle variation are of the same direction. Different types of changes are visually shown in Fig.4(a), where C,U and F denotes the changed class, the unchanged class and false changes caused by Sun angle variation respectively. T are change features to be classified. In fact, the displacement orientation caused by Sun angle variation is closely related to the orientation of shadow. As shown in Fig. 5, for the images after being rectified, the direction of local displacement caused by Sun angle variation is vertical(or with slight variation). In other words, if the displacements are large in both horizontal and vertical directions, the probability of the corresponding pixel is changed is very high; if the displacement is large in only horizontal or vertical direction, the probability of changes caused by Sun angle variation is very high. For a pixel p , the corresponding change vector displacement $w(p) = [u(p), v(p)]$ and the magnitude $m(p) = \sqrt{u(p)^2 + u(p)^2}$, we define the above three change types as follows: $C = \{p|m(p) \geq \tau_1, s(p) \geq \tau_4\}$, $U = \{p|m(p) \leq \tau_2\}$, $F = \{p|m(p) \leq \tau_3, s(p) < \tau_4\}$. $s(p)$ is to represent the degree and the direction(along the horizontal and/or vertical direction(s)) of changes . Motivated by Harris corner detector techniques, $s(p) = u(p) * v(p) - k * (u(p) + v(p))^2$ is used in this paper, k= 0.04. Since the changes F is not the real changes of interest, we add it to the unchanged class, i.e., $U = \{p|m(p) \leq \tau_2\} \cup \{p|m(p) \leq \tau_3, s(p) < \tau_4\}$.



(a) (b)

Figure 4: Different characteristics with respect to different types of changes. (c)distributions of change features in Euclidean coordinates. (d)distributions of change features in polar coordinates.

The task of search space is thus to determine τ_1 - τ_4 and discriminate C from U. As shown in Fig.4(b), the representation of change vector field in polar coordinate is more convenient for computation. The objective function is to make the distance between the changed class and the unchanged class maximized. Considering the complex statistical distributions of the changed class and the unchanged class, SVM(Support Vector Machine), a distribution-free classifier, is used to classify the object-specific change features. However, the traditional SVM needs training samples labeled beforehand, and the manual labeling fashion is not practical for real applications. In this paper, this problem is addressed by the progressive transductive SVM, which is implemented by the following search strategy.

3.4 Search strategy

The search strategy used in this paper is composed of two steps: initial classification and refined classification. In initial classification, some potential training samples are selected automatically(e.g., the regions C_2, U_2 and F in Fig.4(b)), and the initial change map is determined by the initial separating hyperplane. In refined procedure, the performance is improved gradually by adjusting the training samples(e.g., the regions C_1, U_1 and T in Fig.4(b)) dynamically.

3.4.1 Initial Classification The key to automatic training sample selection is to determine $\tau_i (i = 1, \ldots, 4)$, τ_1-τ_3 are related to $m(p)$, and τ_4 is related to $s(p)$. Since the training samples will be tuned iteratively in the latter procedure, a heuristic approach is used to determine τ_i approximately, i.e., let $\tau_i (i = 1, \ldots, 3)$ be the $\alpha_i \times N$ largest number of $m(p)$, α_i reflects the fraction of the changed/unchanged class, N is the total number of features. In this paper, $\alpha_i = 0.8$, $\alpha_2 = 0.2$, α_3 = 0.5, $\tau_4 = 50$. By this way, the training set $S_{train} = S^c \bigcup S^u$ can be achieved, $S^c = \{(x_p, 1)\}$ and $S^u = \{(x_p, -1)\}$, x_p is the 2-d change feature, $x_p = [m(p), s(p)]$. The initial classifier is then obtained by the inductive SVM, and the initial change map can be achieved by computing the decision function values on all unlabeled examples.

Due to the uncertainty of training samples and the inherent nature of inductive SVM, the initial change map may not be accurate enough. The uncertainty of training samples lies in the fact that the initial training samples are selected based on the approximated thresholds and limited to represent the whole change feature set. The inherent nature of inductive SVM lies in the purpose to optimize the classification performance over all possible future test data, In fact, this is not necessary since we are only interesting in the features extracted from the images being considered. Consequently, the classification accuracy is refined by the progressive classification, which is implemented by the iterative transductive SVM.

3.4.2 Refined Classification Given a set of independent labeled examples $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, $y_i \in \{-1, +1\}$, and unlabeled examples $\mathbf{x}_1^*, \ldots, \mathbf{x}_k^*$ from the same distribution, the aim of transductive SVM(Collobert et al., 2006) is to minimize the following equation over $(y_1^*, \ldots, y_k^*, w, b, \xi_1, \ldots, \xi_n, \xi_1^*, \ldots, \xi_k^*)$:

$$\min(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i + C^*\sum_{j=1}^{k}\xi_j^*) \quad (5)$$

$$s.t. \forall_{i=1}^{n}: \; y_i(w \cdot \phi(\mathbf{x}_i) + b) \leq 1 - \xi_i, \xi_i \geq 0. \quad (6)$$

$$\forall_{j=1}^{k}: \; y_j^*(w \cdot \phi(\mathbf{x}_j^*) + b) \leq 1 - \xi_j^*, \xi_j^* \geq 0. \quad (7)$$

Where the regularization parameters C and C^* control the generalization capabilities, ξ_i and ξ_j^* are positive slack variables enabling to deal with the permitted errors, ϕ is the mapping function. For the change detection algorithm based on the iterative transductive SVM, at each iteration, the non-representative training examples are removed from the training set, and the new representative examples are added from the unlabeled examples to the training set. Based on the hyperplane

$$f(\mathbf{x}_i) = \sum_j \alpha_j y_j K(\mathbf{x}_j, \mathbf{x}_i) + b \quad (8)$$

$$= \sum_j \alpha_j y_j < \phi(\mathbf{x}_j), \phi(\mathbf{x}_i) > + b \quad (9)$$

and Karush-Kuhn-Tucker condition, the training examples $\mathbf{x}_i$ can be partitioned into three different categories according to $g_i = y_i f(\mathbf{x}_i) - 1$(Cauwenberghs and Poggio, 2000): the set S of margin support vectors strictly on the margin($g_i = 0$), the set E of error support vectors exceeding the margin($g_i < 0$), and the remaining set R of reserve vectors exceeding the margin($g_i > 0$). At the next iteration, the set E should be deleted from the training set since its label is inconsistent with the current separating hyperplane, and it contradicts the assumption that the objective function has been minimized. For the unlabeled examples $\mathbf{x}_1^*, \ldots, \mathbf{x}_k^*$, only those lying within the margin band are important for the later classification since the adding of them to the training set may change the separating hyperplane. To keep the whole classification stable, at each iteration, we add the new representative training examples in a pair-wise manner(the number