A DISTANCE-BASED KERNEL CHANGE DETECTION ALGORITHM

MA Guorui, SUI Haigang, LI Pingxiang, QIN Qianqing

National Lab for Information Engineering in Surveying, Mapping & Remote Sensing, 430079
Wuhan University, P. R. China, 086-027-61370963, maguorui_rs@yahoo.com.cn

Commission VII, WG VII/5

ABSTRACT:

This paper proposes a distance-based kernel change detection algorithm (DKCD). The input vectors from two images of different times are mapped into a potentially much higher dimensional feature space via a nonlinear mapping. Which will usually increase the linear separation of change and no-change regions. Then, a simple linear distance measure between two feature vectors of high dimension is defined in features space, which corresponds to the complicated nonlinear distance measure in input space. Furthermore, the distance measure's dot is expressed in the combination of kernel functions and large numbers of dot operations processed in input space not in feature place by combined kernel tactic, which avoids the computational load. Finally this paper takes the soft margin single-class support vector machine to select the optimal hyper-plane with maximum margin. Preliminary results show the distance-based kernel change detection algorithm (DKCD) has excellent performance in speed and accuracy.

1. INTRODUCTION

Change Detection in multiple imagery of the same scene taken at different times is of widespread interest due to a large number of application (Li D. R., 2003; Liao M. S., 2000; Ma J. W., 2004; Fang Z., 1997), especially in remote sensing, medical diagnosis and so on. In the past years, a number of change detection methods have been proposed (Pilon P. G., 1998; Rowe N. C., 2001). Most of these algorithms try to compare the pixels, features, or objects description extracted from two or more images according to a distance measure defined in input space and a detection threshold. A remarkable change occurs where the distance between two input sets is larger than the threshold and no change is detected if distance is smaller than the threshold. Actually a hyper-curve is constructed in input space when the distance is equal to the threshold (if distance measure is liner, a hyper-plane is constructed). The change regions and no-change regions are separated by the hyper-curve. The separation of change and no-change regions in input space is only depended on the input sets and the distance measure. The too high dimensions of input sets and too much complexity of distance measure will seriously affect the change detection performance.

This paper proposes a distance-based kernel change detection algorithm (DKCD) inspired by recent machine learning theory (Burges C. J. C., 1998). The input vectors from two images of different times are mapped into feature space of high dimension via a nonlinear mapping, which will usually increase the linear separation of change and no-change regions. Then, a simple linear distance measure between two feature vectors of high dimension is defined in features space, which corresponds to the complicated nonlinear distance measure in input space. Furthermore, the distance measure's dot is expressed in the combination of kernel functions and large numbers of dot operations processed in input space not in feature place by combined kernel tactic, which avoids the operation burden in high dimension space.

The choice of threshold is another vital step in change detection

algorithm, which is not solved satisfactorily as yet. Supervisory learning methods are advisable, considering the facts that change samples and no-change samples are asymmetry badly and change samples are acquired difficulty and no-change samples are infinite, this paper takes the soft margin single-class support vector machine (Scholkopf B., 2001), translating the choice of threshold into the selection of the optimal hyper-plane with maximum margin. Choosing the suitable kernel function and parameter, we get a high detection accuracy.

2. HYPER-PLANE IN INPUT SPACE

Early change detection methods were based on the difference image or radio image, named simple differencing and image radioing, and such methods are still widespread due to simple operation and good performance (Fan H. S., 2001; Feng D. J., 2004). This section addresses the hyper-curve construction of change and no-change regions in input space using the two traditional change detection methods. Simple differencing change detection method expresses commonly as follows:

$$A(i,j) = 255 - \left| f(i,j) - g(i,j) \right| \qquad (1)$$
$$0 \le g(i,j) \le 255 , \quad 0 \le f(i,j) \le 255$$

where i, j are pixel coordinate, $f(i,j)$, $g(i,j)$ are two images taken at different times, $A(i,j)$ is gray image of simple differencing.

Image radioing change detection method is

$$B(i,j)=\begin{cases}255*f(i,j)/g(i,j) & f(i,j)\leq g(i,j),g(i,j)\neq 0\\255*g(i,j)/f(i,j) & f(i,j)>g(i,j),f(i,j)\neq 0\end{cases}\quad(2)$$

Where $B(i,j)$ is gray image constructed by image radio, the other parameters mean same as above.

Simple differencing and image radioing are two typical and traditional change detection methods, the gray images change trend is approximately the same. If we considerate all gray value, supposing $f(i,j)$, $g(i,j)$ are x,y coordinate axes, changing from 0 to 255, and $A(i,j)$, $B(i,j)$ are $z1,z2$ coordinate axes, we can obtain the gray image three dimension distribution of simple differencing and image radioing expressed in Figure 1 and Figure 2.
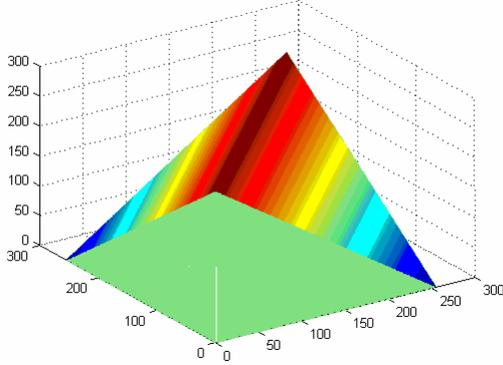

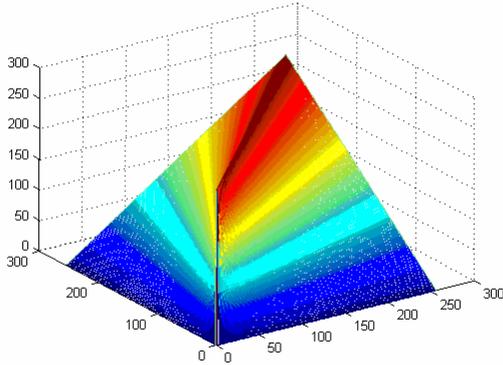Figure 1. Gray distribution of difference images

Figure 2. Gray distribution of radio images

From the above figures, we make out that where no change, the outcomes of simple differencing and image radioing are the same, namely when $x=y$, we have $z1=z2$, and when x,y both are big, $z1$ closes to $z2$, but when one of x,y is very small, $z1$ is differ from $z2$ obviously, suggesting image radioing method magnifies this change. Beside, on the plane that parallels $x=y$, the $A(i,j)$ are

equal everywhere, but the $B(i,j)$ change from 0 to $A(i,j)$ nonlinearly, that is the similarities and differences of simple differencing and image radioing.

The outcome of simple differencing can't reflect absolutely the change of radiation energy in ground scene, because the quantitative operation translation is generally nonlinear. For example, where the gray changes from 50 to 20 and from 250 to 220, the radiation energy changes in ground scene are not equal (Chen S. P.,1998). Image radioing method can boost up some remarkable change and restrain background regions, but which may over-magnify the change sometime. For example, where the gray changes from 200 to 20 and from 20 to 2, image radioing is invalid, but the outcomes of simple differencing are distinct. According to the two methods' shortcoming and people' vision characteristic, merging the two outcomes is one of advisable tactics.

Supposing the thresholds of simple differencing and image radioing are ξ and η respectively, the hyper-plane and the hyper-curve of change and no-change regions in combined input space are:

$$255-\left|f(i,j)-g(i,j)\right|-\xi=0 \quad(3)$$

$$\begin{cases}255*f(i,j)/g(i,j)-\eta=0 & f(i,j)\leq g(i,j),g(i,j)\neq 0\\255*g(i,j)/f(i,j)-\eta=0 & f(i,j)>g(i,j),f(i,j)\neq 0\end{cases}\quad(4)$$

3. KERNEL CHANGE DETECTION ALGORITHM

This paper proposes a distance-based kernel change detection algorithm (DKCD), not the dimensionality but the complexity of the distance measure matters. The input vectors p and q extracted from two images of different times are mapped into a potentially much higher dimensional feature space via a nonlinear mapping, which will usually increase the linear separation of change and no-change regions.

Training samples: $p=\{p_1,\cdots,p_l\}$, $q=\{q_1,\cdots,q_l\}$, l denotes the sample counts. Nonlinear mapping:
$$\Phi(p)=\{\phi_1(p),\phi_2(p),\cdots\}$$
$$\Phi(q)=\{\phi_1(q),\phi_2(q),\cdots\}$$
supposing the feature space is a N-dimensional space.

Following this, a simple linear distance measure between two high dimensional feature vectors is defined in features space, which corresponds to the complicated nonlinear distance measure in input space.

$$\Phi(x)=\{\phi_1(x),\phi_2(x),\cdots\}=\Phi(p)-\Phi(q)$$
$$=\{\phi_1(p)-\phi_1(q),\phi_2(p)-\phi_2(q),\cdots\}\quad(5)$$

The i-th new samples: $\Phi(x_i)=\Phi(p_i)-\Phi(q_i)$,

or $\phi_i(x) = \phi_i(p) - \phi_i(q)$

In the high dimensional space, we assume the new data of change and no-change regions can be separated by a linear hyper-plane (Nello C.,2000):

$$f(x) = \sum_{i=1}^{N} w_i \phi_i(x) + b = \sum_{i=1}^{N} w_i(\phi_i(p) - \phi_i(q)) + b \quad (6)$$

Expressed in dual form (John S. T.,2000.):

$$f(x) = \sum_{i=1}^{l} \alpha_i (\Phi(x_i) \cdot \Phi(x)) + b \qquad (7)$$

Furthermore, the distance measure's dot product is expressed in the combination of kernel functions and large numbers of dot product operations processed in input space not in feature place by combined kernel tactic, which avoids the operation burden in high dimension space.

$$\begin{aligned} (\Phi(x) \cdot \Phi(x')) &= ((\Phi(p) - \Phi(q)) \cdot (\Phi(p') - \Phi(q'))) \\ &= (\Phi(p) \cdot \Phi(p')) - (\Phi(p) \cdot \Phi(q')) - \\ (\Phi(q) \cdot \Phi(p')) &+ (\Phi(q) \cdot \Phi(q')) \qquad (8) \\ &= K(p,p') - K(p,q') - K(q,p') + K(q,q') \\ &= K(x,x') \end{aligned}$$

Thus, the linear algorithm which can be carried out in terms of dot products is made nonlinear by substituting a priori chosen kernel. The use of kernel enables us to evaluate the dot product in the feature space without explicitly computing their coordinates.

The choice of threshold is another vital step in traditional change detection algorithm, which is not solved satisfactorily as yet. Supervisory learning methods are advisable, considering the facts that change samples and no-change samples are asymmetry badly and change samples are acquired difficulty and no-change samples are infinite, this paper takes the soft margin single-class support vector machine (SVM), translating the choice of threshold into the selection of the optimal hyper-plane with maximum margin, which results in:

$$\max_{w,\xi,\rho} \quad -\frac{1}{2}\|w\|^2 - \frac{1}{vl}\sum_{i=1}^{l}\xi_i + \rho \qquad (9)$$
$$st. \quad (w \cdot \phi(x_i)) \geq \rho - \xi_i \quad \& \quad \xi_i \geq 0, \rho \geq 0, v \in (0,1]$$

This convex quadratic optimization problem is solved by introducing Lagrange multipliers, $a_i, i = 1, \cdots, l$, yielding the dual optimization problem

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j K(x_i, x_j) \qquad (10)$$
$$s.t. \quad 0 \leq \alpha_i \leq \frac{1}{vl}, \quad \sum_{i=1}^{l} \alpha_i = 1$$

This dual problem can be solved with standard QP routines, then the nonlinear decision function has a SV expansion (nonzero a_i are called SV)

$$f(x) = \mathrm{sgn}\{\sum_{i=1}^{l}\alpha_i K(x_i,x) - \rho\} = \mathrm{sgn}\{\sum_{i=1}^{l}\alpha_i K(p_i,q_i,p,q) - \rho\} \quad (11)$$

The offset ρ can be recovered by exploiting that for any a_i which is not at the upper or lower bound, the corresponding (p_i, q_i) satisfies

$$\rho = (w \cdot \Phi(x_i)) = \sum_{j=1}^{l}\alpha_i K(x_j, x_i) = \sum_{j=1}^{l}a_i K(p_j, q_j, p_i, q_i) \quad (12)$$

For a given testing data (p,q), Substitute it to the decision function, $f(x) \geq 0$ for a significant change and $f(x) < 0$ otherwise.

4. EXPERIMENT AND ANALYSIS

Figure 3 shows a pair of Ikonos images in Iraq presidential palace using to test the kernel change detection algorithm, image size is 490*410 pixels, taken at 10/9 2002 and 1/4 2003 respectively. Compute environment is window XP, C++, Pentium 2.8GHz, 512M RAM. The experiment steps as follows:
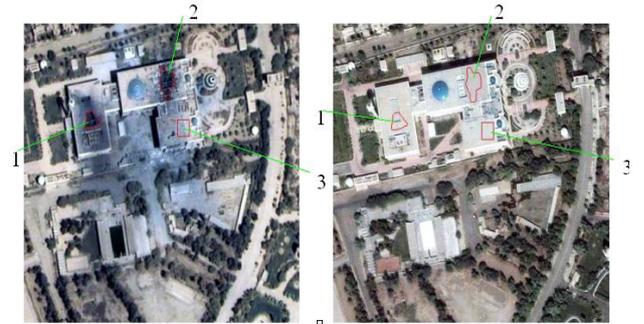

Figure 3. Images and samples of Iraq presidential palace

1)   Training samples selection: change regions are selected by self-auto or handiwork, as shown in regions marked number 1 in Figure 3, there are 321 training samples, the (R、G、B)

components of two times are selected as the elements of each sample, that is a six dimension input space.

2) Testing samples selection: other change regions are selected for positive testing samples by the same method, as shown in regions marked number 2 in Figure 3, including 692 positive testing samples, then no-change regions are selected for negative testing samples, as shown in regions marked number 3 in Figure 3, the number is 84, the sum of testing samples is 976. The elements of testing samples are same as the training samples. To avoid the elements in greater numeric ranges dominate those in smaller numeric ranges and be convenient for calculation, we linearly scaling each elements to the range [-1,1].

3) Model selection and training: model selection consists of the selection of SVM, kernel function and kernel parameters selection. This paper using the single-class v-SVM to training, RBF is a reasonable first choice for single kernel function, then the combined kernel function is

$$K(x,x')=K(p,q,p',q')=\exp\{-\gamma|p-p'|^2\}-\exp\{-\gamma|p-q'|^2\}$$
$$-\exp\{-\gamma|p'-q|^2\}+\exp\{-\gamma|q-q'|^2\} \quad (13)$$

The model has two parameters that have to be adjusted, the maximum fraction of training error, v, and the kernel parameter, γ. Because the training samples are credible change regions, the parameter v can be set priori to the highest allowable fraction of miss-classification of the target class. Here only a 1% classification error is allowed on the training data, $v = 0.01$. There are several possible criterions for selecting γ such as minimizing the number of support vectors, maximizing the margin of separation from the origin and maximizing the detection accuracy. Figure 4 and Figure 5 and Figure 6 show three plots of the different criteria as a function of γ for the Iraq images.
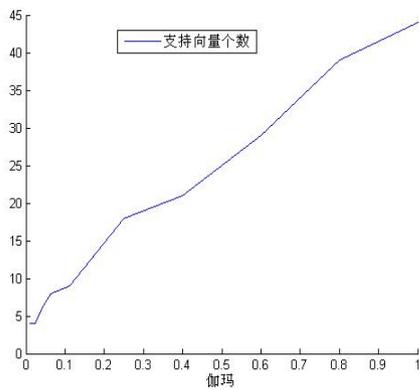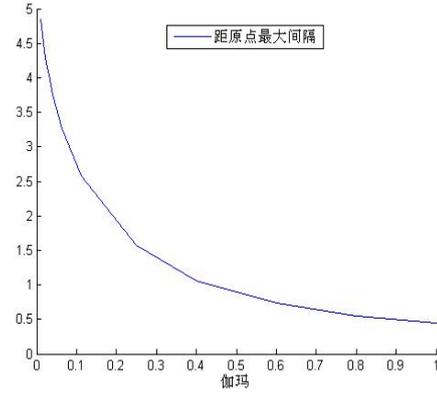

Figure 4. Number of support vectors as function of γ

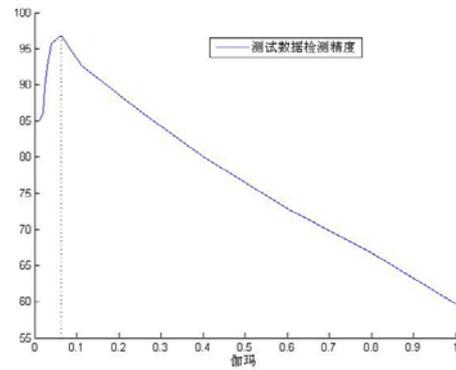Figure 5. Margin from the origin as function of γ

Figure 6. Detection accuracy as function of γ

From the Figure 4 and Figure 5, we see that with higher the parameter γ the number of support vectors increases and the margin of separation from the origin decreases, it is seem that smaller the parameter γ, better the performance, however, in Figure 6, Detection accuracy curve of testing data has a peak 96.8% at $\gamma = 0.0625$, the reason is that the parameter γ determines the shrinkage of decision-making plane in feature space, smaller the parameter γ, bigger the decision-making plane, higher the detection accuracy of positive testing samples, but lower the detection accuracy of negative testing samples because of including many non- remarkable changes regions. For a certain parameter γ, the best detection accuracy of all testing samples is obtained.

4) Change detection: Each pixel of needing detection is predicted with the selected v and optional γ, the change pixel is marked, the boundary of change regions is obtained following some post-processing step. As shown in Figure 7, the kernel change detection of this paper is insensitive to the illumination, the shadow, even the non- remarkable change created by dust but is good at the damage change by attack.

Figure 7. Kernel change detection result

5. CONCLUSION

This article proposed a novel distance-based kernel change detection algorithm. The main thought is that the two vectors in input space are mapped into feature space of high dimension via a nonlinear mapping, then, a new vector including change information as many as possible is obtained by comparing two high dimension vectors using traditional change detection methods, finally, a optimal hyper-plane with maximum margin on the new vector is constructed by soft margin single-class support vector machine. Where, the nonlinear mapping and the optimal hyper-plane are achieved implicitly by the distance kernel function. The combine kernel function increases the linear separation of change and no-change regions while avoids the computational load. Preliminary results show the distance-based kernel change detection algorithm (DKCD) has excellent performance in speed and accuracy.

REFERENCES

Burges C. J. C., 1998. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2(2), pp.121-167.

Chen S. P., Tong Q. X., Guo H. D., 1998. The Study of Remote Sensing Mechanisms. Beijing, Science, pp. 310-312.

Fan H. S., Ma A. N., Li J., 2001. Case Study on Image Differencing Method for Land Use Change Detection Using Thematic Data in Renhe District of Panzhihua. Journal of Remote Sensing, 5(1) , pp.75-80.

Fang Z., Zhang J. Q., Zhang Z. X., 1997. Change Detection Based on Airial Image of Urban Area. Journal of Wuhan Technical University of Surveying and Mapping, 22 (3) , pp. 240-244.

Feng D. J., Li Y. S., Deng F., 2004. Detecting Change Information Automatically by Methodeof Wavelet Coefficient Difference. Remote Sense Information, 2, pp.13-15.

John S. T., 2000. Nello Cristianini. Kernel Methods for Pattern Analysis. Cambridge University, pp. 286-288.

Li D. R., 2003. Change Detection from Remote Sensing Images. Geomaticas and Information Science of Wuhan University, 28(3) , pp. 7-12.

Liao M. S., Zhu P., Gong J. Y., 2000. Multivariate Change Detection Based on Canonical Transformation. Journal of Remote Sensing, 4(3) , pp. 197-201.

Ma J. W., Tian G. L., Wang C. Y., 2004. Yan Shouxun. Review of the Development of Remote Sensing Change Detection Technology. Advance in Earth Sciences, 19(2) , pp. 192-196.

Nello C., John S. T.,2000. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University, pp. 159-160.

Pilon P. G., Howarth P. J., Bullock R. A. and Adeniyi P. O.,1998. An Enhanced Classification Approach to Change Detection in Semi-arid Environments. Photogrammetric Engineering and Remote Sensing, 54, pp.1709-1716.

Rowe N. C., Grewe L. L.,2001. Change Detection for Linear Features in Aerial Photographs Using Edge-Finding. IEEE Transaction on Geo-science and Remote Sensing, 39(7) , pp.1608-1612.

Scholkopf B., Platt J. C., Shawe-Taylor J., Smola A. J.,2001. Estimating the Support of a High-dimensional Distribution. Neural Computation, 13(7) , pp.1443-1471.