

A Kernel-Based Similarity Measuring for Change Detection in Remote Sensing Images

Xiaodan Shi^{a,*}, Guorui Ma^a, Fenge Chen^b, Yanli Ma^a

^a State Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129
Luoyu Road, Wuhan, Hubei, 430079, China-(sxdrs,mgr,mayanli)@whu.edu.cn

^b Department of Statistics, Wuhan University of Technology, 129 Luoshi Road, Wuhan, Hubei, 430079, China-
lixueyuanchen@whut.edu.cn

Commission VII, WG VII/5

KEYWORDS: change detection, kernel function, similarity measure, probability density

ABSTRACT

This paper presents a kernel-based approach for the change detection of remote sensing images. It detects change by comparing the probability density (PD), expressed as kernel functions, of the feature vector extracted from bi-temporal images. PD is compared by defined kernel functions without immediate PD estimation. This algorithm is model-free and it can process multidimensional data, and is fit for the images with rich texture in particular. Experimental results show that overall accuracy of the algorithm is 98.9%, a little bit better than that of the change vector analysis and classification comparison method, which is 96.7% and 95.9% respectively.

1 INTRODUCTION

Change information of the earth's surface is becoming more and more important in monitoring local, regional and global resources and environment (Woodcock 2001). Change detection technology of multi-temporal remote sensing images has special significance in GIS data updating, disaster monitoring and evaluating, project construction monitoring, military and war inspection and evaluation, and land use and land cover (LULC) and so on (Coppin and Bauer 1994, Muchoney and Haack 1994, Pilon *et al.* 1988, Ridd and Liu 1998, Roy *et al.* 1991, Sader *et al.* 1991, Vogelmann 1988, Du *P. et al.* 2014).

Typical change detection methods consist of two steps: first, assume that an image complies with a certain distribution, and parametric method or nonparametric method is adopted in accordance with the distribution model to estimate the probability density (PD) or conditional PD for the image. In practice, difference image of two-temporal images share the same distribution model as the original images, therefore only the PD or conditional PD of the change or no-change region in the difference images need to be estimated; secondly, detect change in accordance with dissimilarity measure of conditional probability to obtain the optimal separation threshold of change or no-change by Bayesian decision with the help of prior knowledge (Pham 2002). Using polar coordinate theory and experimental method, Bovlo and Bruzzene (2007) proved that after radiometric correction and geometric correction, the simple Gaussian model is more appropriate than any other complex model for the no-change class; models adopted for the change class vary with image content. The risk of false assumption frequently exists in the model-based method, and this kind of method works only for the single-band difference images; when applied to multi-band image processing, dimensional disaster may occur. In addition, error may exist in the result of conditional PD estimation; computation is quite complicated, which makes change detection low processing level, tedious threshold selection and poor detection efficiency (Lu *et al.* 2004, Richard *et al.* 2005, Sui *et al.* 2008).

This paper proposes a model-free kernel-based similarity measure for change detection. Faced with challenges of high-dimensional non-linear data, kernel method does not reduce the dimension. Instead, it tries to upgrade the non-linear hypothesis space to feature space with higher dimensions or even infinite dimensions. Samples become very sparse in feature space, but it smoothes the way of applying linear learning algorithm to it (Scholkopf B. *et al.* 1999, Potin D. *et al.* 2006, Desobry *et al.* 2005, Desobry F. and Davy M. 2004). In this paper a feature transform is implemented with kernel function method by the inner product of samples in feature space after the transform of defined features, which can be

considered as a non-linear version of the corresponding linear algorithm. It has attracted much attention because it avoids dimensional disaster.

The model-free PD estimation method, namely, hyper-plane method of one-class Support Vector Machine (SVM) is first adopted by this paper to estimate the PD of bi-temporal images. The PD is expressed as kernel functions, and compared by defined kernel functions without direct PD estimation, which simplifies the PD estimation required by traditional procedures of change detection

2 PD ESTIMATION OF ONE-CLASS SVM

A direct way of detecting change is to compare the similarity measure or dissimilarity measure of the input vectors of the bi-temporal images. The indexes that are most commonly used to describe the similarity of two vectors are length, distance and angle. It is found that all these indexes can be expressed by inner product, which is an index of similarity measure. In accordance with the definition of kernel function that is the inner product of non-linear mapping in the corresponding feature space, choosing different kernel functions is to select different inner products, which means different standards are adopted to estimate the similarity. Thus kernel function can be used to measure similarity. By designing proper kernel function, the degree of similarity can be estimated among the vectors of bi-temporal images, and thus change can be detected.

Assume that $x = \{x_1, \dots, x_t\}$ is the training sample for input space X ; random training sample x_i and other training samples are independent and identically distributed according to some probability density function $pdf(x)$. The aim of the one-class classifier is to compute the area R_x in the input space X in accordance with training samples. The vector that satisfies the distribution of the probability density $pdf(x)$ will be within the R_x , while the vector that doesn't meet $pdf(x)$ will be outside the R_x . If decision function is used, it can be written as follows:

$$f_x(\cdot) \geq 0, \text{ within } R_x \quad f_x(\cdot) < 0, \text{ outside } R_x \quad (1)$$

The basic idea of the one-class SVM hyperplane algorithm is to first choose a right kernel function and map the one-class samples into a feature space; and then to regard the origin of coordinates in this space as the only point of another type, and to use the hyperplane to separate these two kinds of mapping points with maximum interval between them.

*Corresponding author

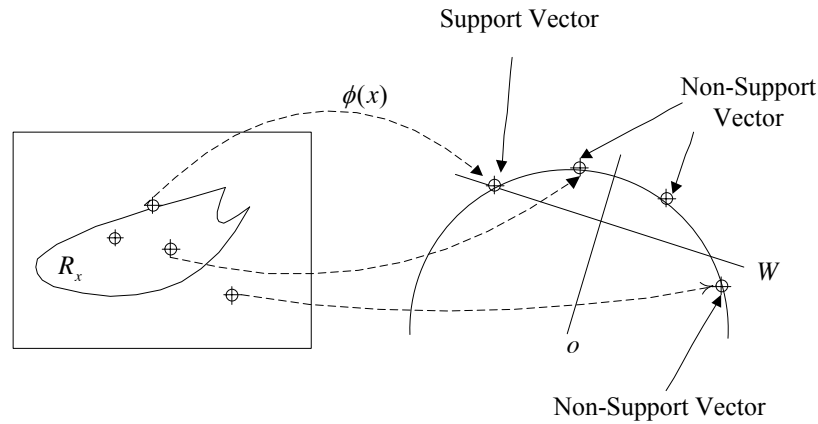


Figure 1 The one-class SVM Mapping

Any hyper plane in high dimension feature space can be written as a set

$$\{\phi(x) \in H \mid \langle w, \phi(x) \rangle_H - \rho = 0\} \quad (2)$$

Assume that $f_x(\cdot) = \langle w, \phi(x) \rangle_H - \rho$ is the decision function, the upper area of the hyperplane corresponds to the inside of R_x ; the lower area corresponds to the outside of R_x ; and hyperplane corresponds to decision function, namely, PD estimation (figure 1).

Figure1 illustrates the map of spatial data points in the original space non-linearly into the high-dimensional feature space. The PD estimation in the original space is converted into the process of seeking optimal separating hyper plane in the feature space. A data point on the hyper plane is non-support vector, while the points mapped on the hyper plane are support vector. The hyper plane can be defined as a combination of these points. The points that are mapped below the hyper plane are interval mistake samples. Of all the possible hyper planes, the one that matches optimal PD estimation satisfy the following optimization criterion:

$$\begin{aligned} \max_{w, \xi, \rho} \quad & -\frac{1}{2} \|w\|^2 - \frac{1}{vl} \sum_{i=1}^l \xi_i + \rho \\ \text{s.t.} \quad & (w \cdot \phi(x_i)) \geq \rho - \xi_i \quad \& \quad \xi_i \geq 0, \rho \geq 0, v \in (0, 1] \end{aligned} \quad (3)$$

The final decision function is

$$f(x) = \text{sgn}\{(w \cdot \Phi(x)) - \rho\} \quad (4)$$

Lagrange multiplier is used to solve the problem of optimization, where coefficient $\alpha_i \geq 0, \beta_i \geq 0$, and the corresponding Lagrange function are given as following

$$\begin{aligned} L_p(w, \xi, \rho; \alpha, \beta) = & \frac{1}{2} \|w\|^2 - \frac{1}{vl} \left(\sum_{i=1}^l \xi_i \right) - \\ & \rho - \sum_{i=1}^l \alpha_i ((w \cdot \Phi(x)) - \rho + \xi_i) - \sum_{i=1}^l \beta_i \xi_i \end{aligned} \quad (5)$$

Set the derivatives with respect to the primal variables w, ξ, ρ , yielding

$$w = \sum_{i=1}^l \alpha_i \Phi(x_i)$$

$$\alpha_i = \frac{1}{vl} - \beta_i \leq \frac{1}{vl}, \sum_{i=1}^l \alpha_i = 1 \quad (6)$$

Substitute them in the decision function, then

$$f(x) = \text{sgn}\left\{ \sum_{i=1}^l \alpha_i K(x_i, x) - \rho \right\} \quad (7)$$

Substituting equation (6) into equation (5) and using kernel function, then obtain the dual problem:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j K(x_i, x_j) \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq \frac{1}{vl}, \sum_{i=1}^l \alpha_i = 1 \end{aligned} \quad (8)$$

If solve the problem of QP, then get the value of α . All the samples of which α_i are non-zero are called Support Vectors (SV). In accordance with any SV, see the following function:

$$\rho = (w \cdot \Phi(x_i)) = \sum_{j=1}^l \alpha_j K(x_j, x_i) \quad (9)$$

The decision function, namely $pdf(x)$, is $f_x(\cdot) = \langle w, \phi(x) \rangle_H - \rho$.

3 CHANGE DETECTION BASED ON KERNEL SIMILARITY MEASURE

Assume that the local pixels of bi-temporal images to be compared are vector sets P and Q respectively. Use two one-class SVMs to train the samples P and Q independently and get their PD functions respectively, yielding two regions R_{xp} and R_{xq} , equivalently, and two hyper planes W_p and W_q in feature space H . These two hyper planes are parameterized by (w_p, ρ_p) and (w_q, ρ_q) respectively. The vectors w_p and w_q define a two dimensional plane, denoted by P , which intersects the hyper sphere S along a circle with center O and radius 1, as depicted in Figure 2. If there is no change in the physic local

area, the vectors w_p and w_q are collinear, in other words, the two hyper planes are overlapped. The more change, the bigger the angle between the two hyper planes. Arc $d_{arc}(c_p, c_q)$ is the angle between the two hyper planes. However, the difference between the two hyper planes can't be completely defined by the angle, because translation exists between them. A better dissimilarity measure is shown as follows

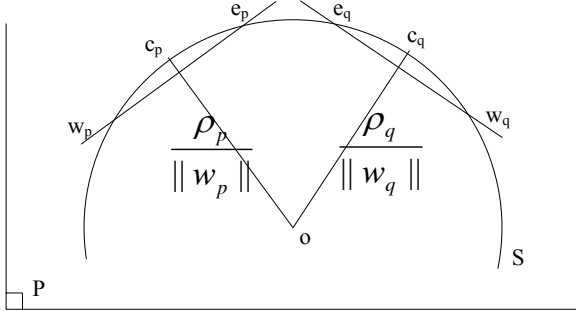


Figure 2 Sketch map of the high-dimensional feature space

Dissimilarity measure for the feature space

$$D_s(p, q) = \frac{d_{arc}(c_p, c_q)}{d_{arc}(c_p, e_p) + d_{arc}(c_q, e_q)} \quad (10)$$

$d_{arc}(c_p, c_q)$ denotes the intra-region distance of bi-temporal images, and $d_{arc}(c_p, e_p)$, $d_{arc}(c_q, e_q)$ denote the inter-region distance of each image respectively, $D_s(p, q)$ is the ratio between the intra-region distance and inter-region distance of bi-temporal images, which is similar to the Fisher criterion function. Big intra-region distance and small inter-region distance is corresponding to the significant change region. The above equations are defined in feature space. A key point of kernel method is that $D_s(p, q)$ must be computed in the input space.

Because the radius of the hyper sphere in feature space is 1 (implemented by kernel function), the arc distance is

$$d_{arc}(c_p, c_q) = \arccos(\langle c_p, c_q \rangle_H) \quad (11)$$

And $c_p = w_p / \|w_p\|_H$, $c_q = w_q / \|w_q\|_H$,

$$e_p = \rho_p / \|w_p\|_H, e_q = \rho_q / \|w_q\|_H$$

The intra-region distance of bi-temporal images is

$$d_{arc}(c_p, c_q) = \arccos\left(\frac{\langle w_p, w_q \rangle_H}{\|w_p\|_H \|w_q\|_H}\right) \quad (12)$$

The weight is expressed as the linear combination of support vectors:

$$w = \sum_{i=1}^l a_i \Phi(x_i) \quad (13)$$

Then

$$\frac{\langle w_p, w_q \rangle_H}{\|w_p\|_H \|w_q\|_H} = \frac{a_p^T K_{pq} a_q}{\sqrt{a_p^T K_{pp} a_p} \sqrt{a_q^T K_{qq} a_q}} \quad (14)$$

Similar calculation can be applied to $d_{arc}(c_p, e_p)$, $d_{arc}(c_q, e_q)$

$$d_{arc}(c_p, e_p) = \arccos\left(\frac{\rho_p}{\sqrt{a_p^T K_{pp} a_p}}\right)$$

$$d_{arc}(c_q, e_q) = \arccos\left(\frac{\rho_q}{\sqrt{a_q^T K_{qq} a_q}}\right) \quad (15)$$

Substitute $d_{arc}(c_p, c_q)$, $d_{arc}(c_p, e_p)$, $d_{arc}(c_q, e_q)$ into the dissimilarity measuring formula $D_s(p, q)$. Then, the similarity measure of the local area can be achieved, and change in this area can be detected by threshold.

In the least probable cases: if $d_{arc}(c_p, c_q)$ tends to be zero, namely, the hyper-planes of two support vectors are approximately the same, change hasn't happened on the bi-temporal images; if $d_{arc}(c_p, e_p)$ and $d_{arc}(c_q, e_q)$ tend to be zero, then

$$\frac{\rho_q}{\|w_q\|} \rightarrow 1, \frac{\rho_p}{\|w_p\|} \rightarrow 1 \quad (16)$$

Find improper regions in accordance with PD, for example, the smooth region, where all the vectors are almost the same, so it is impossible to get the support vector and accurate PD estimation. This can be accomplished by introducing an absolute term $\xi > 0$ behind $d_{arc}(c_p, e_p)$ or $d_{arc}(c_q, e_q)$.

An overall functional diagram of the proposed system is as Figure 3. The steps of change detection method of kernel-based similarity measure are as follows:

- 1) Image pre-processing: such as radiometric corrections and image registration are important and indispensable steps;
- 2) Initialization: select size of the training sets, namely size of blocks, kernel function and its parameter and the threshold;
- 3) Train an one-class SVM on the set p , and obtain the parameter (w_p, ρ_p) ;
- 4) Train another one-class SVM on the set q , and obtain the parameter (w_q, ρ_q) ;
- 5) Compute the kernel functions K_{pp}, K_{pq}, K_{qq} using the support vectors of the bi-temporal images;
- 6) Compute the similarity measure $D_s(p, q)$ defined by kernel function;
- 7) Detect change based on the similarity measure $D_s(p, q)$ and threshold η ;
- 8) Repeat step 3)-7) until all regions are detected.

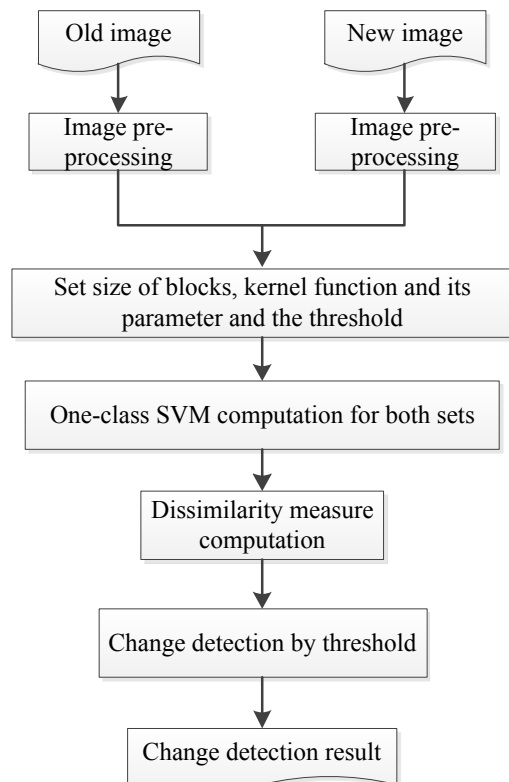


Figure 3 Overall functional diagram of the proposed system

Change is complex and universal, no existing approach is optimal and applicable to all cases. The conventional method of simple difference and threshold is linear method, our method is free of model and no-linear method, which can deal with more complex situation. Conventional class method depends on the samples and the samples selection is very difficult in multispectral image, the change detection method proposed by this paper is a unsupervised method, the samples in our method are not needed, which is similar to the Fisher Analysis. Both methods make use of the intra-region ratio and inter-region ration of bi-temporal images based on the Rayleigh Principle, which is similar to the Fisher criterion. However, the classification is different from Fisher Analysis. The direction of classification plane in this paper is defined by commutating two independent one-class support vectors, while kernel Fisher obtains the biggest variance of the projection vector by optimizing the algorithm. Kernel Fisher is fit for two-type classification. While in practice change conditions is much more complex. They may have different types, more training samples, high dimension and unknown PD and complex shape etc. Kernel Fisher is not applicable, and the method proposed in this paper is more feasible.

4 ALGORITHM FLOW AND EXPERIMENTS

The 2.83 version of LibSvm, developed by Professor Lin Zhiren, has been used in the training of two one-class support vectors. The kernel functions are normalized. Overall accuracy is adopted as evaluation criterion for the performance comparison, which is the ratio of the number of correct samples and all the testing samples. In addition, the accuracy is comprehensively evaluated by integrating Kappa coefficient.

4.1 A Survey on the Research Area and Basic Processing

Two bi-temporal TM images of 1998 and 2000 are adopted. They are constituted by the other six bands except for the sixth infrared band. The bi-temporal images are pre-processed, i.e., precise geometrical correction (precise small-bin differential rectification, rectification accuracy is within 0.2 pixels), normalized radiometric correction and image segment. The images are 496×496 pixels and image resolution is 30 meter/pixel. The colored images of 432 bands are as shown in Figure 3. Typical objects such as water bodies, buildings, roads, mountains and vegetation and so on are included in the testing images. Accuracy test is conducted by selecting 1334 change pixels on the bi-temporal images by hand as positive samples and selecting 1826 no-change pixels as negative samples.



(a)Image in 1998



(b)Image in 2000

Figure 4 bi- temporal images

The experiment presents the change detection results from change vector analysis and post-classification comparison respectively first and evaluates the accuracy. Then select RBF kernel functions, and select different kernel parameters and different block sizes respectively to test the performance of the algorithm in this paper. The results of change vector analysis and classification comparison are compared to each other to check the performance of this algorithm. Some of the results given by the experiment are difference images of change, and separation threshold needs to be selected in accordance with experience.

- (1) Change vector analysis: Take the difference absolute value of the corresponding bands of the bi-temporal images as change vector; a 543-band pseudo color image is as

shown in Figure 5 (a). Take the intensity value of the change vector as a standard to detect change; and the change detection result acquired after bi-value processing is as shown in Figure 5 (b).

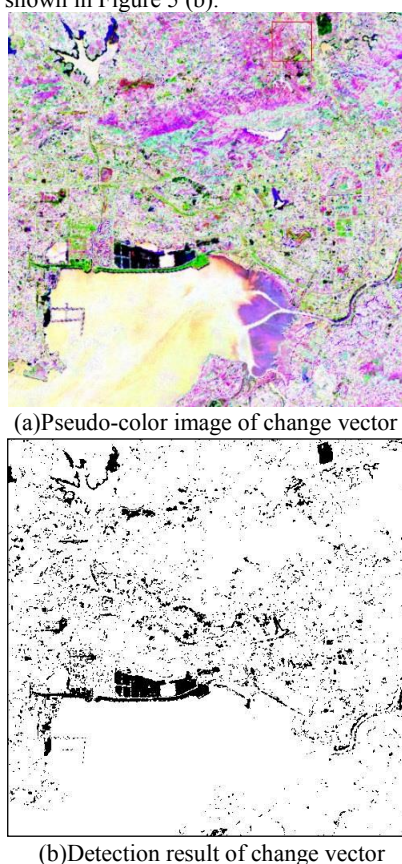


Figure 5 The result of CVA

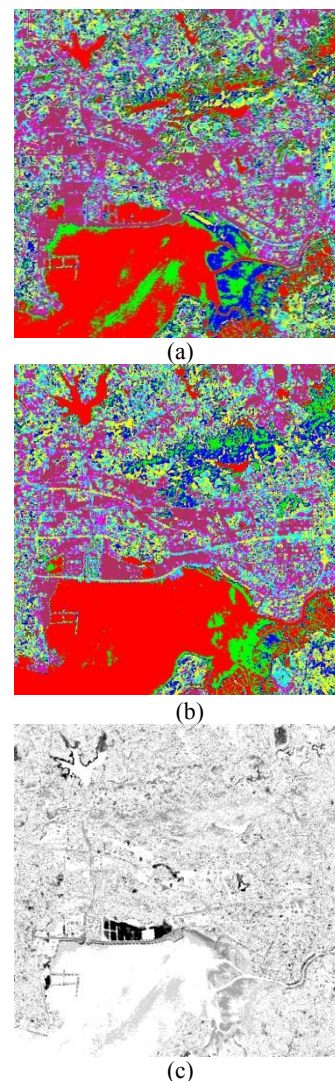


Figure 6 Results of classification and change detection

(2) Classification comparison: least-distance classification and non-supervised classification are applied to the bi-temporal images respectively (IsoData in Envi4.8), and the results are as shown in Figure 6(a) and (b). Compare the bi-temporal images after classification, and the difference is as shown in Figure 6(c).

Table 1 Change detection results

	Change Vector Analysis	Classification Comparison	Our method
Accuracy(%)	96.7	95.9	98.9
Kappa Coefficient	0.945	0.911	0.97

From Figure 5 and Figure 6, it is obvious that both change vector analysis and classification comparison have detected change of the buildings and lake surface, but they got a little difference in such trivial changes as change of residential area and so on. The overall accuracy and Kappa Coefficient of the two methods are as shown in Table 1. Change vector analysis has better performance than non-supervised classification comparison. The reason is that it is difficult to define number of classification in classification comparison and classification and comparison are separate, which results in error accumulation; while change vector analysis has made full use of data of all bands equally well, and no other procedure produces error, so it has high accuracy.

4.2 Change Detection Results with Different Parameters

Use RBF kernel functions for experiment. Because the radius of the hyper sphere in the above algorithm shall be 1, kernel functions are normalized. ν and γ are defined with different values, and then check the change detection results. Generally speaking, ν is an upper bound on the fraction of interval mistake sample points and a lower bound on the fraction of SVs. Its value is between 0 and 1, 0 and 1 are included. γ is the width of the basis function of radial direction. It determines the similarity measure and shape of the decision surface. Change detections are compared by choosing ν and γ with different values. A $\xi = 0.05$ is applied to all experiments. A region of

3*3 window size is processed every time. Namely, 9 samples are selected for one-class support vector training every time. With typical kernel parameters, the results of change difference are as shown in Figure 7. Overall accuracy and Kappa

coefficient of the change detections with different kernel parameters are as shown in Table 2.

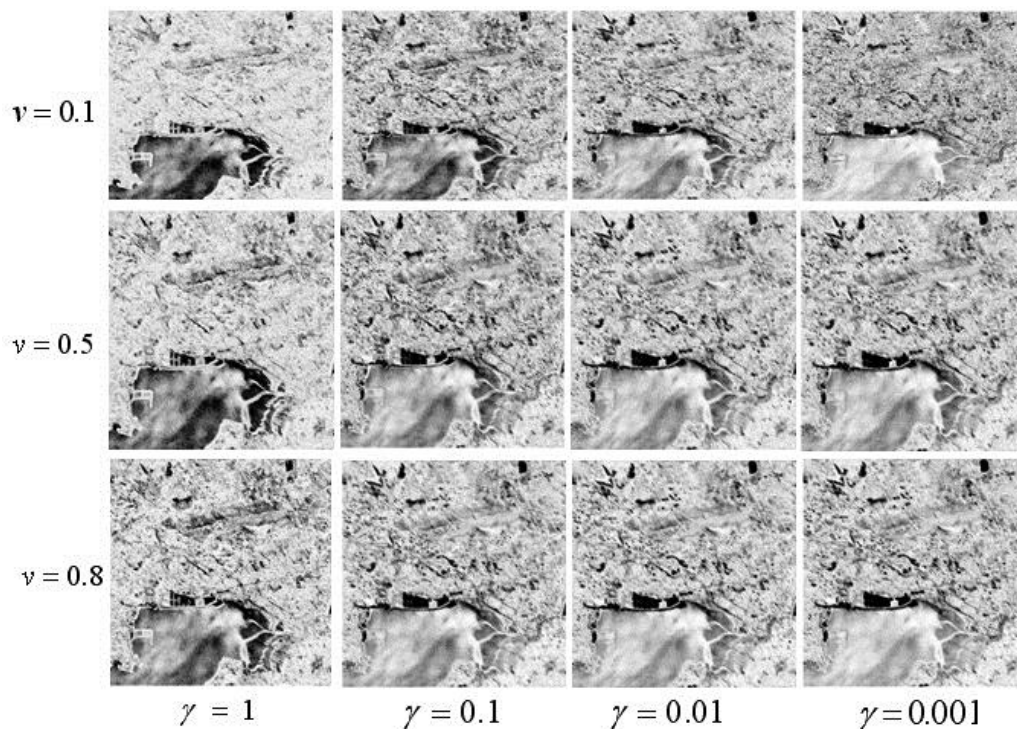


Figure 7 some results with different parameters

Table 2 Change detection results with different parameters

Overall Accuracy: %(κ)	$\gamma = 1$	$\gamma = 0.1$	$\gamma = 0.01$	$\gamma = 0.001$	$\gamma = 0.0001$
$\nu = 0.1$	77.3(0.543)	91.5(0.822)	97.0(0.932)	96.8(0.929)	96.7(0.927)
0.2	79.3(0.579)	95.8(0.907)	97.8(0.949)	97.9(0.951)	95.5(0.902)
0.3	84.2(0.678)	97.2(0.936)	98.1(0.955)	98.5(0.964)	98.5(0.962)
0.4	81.9(0.638)	97.8(0.949)	98.1(0.955)	98.9(0.972)	97.6(0.945)
0.5	83.3(0.664)	98.5(0.962)	98.5(0.963)	98.9(0.971)	98.8(0.969)
0.6	87.9(0.752)	98.4(0.962)	98.8(0.969)	98.6(0.965)	97.6(0.946)
0.7	91.5(0.823)	98.8(0.969)	98.8(0.97)	98.4(0.96)	98.8(0.97)
0.8	94.3(0.879)	98.9(0.97)	98.9(0.972)	98.7(0.967)	98.7(0.967)
0.9	96.9(0.931)	98.9(0.97)	98.9(0.971)	98.7(0.967)	98.6(0.965)

It can be seen from the results of quantitative analysis (Table 2) and qualitative analysis (Figure 7) that this algorithm is not sensitive to ν , because different values of it all can lead to good change detection results. In general, when γ keeps the same value, the larger ν is, a better result it is, because larger ν can result in more support vectors, so the defined hyper plane is more accurate. However, larger ν can also result in longer operation time, because more SVs are taking part in operation. Taking both detection time and effect into account, ν from 0.3 to 0.7 is comparatively better. The table also has shown that γ also has wide numeric area. In experiments, γ has little influence on the number of the SV, so the corresponding operation times are almost the same. However, large γ mainly reflects the overall and typical changes, while small γ can detect more trivial changes, because γ is the width

of the basis function of radial direction, and the smaller it is, more precise PD estimation can be got and more detailed detection results can be achieved. Taking noises into consideration, detection accuracy falls slightly. In accordance with results of quantitative analysis, the best results can be achieved, when γ is between 0.01 and 0.1

4.3 Change Detection Results with Blocks of Different Sizes

Compare the change detection results, using RBF kernel function, and selecting blocks with the size of 3*3, 5*5 and 7*7 respectively. When $\nu = 0.2$ and $\gamma = 0.1$, change difference results of different block sizes are as shown in Figure 8. The

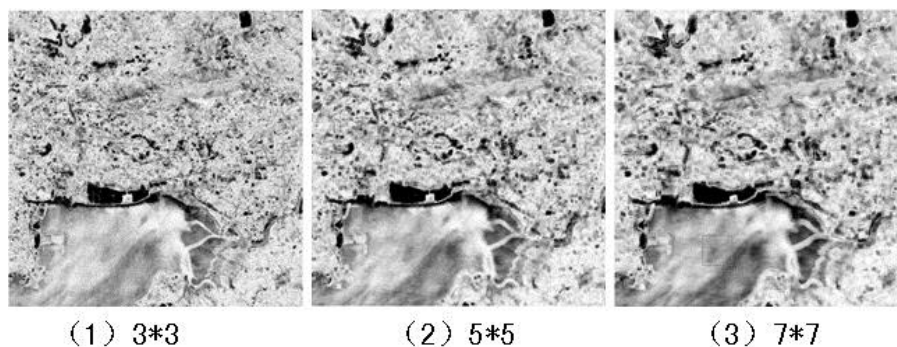


Figure 8 results with some blocks of different sizes

Table 3 Change detection results with blocks of different sizes

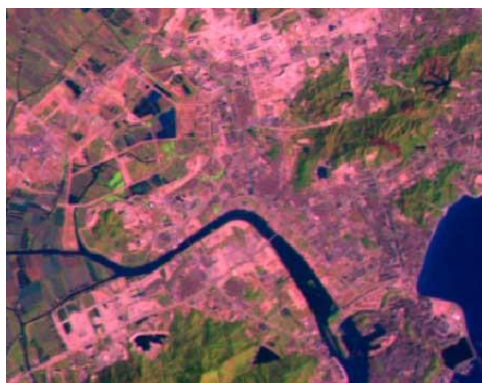
Overall Accuracy (Kappa Coefficient)	3*3	5*5	7*7
$\nu = 0.2, \gamma = 0.1$	95.8(0.907)	95.4(0.9)	91.1(0.809)
$\nu = 0.4, \gamma = 0.01$	98.1(0.955)	98(0.952)	93.8(0.867)
$\nu = 0.6, \gamma = 0.001$	98.6(0.965)	98.4(0.96)	93.9(0.867)
$\nu = 0.8, \gamma = 0.0001$	98.7(0.967)	98.7(0.967)	94.3(0.879)

overall accuracy and Kappa coefficient under different parameters and different block sizes are as shown in Table 3.

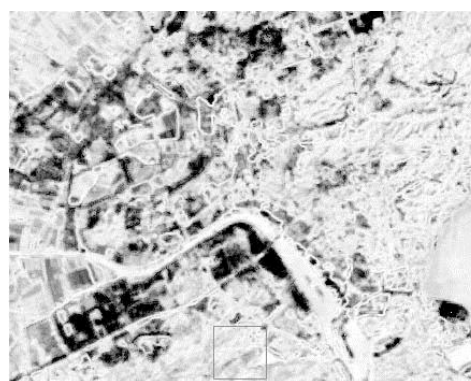
Another two bi-temporal TM images of 1991 and 1996 in Shenzhen are adopted. precise geometrical correction and normalized radiometric correction are needed. Compare the change detection results, using RBF kernel function, and selecting blocks with the size of 3*3, 5*5 respectively. When $\nu = 0.2$ and $\gamma = 0.1$. Change difference results of different block sizes are as shown in Figure 9.



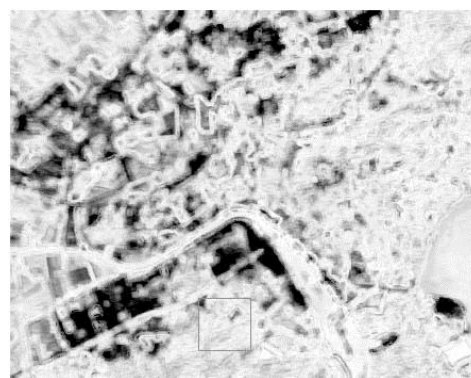
(a)Image in 1991



(b)Image in 1996



(c) results with blocks of 3*3



(d) results with blocks of 5*5

Figure 9 Another bi- temporal images and results

It is seen from the results of quantitative analysis (Table 3) and qualitative analysis (Figure8, Figure 9) that the smaller the block is, the more detailed change can be detected and more accurate results can be achieved. Detection accuracies got from block of 3*3 and block of 5*5 are approximately the same, but the accuracy from the block of 7*7 drops greatly. On the other hand, the larger the block is, more pixels take part in operation,

which will increase operation time obviously, and is not good for change detection. Overall small blocks such as 3*3 or 5*5 are better.

5 CONCLUSION

This paper proposes an algorithm that detects change by comparing the PD of the feature vector of bi-temporal images. The PD comparison is expressed as kernel functions, and change is detected using similarity measuring with kernel functions. PD is compared by defined kernel functions without immediate PD estimation. The results of experiments and theoretical analysis show that this algorithm is free of model; it can deal with any complex situation; it has no real PD estimation and PD comparison is achieved indirectly; it can process multidimensional data and multi-scale data; and is fit for the images with rich texture, the detection results have good visual effects.

ACKNOWLEDGEMENTS

The research is supported by National Key Basic Research and Development Program of China (2012CB719900), Supported by National Natural Science Foundation of China (61001187).

REFERENCES

- WOODCOCK, C.E., MACOMBER, S.A., PAX-LENNEY, M. and COHEN, W.B., 2001, Monitoring large areas for forest change using Landsat: generalization across space-time and Landsat sensors, *Remote sensing of Environment*, 78: 194-203.
- COPPIN, P. and BAUER, M., 1994, Processing of multitemporal Landsat TM imagery to optimize extraction of forest cover change features. *IEEE Transactions on Geoscience and Remote Sensing*, 32: 918-927.
- MUCHONEY, D.M. and HAACK, B.N., 1994, Change detection for monitoring forest defoliation, *Photogrammetric Engineering and Remote Sensing*, 60(10): 1243–1251.
- PILON, P.G., HOWARTH, P.J. and BULLOCK, R.A., 1988, An enhanced classification approach to change detection in semi-arid environments. *Photogrammetric Engineering and Remote Sensing*, 54(12): 1709–1716.
- RIDD, M.K. and LIU J.J., 1998, A comparison of four algorithms for change detection in an urban environment, *Remote Sensing of Environment*, 65(2): 95-100.
- ROY, P.S., RANGANATH, B.K., DIWAKAR, P.G., VOHRA, T.P.S., BHAN, S.K., SINGH, I.J. and PANDIAN, V.C., 1991, Tropical forest mapping and monitoring using remote sensing, *International Journal of Remote Sensing*, 12(11): 2205–2225.
- SADER, S.A., POWELL, G.V.N., and RAPPOLE, J.H., 1991, Migratory bird habitat monitoring through remote sensing, *International Journal of Remote Sensing*, 12(3): 363–372.
- VOGELMANN, J.E., 1988, Detection of forest change in the green mountains of Vermont using multispectral scanner data, *International Journal of Remote Sensing*, 9(7): 1187-1200.
- DU P.J., LIU S., LIU P., TAN K. & CHENG L., 2014, Sub-pixel change detection for urban land-cover analysis via multi-temporal remote sensing images, *Geo-spatial Information Science*, 17(1):26-38.
- BOVLO, F. and BRUZZONE, L., 2007, A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain, *IEEE Transactions on Geoscience and Remote Sensing*, 45(1): 218-235.
- NEMMOUR, H., CHIBANI Y., 2006, Multiple support vector machines for land cover change detection: An application for mapping urban extensions. *International Society for Photogrammetry and Remote Sensing*, 61(2): 125-133.
- LU, D., MAUSEL, P., BRONDIZIO, E., MORAN, E., 2004, Change detection techniques. *International Journal of Remote Sensing*, 25(12): 2365-2407.
- RICHARD, J.R., ANDRA, S., AI-KOFAHI, O., ROYSAM, B., 2005, Image Change Detection Algorithms: A systematic Survey, *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 14(3): 294-307.
- SCHOLKOPF, B., PLATT J.C., SHAWE-TAYLOR, J., SMOLA, A.J., 2001, Estimating the Support of a High-dimensional Distribution. *Neural Computation*, 13(7): 1443-1471.
- SUI, H.G., ZHOU, Q.M., GONG, J.Y., MA, G.R., 2008, Processing of multi-temporal data and change detection, *Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences: 2008 International Society for Photogrammetry and Remote Sensing Congress Book*, Edited by Zhilin Li, Jun Chen and Emmanuel Baltsavias, 227-247.
- SCHOLKOPF, B., MIKA, S., BURGESS, C.J.C., KNIRSCH, P., MULLER, K.R., RATSCH, G., SMOLA, A.J., 1999, Input Space versus Feature Space in Kernel-Based Methods. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 10(5): 1000-1017.
- POTIN, D., VANHEEGHE, P., DUFLOS, E., DAVY, M., 2006, An Abrupt Change Detection Algorithm for Buried Landmines Localization, *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, 44(2): 260-272.
- DESOBRY, F., DAVY, M., DONCARLI, C., 2005, An Online Kernel Change Detection Algorithm. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*.53(8): 2961-2974.
- DESOBRY, F., DAVY, M., 2004, Dissimilarity Measures in Feature Space, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 5: 473-476