

# SAR Image Change Detection Based on Correlation Kernel and Multistage Extreme Learning Machine

Lu Jia, Ming Li, *Member, IEEE*, Peng Zhang, *Member, IEEE*, and Yan Wu

**Abstract**—Designing a kernel function with good discriminating ability and a highly application-adaptive kernelized classifier is the key of many kernel methods. However, not many kernel functions combining directly the bitemporal images' information are designed specifically for change detection tasks. In addition, extreme learning machine (ELM) has not found wide applications in change detection tasks, even though it is a potential kernel method possessing outstanding approximation and generalization capabilities as well as great classification accuracy and efficiency. Therefore, an approach relying on a difference correlation kernel (DCK) and a multistage ELM (MS-ELM) is proposed in this paper for synthetic aperture radar (SAR) image change detection. First, a DCK function is constructed specifically for change detection by measuring the “distance” between any two pixels. The DCK function depicts the cross-time similarities between couples of bitemporal image patches at any cyclic shifts with a kernel correlation operation and the high-order spatial distances between two differently located pixels with an algebraic subtraction. The DCK function possesses strong noise immunity and good identification of changed areas simultaneously. Second, an MS-ELM classifier is constructed for obtaining the change detection result. In MS-ELM, the hidden nodes and weights between the hidden and output layers are updated stage by stage by improving the kernel functions that compose them. Each stage of the MS-ELM is a standard kernel-ELM, and the DCK function is utilized in the first stage. The regenerative kernel functions incorporate the output spatial-neighborhood information of the previous stage for enhancing remarkably the MS-ELM's discriminating ability and noise resistance. The converged result at the last stage of MS-ELM is the final change detection result. Experiments on real SAR image change detection demonstrate the effectiveness of the DCK function and the MS-ELM algorithm, particularly its good identification of changed areas and strong robustness against noise in SAR images.

**Index Terms**—Change detection, difference correlation kernel (DCK) function, multistage extreme learning machine (ELM) (MS-ELM), spatial-neighborhood information, synthetic aperture radar (SAR) image.

Manuscript received October 16, 2015; revised March 20, 2016 and May 11, 2016; accepted June 3, 2016. This work was supported in part by the Natural Science Foundation of China under Grants 61272281, 61271297, and 61301284, by the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20130203120006, by the National Ministries and Research Foundation under Grant 9140A07020913DZ01001, and by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2015JM6288.

L. Jia, M. Li, and P. Zhang are with the National Laboratory of Radar Signal Processing and the Collaborative Innovation Center of Information Sensing and Understanding, Xidian University, Xi'an 710071, China (e-mail: liming@xidian.edu.cn).

Y. Wu is with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: ywu@mail.xidian.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2016.2578438

## I. INTRODUCTION

CHANGE detection is a vital research branch of remote sensing image understanding [1]. It aims at identifying changes in images of the same scene taken at different times. It has been applied successfully to civil and military applications, for example, disaster management [2], land cover data set updating [3], urban planning [4], man-made target detection [5], and so on. A synthetic aperture radar (SAR) system creates images by illuminating a target scene with successive pulses of radio waves and receiving and processing the echoes [6]. It operates day and night and under any weather conditions. In the light of the fact that a large number of high-quality SAR images have been obtained from the well-developed spaceborne or airborne SAR systems, there is a good application prospect of developing SAR image change detection techniques. However, the multiplicative speckle noise that inherently contaminates the SAR images makes SAR image change detection a challenging work. Therefore, developing SAR image change detection techniques that are robust against speckle noise is a desperate need now.

SAR image change detection techniques have drawn great attention recently. Unsupervised methods are popular because of their avoidance of human intervention. In most unsupervised methods, producing a difference image to indicate the extent of the changes is the first step. Ratioing [7] is one of the most widely used methods for generating the difference image since its logarithmic pattern can transform the multiplicative speckle noise of SAR images into an additive one. The MS-ITCD image [8] which extracts information-theoretic features through mean shifting the 2-D scatterplot between bitemporal images is another effective one. Given the difference images, binary classifiers, such as the ones based on the multiscale analysis, model, and field theory, should be carried out for obtaining the change detection results. Classifiers based on multiscale analysis gain good performances by extracting features from the approximate and detailed subimages and fusing the features optimally. The discrete wavelet transform (DWT) [9], undecimated DWT [10], and nonsubsampled contourlet transform [11] have found successful applications. Methods based on model analyze SAR images with proper probability distributions or statistical characteristics. The generalized Gaussian distribution [12] and Gamma distribution [13] are the commonly used ones for SAR image change detection. Methods involving the field theory obtain smooth change maps and strong noise resistance by taking the spatial–contextual information as well as the nonstationary and non-Gaussian characteristics of SAR images into consideration. The Markov random

field [2], triplet Markov field [14], and conditional random field [15] are the typical ones. In addition, methods based on the Kullback–Leibler divergence [16], graph-cut theory [17], Radon transform and Jeffrey divergence [18], and other theories [19] have also been successfully applied to SAR image change detection. Occasionally, some prior information is available. To our knowledge, supervised methods theoretically outperform unsupervised ones by considering the prior information.

Supervised methods excel in exploiting the prior information for obtaining an exhaustive description of the changes in images. The simplest supervised change detection method implements a globally optimal threshold on a difference image [7], [8] through trial and error. Kernel methods [20] are other ones utilizing pixels' labels known in advance as the prior information. Two keys, the kernel function and the kernelized classifier, compose the kernel methods. They have three advantages: 1) they have unusual ability in dealing with nonlinear classification; 2) they avoid the curse of dimensionality when the dimension of the input features is extremely high; and 3) the kernel function can be constructed flexibly concerning the practical demands. The aforementioned three merits prove the capability of kernel methods to do nonlinear SAR image change detection. However, kernel methods have not found widespread applications for SAR image change detection. In 2008, Camps-Valls *et al.* [21] propose a supervised kernel-based support vector machine (SVM) method for remote sensing image change detection, in which a difference-kernel function and a ratio kernel (RK) function are constructed. The two kernel functions have strong universality for remote sensing image change detection, and the RK function has considerable adaptability to SAR images. After that, we propose a SAR image change detection method [22] based on an iterative label-information composite (LIC) kernel and SVM. The LIC kernel is obtained by combining the RK function with the spatial information extracted from the output labels. The LIC kernel possesses strong noise immunity and good preservation of edge locations of changed areas simultaneously. Moreover, a semisupervised cluster-neighborhood kernel method [23] is proposed. It exploits the information of unlabeled pixels for enhancing the kernel's noise immunity. All aforementioned kernel methods are founded on the RK function [21] and SVM. Therefore, there is much space to develop a specific kernel function for SAR image change detection. Moreover, as we know, extreme learning machine (ELM) [24] is a promising kernelized classifier possessing universal approximation capability, almost optimal generalization ability, high learning accuracy, and fast learning speed. It is encouraging to use ELM and its improvements for change detection tasks.

In view of the aforementioned analysis, in this paper, we develop a supervised SAR image change detection method based on a difference correlation kernel (DCK) function and a multistage ELM (MS-ELM) classifier. On the one hand, a DCK function which includes a kernel correlation operation followed by an algebraic subtraction is constructed specifically for SAR image change detection. The kernel correlation computes the cross-time similarities between coupled bitemporal image patches at arbitrary cyclic shifts, and the algebraic subtraction computes the distance between any two pixels at

different locations. Both steps utilize the spatial information for providing strong noise immunity for the DCK function, and the kernel correlation provides the DCK with good identification of changed areas, even small dotlike changes. The construction processes of the DCK and RK functions are converse, and the merits of the DCK function can be analyzed theoretically and experimentally. On the other hand, an MS-ELM classifier is implemented on the DCK function for earning the change detection result. MS-ELM has a multistage structure in which the hidden nodes and the weights between the hidden and output layers are updated stage by stage. The updating is realized by regenerating the kernel functions composing the hidden nodes and weights with the spatial-neighborhood information extracted from the output of the previous stage. The last stage can be obtained when the output of the MS-ELM converges, and the converged result is known as the final change detection result. The spatial-neighborhood information in the output space improves further the discriminating ability, noise immunity, and regional smoothness of MS-ELM. The proposed DCK-based MS-ELM method has a similar strategy to the LIC kernel method [22]. In this paper, an in-depth theoretical and experimental comparison between the LIC kernel and the DCK-based MS-ELM is given for showing the characters and merits of the proposed method more elaborately and completely. Moreover, for illustrating further the validity of the proposed method, the DCK function is compared with the RK function, and SVM, ELM, and MS-ELM are compared as classifiers. Two supervised thresholding methods, manually thresholding the log-ratio image and the MS-ITCD image [8], are also tested. Experiments on four real SAR data sets prove that our method does have vigorous discriminating ability, strong robustness against noise, and good identification of changed areas for SAR image change detection.

The organization of this paper is as follows. Section II introduces the fundamentals of the kernel methods and the standard ELM. Section III elaborates the proposed method, including the construction of the DCK function, the formation of the MS-ELM, and the complete steps of the method. After that, a theoretical comparison between the LIC kernel method and the proposed method is given for providing an in-depth explanation of the characters of these two methods. Next, experimental results are presented. The graphical comparison between the DCK function and the RK function is shown first, followed by the real SAR image change detection results and the parameter sensitivity analysis. Finally, Section V gives the conclusion.

## II. KERNEL METHODS AND ELM

The rationales behind the proposed method are the kernel methods and the ELM. In this part, we introduce the basics, including the definitions, characteristics, formulas, and merits, of the kernel methods and the ELM.

### A. Kernel Methods

Broadly speaking, real SAR image processing is nonlinear [6]. Kernel methods offer a framework for machine learning with nonlinear data. When dealing with a nonlinear algorithm

in the input space  $R^d$ , kernel methods embed the samples  $\{x_i\}_i$  defined over  $R^d$  into a high-dimensional Hilbert space  $H^d$  (also known as the kernel space or feature space), and build a linear algorithm therein. In this way, a linear learning machine which is nonlinear with respect to the input space is obtained. The mapping function is commonly represented as  $\phi : R^d \rightarrow H^d$ . In practice, the mapping function is usually hard to determine, and it is a rather heavy load to compute the sample coordinates in  $H^d$  [20]. Therefore, the kernel trick is utilized: If an algorithm was expressed with dot products in the input space, its nonlinear or kernel version could be obtained only with the dot products among mapped samples. Thus, the kernel function  $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$  between pairwise samples is obtained by computing the inner product between mapped samples. The introduction of kernel functions makes kernel methods suitable for handling high-dimensional input data effectively and efficiently.

For classification, kernel values indicate the similarities, also the “distances,” between pixels. The discriminating ability to distinguish pixels is a standard by which we judge the quality of the corresponding kernel function. Therefore, defining a kernel function with strong discriminating ability is the first step of many kernel methods. The flexibility of defining a kernel function according to practical needs makes kernel methods valuable for application-driven learning tasks. The valid kernels are those that fulfill Mercer’s theorem [25]. The most common valid kernels are the linear kernel  $K(x_i, x_j) = x_i \cdot x_j$ , the polynomial kernel  $K(x_i, x_j) = (x_i \cdot x_j + 1)^e$ ,  $e \in Z^+$ , and the radial basis function (RBF) kernel  $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2/2\zeta^2)$ ,  $\zeta \in R$ . Given two Mercer’s kernels, their direct sum, tensor product, and scaling are also valid kernels [25].

By reviewing the literature, we find that the original RK function [21] has high effectiveness for SAR image change detection. By deforming the basic structure of the RK function with additional information, for example, the spatial information in the output space, we propose the advanced LIC kernel function in [22]. Forsaking the former habit of deforming the RK’s structure, in this paper, we construct a DCK function which is independent of the RK function. The construction procedure and the merits of the DCK function are elaborated in Section III-A. Moreover, the graphical exhibition of kernel values in Fig. 8 illustrates that the DCK function possesses good identification of changed areas as well as strong noise resistance in dealing with SAR images.

## B. ELM

ELM is a kernel method for classifying nonlinearly scattered samples [24]. It is proposed for training the “generalized” single hidden layer feedforward neural networks (SLFNs). The hidden nodes of the ELM are not asked to be neuron alike, and they can be randomly initiated and fixed without tuning. In other words, only the weights between the hidden layer and the output layer need to be calculated, which results in a fast training speed. The universal approximation capability and the almost optimal generalization bound of ELM have been proved in [26]–[28]. In [24], a unified ELM is proposed. It not only provides a unified framework for diverse applications, for example, classification

and regression, but also combines various popular networks, such as the SLFNs, LS-SVM, and P-SVM. For the case with a single output node, the unified ELM aims for obtaining the smallest training error  $\xi$  and the smallest norm of the output weight  $\beta$

$$\text{minimize} : L = \frac{1}{2} \|\beta\|^2 + C \cdot \frac{1}{2} \cdot \sum_s \xi_s^2 \quad (1)$$

$$\text{subject to} : \mathbf{h}(x_s)\beta = T_s - \xi_s, s = 1, \dots, N_s. \quad (2)$$

$x_s$  is a training sample, and  $T_s$  is its label.  $N_s$  is the number of the training samples.  $C$  is a balance parameter.  $\mathbf{H}$  and  $\mathbf{T}$  are the hidden layer matrix and the output matrix, respectively

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(x_1) \\ \vdots \\ \mathbf{h}(x_{N_s}) \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} T_1 \\ \vdots \\ T_{N_s} \end{bmatrix}, \quad \mathbf{H} \cdot \beta = \mathbf{T}. \quad (3)$$

For the case where the number of the training samples is not huge, the output of the unified ELM [24] is

$$f(x_i) = \mathbf{h}(x_i) \mathbf{H}^T \left( \frac{\mathbf{I}}{C} + \mathbf{H} \mathbf{H}^T \right)^{-1} \mathbf{T}. \quad (4)$$

$x_i$  is a test sample.  $T$  is the matrix transpose. For binary classification, the output label of  $x_i$  is

$$Y(x_i) = \text{sign}(f(x_i)). \quad (5)$$

The hidden nodes  $\mathbf{h}(\bullet)$  can be almost all nonlinear piecewise continuous functions available [24]. Varied hidden nodes correspond to varied feature mappings. When hidden nodes are unknown, the authors of [24] prove that kernel functions that satisfy Mercer’s theorem can be utilized, and thus the kernel-ELM. Theoretically, all Mercer’s kernels can be used in the kernel-ELM [29]. The output of the kernel-ELM is

$$f(x_i) = \begin{bmatrix} K(x_i, x_1) \\ \vdots \\ K(x_i, x_{N_s}) \end{bmatrix}^T \left( \frac{\mathbf{I}}{C} + \mathbf{K}_{\text{train}} \right)^{-1} \mathbf{T} \quad (6)$$

$$\mathbf{K}_{\text{train}} = \mathbf{H} \mathbf{H}^T : K_{\text{train}}(x_s, x_{s'}) = \mathbf{h}(x_s) \mathbf{h}(x_{s'})^T = K(x_s, x_{s'}). \quad (7)$$

$x_s$  and  $x_{s'}$  are training samples.

As a unified learning framework of the SVM, LS-SVM, and P-SVM, ELM outperforms the other three methods in both classification accuracy and efficiency [24]. Recently, many variants of ELM have been proposed, for example, the incremental ELM [27], the online sequential ELM [30], the weighted ELM [31], and the semisupervised ELM [32]. These improvements can be applied to diverse applications, such as computer vision, image processing, medical diagnosis, system modeling, and prediction. All these improvements are realized on the original ELM, not the kernelized one. In this paper, we improve the kernel-ELM for enhancing its adaptability to SAR image change detection.

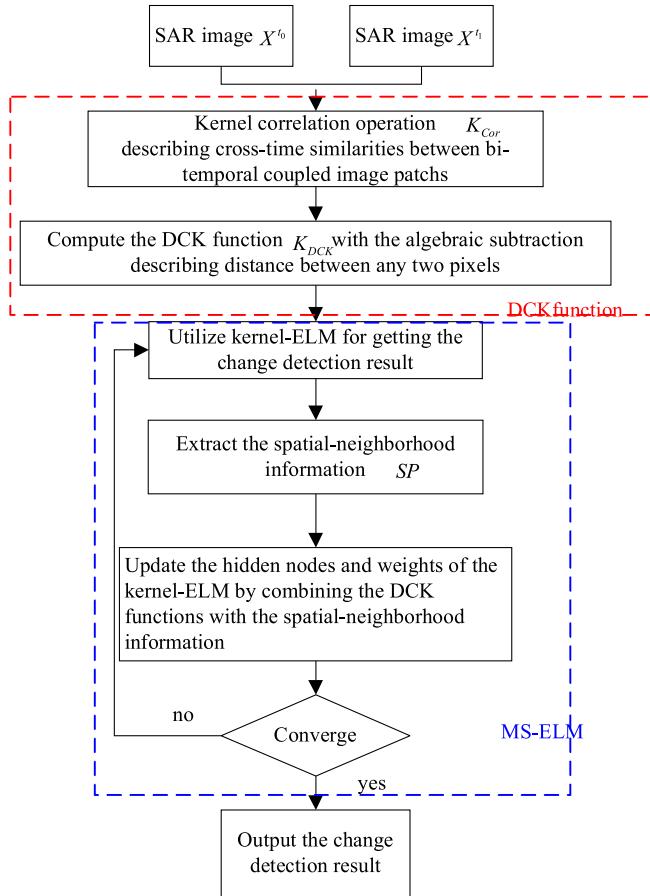


Fig. 1. Flowchart of the proposed DCK-based MS-ELM change detection method.

### III. SAR IMAGE CHANGE DETECTION WITH DCK FUNCTION AND MS-ELM

In this section, theories and characteristics of the DCK-based MS-ELM algorithm are introduced. Note that we design the DCK function and the MS-ELM specifically for SAR image change detection since both of them have unusual ability in handling noise-contaminated SAR images by taking the spatial-neighborhood information into consideration. The construction of the DCK function relies on a kernel correlation measuring the similarities between coupled bitemporal image patches and a subtraction measuring distances between differently located pixels. Its construction is converse to that of the RK function in some degree. The MS-ELM has a multistage structure in which its hidden nodes and weights are updated stage by stage on the basis of the kernel functions. Both DCK function and MS-ELM possess strong noise immunity by absorbing the spatial information, and the DCK has good identification of changed areas because of the kernel correlation. Given the bitemporal coregistered SAR images  $\{X^{t_0}, X^{t_1}\}$ , Fig. 1 shows the flowchart of the proposed DCK-based MS-ELM change detection method. Principals and merits of the DCK function and the MS-ELM are introduced in detail in this section. Complete steps of the proposed method are elaborated. Finally, theoretical comparison between the LIC kernel and the DCK-based MS-ELM is given for providing deep explanation of characters of the proposed method.

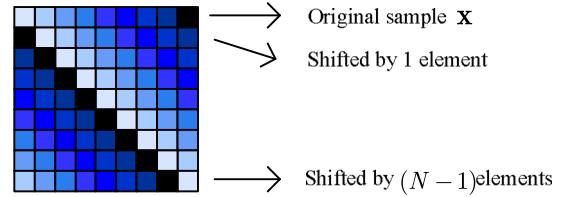


Fig. 2. Illustration of the circulant matrix.

#### A. DCK Function

The notion of the DCK function is inspired by the work of Henriques *et al.* [33] in which a kernel correlation operation is proposed for high-speed video target tracking. Henriques proves that, when the translated and scaled image patches are used as the training samples, the kernel regression is equivalent to a kernel correlation. As a kernelized version of the linear correlation filter, the kernel correlation measures the correlations, i.e., similarities, between the coupled bitemporal patches at all cyclic shifts. Moreover, he brings up a fast implementation of the kernel correlation. Experiments show the validity of the fast kernel correlation in measuring similarities between image patches at different time points [33]. Inspired by the aforementioned analysis, we build a DCK function on the basis of the fast kernel correlation. Given a vector  $\mathbf{x} = [x_1, x_2, \dots, x_N]$  representing an object patch of interest, 1-D translations of it can be modeled with a cyclic shift operation  $P$ , which results in the circulant matrix  $\bar{\mathbf{X}}$

$$\begin{aligned}
 \bar{\mathbf{X}} &= \{P^u \mathbf{x} \mid u = 0, 1, \dots, N - 1\} \\
 P &= \begin{bmatrix} 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \\
 \bar{\mathbf{X}} &= \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_N \\ x_N & x_1 & x_2 & \cdots & x_{N-1} \\ x_{N-1} & x_N & x_1 & \cdots & x_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_2 & x_3 & x_4 & \cdots & x_1 \end{bmatrix}. \quad (8)
 \end{aligned}$$

Obviously, the circulant matrix is specified by the generating vector  $\mathbf{x}$  in the first row. The visual illustration of the circulant matrix  $\bar{\mathbf{X}}$  is given in Fig. 2.

With the aforementioned definitions, Henriques *et al.* [33] define a kernel correlation as follows:

$$\mathbf{k}^{\mathbf{z}\mathbf{x}} = \{k_u^{\mathbf{z}\mathbf{x}} = k(\mathbf{z}, P^{u-1} \mathbf{x})\}_u = k(\mathbf{z}, \bar{\mathbf{X}}). \quad (9)$$

The kernel correlation measures the kernelized correlation between vectors  $\mathbf{z}$  and  $\mathbf{x}$  at all relative shifts. Taking the advantage of the fact that the convolution of two patches is equivalent to the elementwise product in the Fourier domain [33], the kernel correlation  $\mathbf{k}^{\mathbf{z}\mathbf{x}}$  at all shifts can be calculated fast at once. Three fast kernel correlation operations, the dot-product, polynomial, and Gaussian kernel correlation, are derived in [33]. Because of

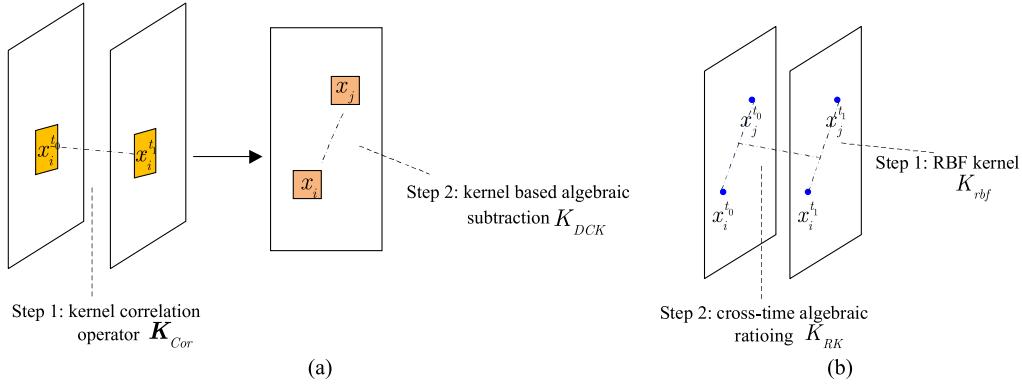


Fig. 3. Sketch maps of the construction procedures of two kernel functions. (a) DCK function. (b) RK function.

its good performance and widespread adaptability, the Gaussian kernel correlation is usually used

$$\mathbf{k}^{\mathbf{zx}} = \exp\left(-\frac{1}{\zeta^2} (\|\mathbf{z}\|^2 + \|\mathbf{x}\|^2 - 2F^{-1}(\hat{\mathbf{z}} \odot \hat{\mathbf{x}}^*))\right). \quad (10)$$

$F^{-1}$  is the inverse DFT,  $\odot$  is the elementwise product, and  $\zeta$  is the kernel width parameter.

In the light of the fast and effective implementation of the kernel correlation on measuring cross-time similarities, we try to construct a DCK function for bitemporal SAR image change detection. For change detection tasks, two images  $\{X^{t_0}, X^{t_1}\}$  taken at the same scene but different time points are available. Supposing that  $\{x_i^{t_0}, x_i^{t_1}\}$  is the  $i$ th bitemporal pixel couple,  $\{x_i^{t_0}, x_i^{t_1}\} = \{x_{ip}^{t_0}, x_{iq}^{t_1} | \forall ip \in N_i^{t_0}, iq \in N_i^{t_1}\}$  is the corresponding neighborhoods, and  $N$  is the neighborhood size, we use the kernel correlation for computing the cross-time correlation between the two neighborhoods at all relative shifts

$$\mathbf{K}_{\text{Cor}}(\mathbf{x}_i^{t_0}, \mathbf{x}_i^{t_1}) = \mathbf{k}^{\mathbf{x}_i^{t_0} \mathbf{x}_i^{t_1}}. \quad (11)$$

By substituting the Gaussian kernel correlation into (11), we obtain

$$\begin{aligned} \mathbf{K}_{\text{Cor}}(\mathbf{x}_i^{t_0}, \mathbf{x}_i^{t_1}) \\ = \exp\left(-\frac{1}{\zeta^2} (\|\mathbf{x}_i^{t_0}\|^2 + \|\mathbf{x}_i^{t_1}\|^2 - 2F^{-1}(\hat{\mathbf{x}}_i^{t_0} \odot \hat{\mathbf{x}}_i^{t_1}))\right). \end{aligned} \quad (12)$$

$\hat{\mathbf{x}}_i^{t_0}(\hat{\mathbf{x}}_i^{t_1})$  is the DFT of the vector  $\mathbf{x}_i^{t_0}(\mathbf{x}_i^{t_1})$  which is made up of the neighbors of  $x_i^{t_0}(x_i^{t_1})$ . The kernel correlation depicts the cross-time correlation between bitemporal image patches (defined by the neighborhood). It has the size of  $1 \times N$ , and we record it as  $\mathbf{K}_{\text{Cor}}(\mathbf{x}_i^{t_0}, \mathbf{x}_i^{t_1}) = \{K_{\text{Cor}}^i(m)\}_{m \in N}$ . For change detection tasks, large values in  $\mathbf{K}_{\text{Cor}}(\mathbf{x}_i^{t_0}, \mathbf{x}_i^{t_1})$  depict that pixels  $x_i^{t_0}$  and  $x_i^{t_1}$  are highly correlated and no change occurs between them. On the contrary, small values in  $\mathbf{K}_{\text{Cor}}(\mathbf{x}_i^{t_0}, \mathbf{x}_i^{t_1})$  mean that change may occur between  $x_i^{t_0}$  and  $x_i^{t_1}$ . Not only the pixel couple but also their neighbors, i.e., spatial information, at all relative shifts are considered in  $\mathbf{K}_{\text{Cor}}$ .

With (12), we define a DCK function for describing the distance between two pixels at different locations. It is

$$\begin{aligned} K_{\text{DCK}}(x_i, x_j) = & \left| \frac{1}{N} \sum_{m \in N} K_{\text{Cor}}^i(m) - \frac{1}{N} \sum_{m \in N} K_{\text{Cor}}^j(m) \right| \\ & + \lambda \delta(x_i, x_j) \end{aligned} \quad (13)$$

where  $\delta(A, B) = 1$ , if  $A = B$ , and  $\delta(A, B) = 0$ , if  $A \neq B$ .  $x_i$  and  $x_j$  are two pixels at different locations. Parameter  $\lambda$  is used for making the kernel function positive definite.  $N$  is the neighborhood size. Obviously, we use the spatial-neighborhood information which can be regarded as high-order features and can definitely increase the noise resistance of the kernel function.

Equations (12) and (13) compose the DCK function. In this paper, the RK function [21] is tested for comparison. The formula of the RK function is given hereinafter. The RBF kernel is utilized in the RK function

$$\begin{aligned} K_{\text{RK}}(x_i, x_j) &= \frac{K_{\text{rbf}}(x_i^{t_0}, x_j^{t_0})}{K_{\text{rbf}}(x_i^{t_1}, x_j^{t_1})} + \lambda \delta_{ij} \\ K_{\text{rbf}}(x_i, x_j) &= \exp(-\|x_i - x_j\|^2 / 2\zeta^2), \zeta \in R. \end{aligned} \quad (14)$$

Based on the formulas of the DCK and RK functions, we analyze the differences between these two kernels. First, to some extent, the DCK function and the RK function have converse construction procedures. Sketch maps of the construction procedures of the DCK and RK functions are shown in Fig. 3. For the DCK function, the kernel correlation between coupled image patches at the same location but different times is computed first. Then, the algebraic subtraction between pixels at different locations is carried out. For the RK function, the RBF kernel is computed first with pixels at the same time but different locations. Then, the algebraic ratio is implemented with the cross-time RBF kernels.

Second, it is found that values of the DCK function and the RK function have converse tendency. If pixels  $x_i$  and  $x_j$  belong to the same class (changed or unchanged), their DCK value is small. If pixels  $x_i$  and  $x_j$  belong to different classes (one is changed, and the other is unchanged), their DCK value is relatively large. However, the RK function produces the contrary. Pixels in the same classes have large RK values, and pixels in different classes have small RK values. The phenomenon is caused by the converse construction procedures analyzed earlier. To clarify the phenomenon further, Fig. 4 shows the kernel matrices of the “Hidden tanks” images. Both the DCK and RK matrices are computed with 120 training samples selected randomly. The information of the “Hidden tanks” images and the locations of the training samples are shown in Table I and Fig. 8(a1). Fig. 4(a) and (b) shows the kernel matrices obtained

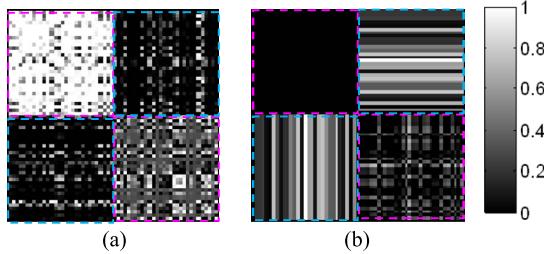


Fig. 4. Exhibition of kernel matrices. (a) RK matrix. (b) DCK matrix.

TABLE I  
KNOWLEDGE OF REAL SAR IMAGE

Images	Hidden tanks	Japan city flooding	Sweden air-vehicle	Indonesia flooding
Image source	LYNX-S AR	Terra SAR-X	CARABAS -II	ERS-I
Date	--	2010.10.20	2002.04.05	1994.2.16
	--	2011.06.05		1994.3.6
ENL	4 look	1 look	3 look	8 look
Size	254×238	262×236	240×188	256×256
Polarization	VV	VV	HH	VV
Changes	Exposure of hidden tanks	Flood event	Take-off of air-vehicle	Flood event

with the RK function and the DCK function, respectively. In Fig. 4(a) and (b), the up-left and down-right pink rectangles show kernel values between samples in the same classes, and the down-left and up-right blue rectangles show kernel values between changed and unchanged samples. Obviously, the RK matrix possesses large values for pixels in the same classes and small values for pixels in different classes, whereas the DCK matrix possesses small values for pixels in the same classes and large values for pixels in different classes. This phenomenon is in accordance with the converse construction procedures. Since change detection aims at classifying pixels into changed or unchanged classes, it is expected that the difference between values in pink rectangles and values in blue rectangles should be large enough for providing the kernel matrices with the strong discriminating ability. According to Fig. 4(a) and (b), both the RK and DCK functions are valid for discriminating the pixels, even if their measurements have contrary tendencies. In a word, both the RK and DCK functions offer acceptable “distances” (or similarities) between pixels.

Finally, these two kernels have different performances. The DCK function utilizes the high-order spatial neighborhood, whereas the RK function does not. Theoretically, the spatial information could improve the DCK’s immunity to noise. Moreover, the DCK employs the kernel correlation which could experimentally provide good identification of changed areas (even small dotlike changes). The graphical exhibition of kernel values is given in Fig. 8 to show the two advantages, strong noise and good identification of changed areas, of the DCK function. With its explicit physical meaning, effective “distance” measurement, strong noise immunity, and good discriminating ability, the DCK function is a good candidate for SAR image change detection.

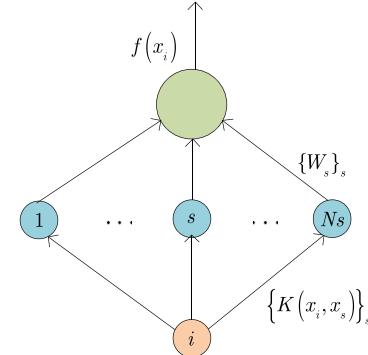


Fig. 5. Structure of kernel-ELM.

### B. MS-ELM

In Section II-B, the basic formulas and properties of the kernel-ELM are summarized. With its high classification accuracy and efficiency, kernel-ELM is competent for change detection tasks. For improving the noise resistance of kernel-ELM in dealing with SAR images, we propose an MS-ELM by considering the spatial information in the output space. First, we draw the structure of the traditional kernel-ELM by referring to (6). It shows why and how the kernel-ELM can be transformed into MS-ELM. It is obvious from (6) that two keys compose the target function of the kernel-ELM. They are the hidden nodes and the weights between the hidden and output layers. Given a test pixel  $x_i$  and the training pixels  $\{x_s, s = 1, \dots, N_s\}$ ,  $\{K(x_i, x_s)\}_s$  which are computed between the test and training pixels are regarded as the hidden nodes, and  $\mathbf{W} = \{W_s\}_s = ((\mathbf{I}/C) + \mathbf{K}_{\text{train}})^{-1}\mathbf{T}$  contains the weights. Note that  $\mathbf{K}_{\text{train}}$  is the kernel matrix computed with training samples.  $\mathbf{T}$  contains labels of training samples that are known in advance. By referring to [29], the structure of the kernel-ELM is shown in Fig. 5. Evidently, both hidden nodes and weights are determined only by the kernel functions. Therefore, the kernel-ELM can be improved by just deforming the kernel functions. For improving the noise immunity of kernel-ELM in handling SAR images, we consider incorporating the spatial information in local neighborhoods into the kernel functions. The explicit construction of MS-ELM is elaborated in the following part. By the way, both RK and DCK functions adapt to the kernel-ELM in change detection tasks.

In this part, the principle of the MS-ELM is elaborated. On the whole, the MS-ELM is obtained by improving kernel functions of the kernel-ELM with spatial-neighborhood information stage by stage. We first draw the structure of the MS-ELM in Fig. 6 for exhibiting its construction procedure clearly. At each stage of the MS-ELM, a standard ELM is implemented. As is shown in Fig. 6, the kernel functions at the  $(b+1)_\text{th}$  stage are generated by combining the kernel functions at the first stage and the spatial information extracted at the  $b_\text{th}$  stage. That is, the kernel functions are updated and regenerated by deforming the original kernel function with the spatial-neighborhood information extracted from the output of the previous stage. In this way, the hidden nodes and weights between hidden nodes and outputs are updated iteratively, and

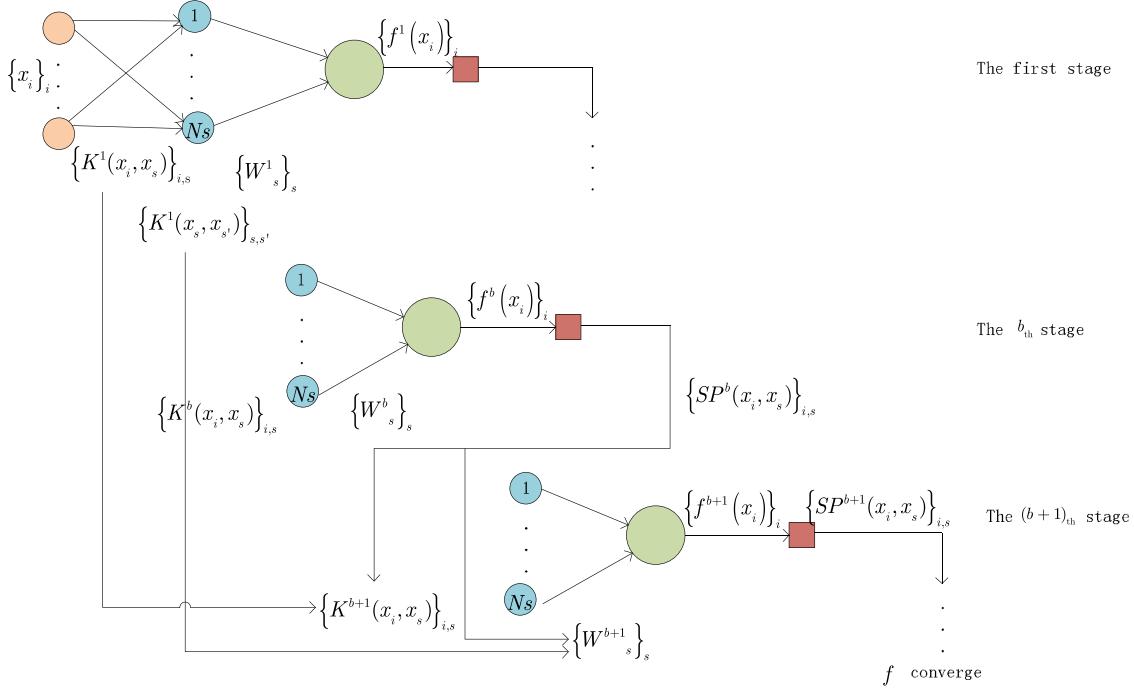


Fig. 6. Structure of MS-ELM.

thus, a new multistage kernel-ELM is constructed. In this paper, the original kernel functions used at the first stage are the DCK functions. Because of the contrary behavior of the RK function to the DCK function, the RK function can be applied to MS-ELM if and only if its way of representing similarities (see Fig. 4) is reversed.

On the basis of the visual structure of the MS-ELM in Fig. 6, we give the detailed derivation of MS-ELM as follows. Given a test pixel  $x_i$ , hidden nodes and the weights between the hidden and output layers in the  $b_{\text{th}}$  stage are

$$\{K^b(x_i, x_s)\}_s, \text{ and } \mathbf{W}^b = \{W_s^b\}_s = \left( \frac{\mathbf{I}}{C} + \mathbf{K}^b_{\text{train}} \right)^{-1} \cdot \mathbf{T}, \\ \mathbf{K}^b_{\text{train}} = \{K(x_s, x_{s'})\}_{s,s'}. \quad (15)$$

$b > 1$  is the stage number.  $x_s$  and  $x_{s'}$  are training pixels, and  $s, s' = 1, \dots, N_s$ .  $\mathbf{T}$  is defined in (3).  $C$  is defined in (1). For  $x_i$ , we implement the standard kernel-ELM for obtaining the output  $f^b(x_i)$  in stage  $b$

$$f^b(x_i) = \begin{bmatrix} K^b(x_i, x_1) \\ \vdots \\ K^b(x_i, x_{N_s}) \end{bmatrix} \cdot \mathbf{W}^b. \quad (16)$$

With all test pixels' outputs  $\{f^b(x_i)\}_i$ , the spatial-neighborhood information around  $x_i$  and related to  $x_s$  can be extracted using the following expression:

$$SP^b(x_i, x_s) = \left( \frac{1}{N} \sum_{i \in N_i} f^b(x_{il}) \right) \cdot T_s \cdot (-1). \quad (17)$$

$T_s$  is the label of the training pixel  $x_s$ .  $N_i$  indicates the neighborhood of  $x_i$ , and  $N$  is the neighborhood size. For simplicity, we make the neighborhood sizes in (17) and (13) the same.  $SP^b(x_i, x_s)$  is the average output in the neighborhood of  $x_i$ .

Then, the hidden nodes and weights in stage  $(b+1)$  are updated with

$$K^{b+1}(x_i, x_s) = \eta \cdot K^1(x_i, x_s) + (1-\eta) \cdot SP^b(x_i, x_s) \quad (18)$$

$$\mathbf{W}^{b+1} = \{W_s^{b+1}\}_s = \left( \frac{\mathbf{I}}{C} + \mathbf{K}^{b+1}_{\text{train}} \right)^{-1} \cdot \mathbf{T} \quad (19)$$

$$\mathbf{K}^{b+1}_{\text{train}} = \{K^{b+1}(x_s, x_{s'})\}_{s,s'} \\ = \{\eta \cdot K^1(x_s, x_{s'}) + (1-\eta) \cdot SP^b(x_s, x_{s'})\}_{s,s'}. \quad (20)$$

$\eta$  is a balance parameter that constitutes a tradeoff between the original kernels and the output spatial information. In our method, the DCK function [see (13)] is utilized at the first stage for computing the hidden nodes  $\{K^1(x_i, x_s)\}_s$  and the weights  $\mathbf{W}^1$ . With the hidden nodes and weights in stage  $(b+1)$ , the classification output  $f^{b+1}(x_i)$  in stage  $(b+1)$  can be obtained with (16). The iterative procedure continues [like the one from stage  $b$  to  $(b+1)$ ] until the output converges, and then, we get the final stage. For change detection, (5) is carried out with the converged output for obtaining the final change detection results.

Equations (18)–(20) show that the spatial-neighborhood information  $SP$  is the key for updating the kernel functions. Now, let us analyze the practicability and validity of the output spatial-neighborhood information  $SP$ , and then the validity of the MS-ELM algorithm. According to (5), it is obvious that, at each stage, the classification output  $f(x_i)$  and the label are positive or negative simultaneously. When a test pixel  $x_i$  and a training pixel  $x_s$  belong to the same class, i.e.,  $f(x_i)$  and  $T_s$  have the same sign, their  $SP^b(x_i, x_s)$  value is negative; when  $x_i$  and  $x_s$  belong to different classes, their  $SP^b(x_i, x_s)$  value is positive. It is shown in Fig. 4(b) that, for obtaining better discriminating ability, the DCK values between pixels

in the same classes should be as small as possible and pixels in different classes should have large DCK values. Hence, the DCK and  $SP$  describe the “distances” between pixels with the same tendency. Since the output at the previous stage is a kind of prior information, adding the output spatial-neighborhood information to the DCK function could enhance its discriminating ability. For MS-ELM, the multistage structure soundly utilizes the neighborhood information in the previous output for improving the hidden nodes and weights of the next stage. The neighborhood information is a kind of high-order spatial information which could increase the noise immunity and the regional smoothness of the method. Therefore, the MS-ELM possesses strong discriminating ability, good regional smoothness, and excellent noise immunity simultaneously.

### C. Complete Steps of Proposed DCK-Based MS-ELM Change Detection Method

With the explicit construction formulas of the DCK function and the MS-ELM, the complete implementation of the proposed change detection method is interpreted in detail in this part, which is an elaboration of Fig. 1. For a practical implementation related to SAR images, we only utilize the gray values. It is because the gray values indicate holistically the intensities of the changes. Moreover, the gray values are severely contaminated by the speckle noise, which could reveal the strong noise immunity of the proposed method.

- 1) Given the bitemporal SAR images, extract their gray values as the intensity features. Note that the intensity features should be normalized before the DCK function is constructed.
- 2) Select training samples  $x_s, s = 1, \dots, N_s$ , manually.
- 3) With all bitemporal pixels (both training and test pixels) and their neighborhoods, construct the DCK function using (12) and (13). In this step, two kernels are built. One describes similarities between two training samples [ $K^1_{\text{train}}$  in (15)], and the other describes similarities between training and test pixels [ $K^1(x_i, x_s)$  in (15)].
- 4) Let the stage  $b = 1$ . Compute the weights  $\mathbf{W}^b$  between the hidden nodes and the outputs with (15).
- 5) Input  $\{K^b(x_i, x_s)\}_{i,s}$  and  $\mathbf{W}^b$  into (16) for obtaining the outputs  $\{f^b(x_i)\}_i$  at stage  $b$ .  $i$  and  $s$  are the suffixes of the test and training samples.
- 6) With the outputs, extract the spatial-neighborhood information  $\{SP^b(x_i, x_s)\}_{i,s}$  at stage  $b$  with (17).
- 7) Update  $\mathbf{W}^{b+1}$  and  $\{K^{b+1}(x_i, x_s)\}_{i,s}$  at stage  $(b + 1)$  using (18)–(20).
- 8) If  $\{f^b(x_i)\}_i$  converges, implement the final steps; if not,  $b = b + 1$  and go back to step 5).
- 9) Output the final change map using  $\{Y(x_i)\}_i = \{\text{sign}(f^b(x_i))\}_i$ .

### D. Comparison Between LIC Kernel Method and Proposed DCK-Based MS-ELM Method

In previous work, we propose an iterative LIC kernel [22] change detection method. The LIC kernel is constructed

successfully by linearly combining the RK function with the output-space label-neighborhood information extracted under the supervision of anisotropic texture analysis. It is implemented in an iterative mode which updates, in turn, the change map outputted from the SVM and the LIC kernel function to gain the optimal converged result. Therefore, the LIC kernel and the DCK-based MS-ELM methods employ rather similar strategies. In this part, we summarize the similarities and differences between these two methods for providing an in-depth explanation of characters of them. For more details of the LIC kernel method, please refer to [22].

Similarities between these two methods lie in their implementation strategies. First, they both incorporate spatial information into kernel functions for providing strong noise immunity. Then, they both employ iterative procedures for updating kernel functions iteratively, and the iterations end when the change detection results converge.

However, they have the following differences. First, they utilize different kernel functions. The LIC kernel is built on the RK function, whereas the DCK function is independent of the RK function. The advantage of the DCK function over the RK function provides the proposed method with strong identification of changed areas. It can be seen from the experiments (see Section IV-B and C) that even small changes like dots can be identified by the DCK function. Second, they utilize different classifiers. The LIC method utilizes an iterative SVM, whereas the proposed one utilizes an iterative MS-ELM. Obviously, the MS-ELM possesses stronger noise immunity and higher efficiency over the iterative SVM. Finally, they extract different spatial information. In the LIC kernel method, the spatial information is extracted under the supervision of the anisotropic texture analysis. That is, in edges, the spatial information in specific orientations is extracted. In this way, the LIC kernel produces change maps with strong noise resistance as well as good identification of edge locations. However, the whole neighborhood information is extracted in the DCK-based MS-ELM method. It provides strong noise immunity, but it blurs the edges. Moreover, the anisotropic texture analysis increases the computational burden of the LIC kernel method.

The aforementioned differences lead to different performances. For large changed areas, the LIC kernel has good regional smoothness, strong noise immunity, and precise edge locations. It derives from the effective extraction of the spatial information under the supervision of the anisotropic texture analysis. However, for small changed pieces, the anisotropic texture analysis loses efficacy. Therefore, small changed regions, particularly those like dots, may be blurred. For the DCK function, the kernel correlation enhances its ability in identifying small changes like dots. Even though the spatial information is utilized, the dotlike changes can still be identified. Since the DCK-based MS-ELM only considers the whole neighborhood information, it cannot preserve the edge locations well when dealing with large changed areas. In a word, the LIC kernel method is good at resisting noise and preserving edge locations, whereas the DCK-based MS-ELM method is good at resisting noise and identifying small changes (even dotlike). Experimental results in the next section verify their different performances further.

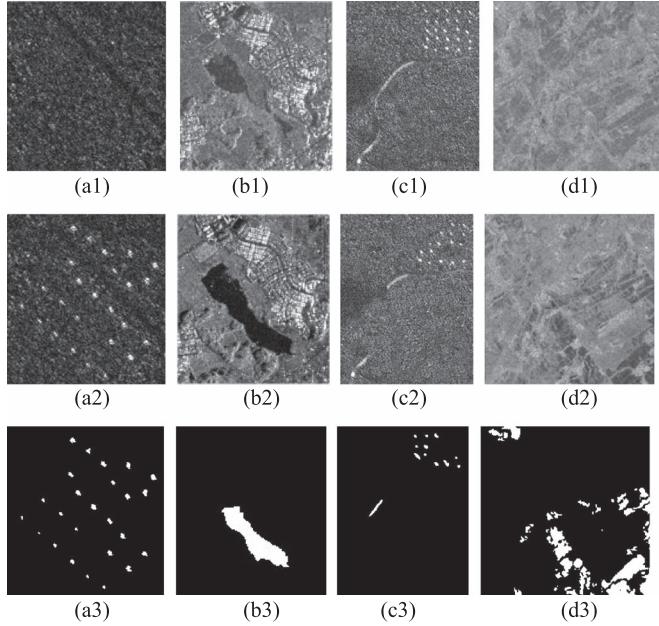


Fig. 7. Real SAR images utilized. (a1)–(d1) Pre-event images. (a2)–(d2) Postevent images. (a3)–(d3) Reference images. (a1)–(a2) “Hidden tanks” images. (b1)–(b2) “Japan city flooding” images. (c1)–(c2) “Sweden air-vehicle” images. (d1)–(d2) “Indonesia flooding” images.

#### IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, experimental results are shown to illustrate the validity of the proposed change detection method. First, we view the experimental setup, including the information of the real SAR images used and the parameters predefined. Characteristics of the changes in these images are introduced, which is the motivation behind the different performances of the compared methods. Then, values of the DCK and RK functions are shown as images to illustrate the strong noise immunity and good identification of changed areas of the DCK function. Next, subjective and objective results of seven supervised change detection methods are analyzed and compared, which show further the strong noise immunity and good regional smoothness of the MS-ELM. Finally, the sensitivity of the proposed method to two parameters is analyzed for showing the robustness of the method.

##### A. Experimental Setup

Four sets of real SAR images shown in Fig. 7 are utilized for illustrating the effectiveness and adaptability of the proposed kernel method. Prior information of the real SAR images applied is shown in Table I. In the “Hidden tanks” data set, 25 hidden tanks are exposed. Consequently, 25 small dotlike changes occur. Similarly, in the “Sweden air-vehicle” data set, 11 small dotlike changes exist because of the takeoff of the air-vehicles. These two images are applied for illustrating the good identification of changed areas (particularly small changes) of the DCK function. A large and smooth changed area exists in the “Japan city flooding” data set. It is utilized for showing the strong noise immunity of the method. However, an area with light color may bring some false detections. It shows further the greater identification of changed areas of the DCK function over the RK function. Both large and small changes

TABLE II  
PARAMETERS USED

Images	Hidden tanks	Japan city flooding	Sweden air-vehicle	Indonesia flooding
Neighborhood-size $N$	$3 \times 3$	$5 \times 5$	$3 \times 3$	$3 \times 3$
Number of samples per class	60	60	60	60

exist in the “Indonesia flooding” data set. It shows the overall ability of the proposed method in resisting noise and preserving small changes.

In the DCK and RK functions, the kernel width  $\zeta$  [see (12) and (14)] needs to be tuned. In SVM, the regularization parameter  $C$  needs to be tuned [21]. In MS-ELM, the parameter  $C$  in (6) and the balance parameter  $\eta$  in (18) and (20) are adjustable. In all experiments, these three parameters are tuned with the training data through a tenfold cross-validation [34] in the range  $\zeta = \{10^{-3}, \dots, 10^3\}$ ,  $C = \{10^{-3}, \dots, 10^3\}$  and  $\eta = [0, 1]$  for finding the best parameter combination. Moreover, two parameters are selected manually. They are the neighborhood size in (13) and (17), from which the spatial information is extracted, and the number of training samples per class. Performances varying these two parameters will be discussed in the final part of this section. Moreover, results in Sections IV-B and C are obtained with the parameters in Table II. In all experiments, the training samples are selected randomly across the images, and all pixels minus the training samples are tested. The bitemporal images are coregistered first. The registration precision is within one pixel. The reference images are defined manually.

##### B. Illustration of Kernel Functions

It is shown in Section III-A that, even though the DCK and the RK functions have converse construction procedures, they both measure the “distances” (or similarities) between pixels in a reasonable and reliable way. In other words, they both possess good discrimination between changed and unchanged samples. Moreover, it is mentioned that the kernel correlation could improve the DCK’s identification of changed areas and the spatial-neighborhood information could increase the DCK’s robustness against noise. In this section, the two merits are confirmed visually.

In Fig. 8, the DCK and RK values of real SAR images are shown as images. In Fig. 8(a1)–(d1), training samples randomly selected are marked on the reference images. The red stars indicate 60 unchanged training samples, and the green stars indicate 60 changed samples. Since kernel functions describe the “distance” between any two pixels, kernel values in Fig. 8(a2)–(d3) are the average “distances” between all pixels and the unchanged training samples. For a test pixel  $x_i$  and a training sample  $x_s$ , the following kernel function is computed:

$$K(x_i, \mathbf{x}_{NC_{train}}) = \frac{1}{|NC_{train}|} \sum_{s \in NC_{train}} K(x_i, x_s). \quad (21)$$

$NC_{train}$  indicates the set of the suffixes of the unchanged training samples, and  $|NC_{train}|$  counts its number.  $K(x_i, \mathbf{x}_{NC_{train}})$  is the average “distance” between  $x_i$  and the unchanged training

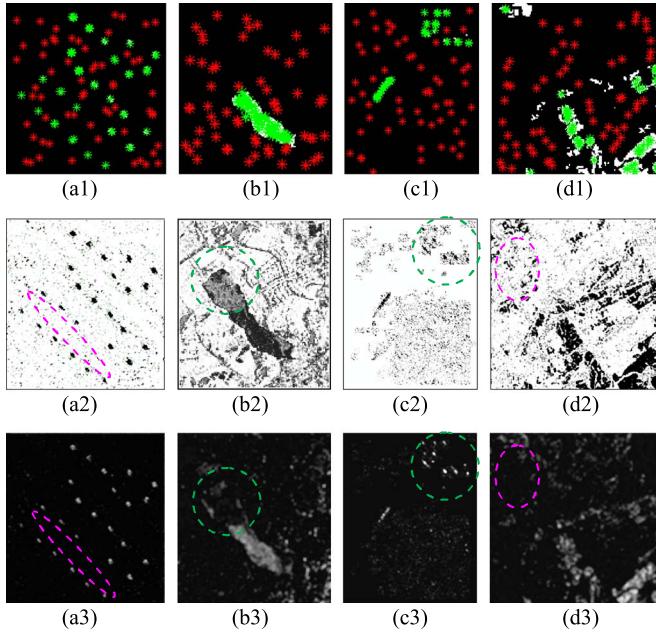


Fig. 8. Illustration of kernel values. (a1)–(d1) Training samples on reference images. (a2)–(d2) RK values. (a3)–(d3) DCK values. (a1)–(a3) “Hidden tanks” images. (b1)–(b3) “Japan city flooding” images. (c1)–(c3) “Sweden air-vehicle” images. (d1)–(d3) “Indonesia flooding” images.

samples. Substituting (13) and (14) into (21), we got the following DCK and RK values, respectively:  $\{K_{\text{DCK}}(x_i, \mathbf{x}_{NC_{\text{train}}})\}_i$  and  $\{K_{\text{RK}}(x_i, \mathbf{x}_{NC_{\text{train}}})\}_i$ . The two kernels are shown graphically in Fig. 8(a3)–(d3) and (a2)–(d2).

First, it can be seen from Fig. 8(a2)–(d2) that the RK function produces large values between unchanged test and training pixels and small values between changed test and unchanged training pixels. The DCK function produces the converse according to Fig. 8(a3)–(d3). The phenomenon, the converse measurement of “distances,” stems from the converse construction procedures in Fig. 3 and is consistent with the kernel matrices in Fig. 4. Obviously, both kernels can separate the changed and unchanged pixels. Second, kernel values in several typical areas are analyzed for showing the advantages of the DCK function. On the one hand, pink ellipses in Fig. 8(a2), (a3) and (d2), (d3) show that the DCK offers clearer unchanged areas, which means that it has stronger noise immunity compared with the RK function. The great suppression of the degrading effects of noise lies in the high-order spatial-neighborhood information utilized in the DCK [see (12) and (13)]. On the other hand, green ellipses in Fig. 8(b2), (b3) and (c2), (c3) show that the DCK function identifies the changed areas more precisely, particularly small dot changes. This strong identification derives from the kernel correlation operation [see (12)] which could highlight the changes that may be lost by the RK function. In summary, the DCK function has two merits, good identification of changed areas and strong noise immunity, simultaneously.

### C. Change Detection Results on Real SAR Images

For illustrating the validity of the proposed DCK-based MS-ELM algorithm, seven change detection methods are tested and compared. They are the supervised thresholding method

with the log-ratio images and the MS-ITCD images [8], the LIC kernel method [22], the RK-based SVM [21], the DCK-based SVM, the DCK-based ELM, and the DCK-based MS-ELM. A manually selected global threshold is usually utilized for classifying the difference images into binary maps. Logratioing is the most commonly used operation for indicating changes in SAR images. The MS-ITCD image relies on the information-theoretic features for detecting structural changes and being robust against statistical changes [8]. The DCK- and RK-based SVMs show the better identification and noise immunity of the DCK over RK. The DCK-based SVM, ELM, and MS-ELM show the higher accuracy and efficiency of ELM over SVM and the stronger noise immunity of MS-ELM over ELM. Moreover, the LIC kernel method is tested because of its similar implementation strategy but different performance (shown in Section III-D). The experimental results verify further the similarities and dissimilarities between the LIC kernel and the DCK-based MS-ELM methods. In the following, the performances of the seven methods are shown and analyzed.

First, we evaluate the change detection results subjectively. The subjective results are shown as change maps (binary maps). Black pixels on a change map form the unchanged areas, and white pixels form the changed areas. The change maps of the seven methods are shown in Fig. 9. Fig. 9(a1)–(d1) shows the change detection results obtained by classifying the log-ratio images with the globally optimal thresholds defined manually. Evidently, they are contaminated severely by noise since no spatial information is utilized for suppressing the speckle noise. Similarly, Fig. 9(a2)–(d2) shows the change detection results obtained by optimally thresholding the MS-ITCD images. The utilization of the local means as well as the global statistics of the 2-D scatterplot provides the MS-ITCD images with relatively stronger noise immunity compared with the log-ratio images. However, its performance is still constrained by the simple thresholding method. The last five methods are kernel methods. Their results of a data set [e.g., Fig. 9(a3)–(a7)] are acquired in the same situation and with the same set of training samples. If the samples were changed, the experimental results would be altered accordingly. Fig. 9(a4)–(d4) and (a5)–(d5) compares the DCK and RK functions with real SAR image change detection results. First, areas in the red ellipses show that the DCK function produces fewer false alarms, i.e., better noise immunity, compared with the RK function. It comes from the utilization of the spatial-neighborhood information in DCK. Second, areas in the green ellipses show that the DCK identifies changed areas more precisely. It comes from the better identifying ability of the kernel correlation. The results are consistent with the kernel values shown in Fig. 8. Fig. 9(a5)–(d5) and (a6)–(d6) compares the performances of the SVM and ELM. It is obvious that the ELM has equal or better performance compared with the SVM. It is caused by the excellent approximation and generalization abilities of ELM [24]. Fig. 9(a6)–(b6) and (a7)–(b7) shows the advantage of the MS-ELM over ELM. Obviously, the MS-ELM produces rather good change maps. Taking the results in yellow ellipses for example, the MS-ELM produces smoother unchanged regions compared with the ELM. It means that the MS-ELM produces fewer false alarms and it suppresses the bad effects of noise

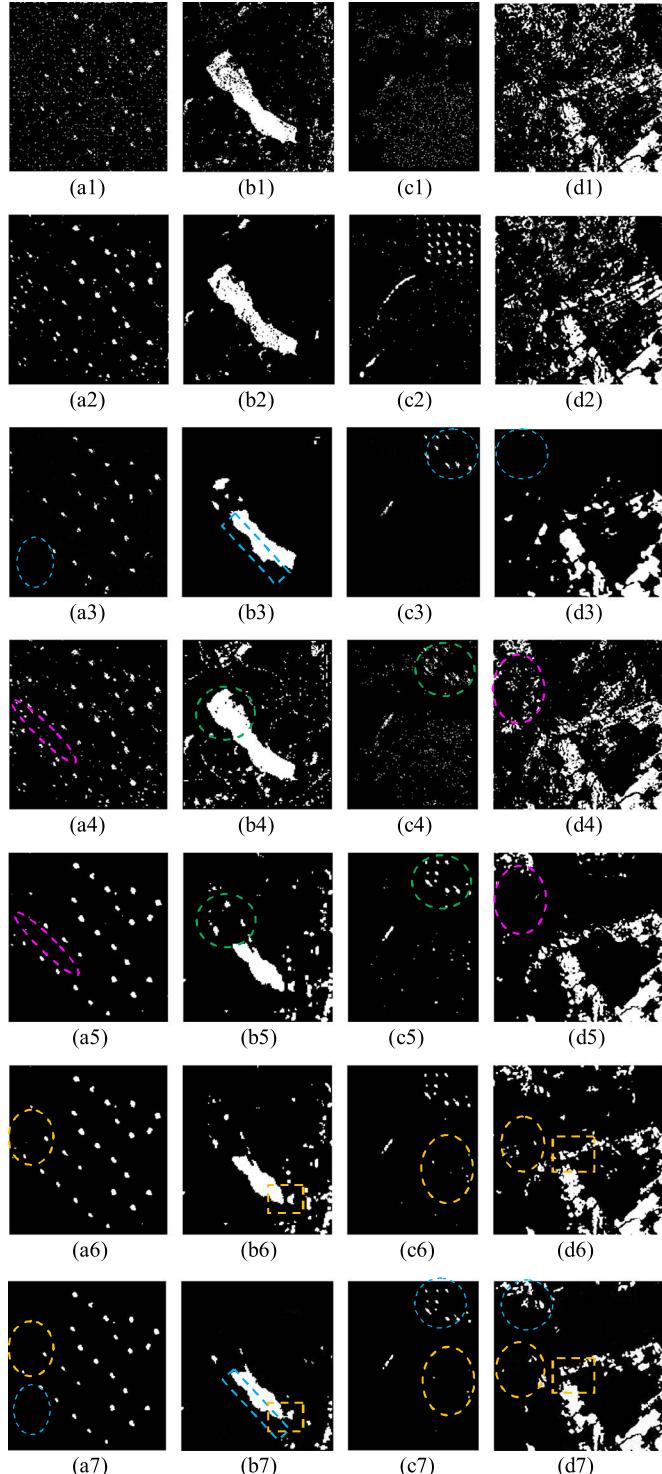


Fig. 9. Two-temporal SAR image change detection results. (a1)–(a6) “Hidden tanks” images. (b1)–(b6) “Japan city flooding” images. (c1)–(c6) “Sweden air-vehicle” images. (d1)–(d6) “Indonesia flooding” images. (a1)–(d1) Change maps resulting from thresholding the log-ratio images. (a2)–(d2) Change maps resulting from the LIC kernel method. (a3)–(d3) Change maps resulting from the RK-based SVM. (a4)–(d4) Change maps resulting from the DCK-based SVM. (a6)–(d6) Change maps resulting from the DCK-based ELM. (a7)–(d7) Change maps resulting from the DCK-based MS-ELM.

to greater extent than the ELM. In some manner, the strong noise immunity demonstrates the strong discriminating ability of the MS-ELM. The results in yellow rectangles show that the

MS-ELM keeps changed areas detected well. The advantage of the MS-ELM over ELM comes from the multistage structure in which the output-neighborhood information is utilized for improving its noise immunity, regional smoothness, and discriminating ability. Fig. 9(a3)–(d3) and (a7)–(d7) provides the experimental comparison between the LIC kernel and DCK-based MS-ELM methods. Blue rectangles show that the LIC kernel has good preservation of edge locations for large changed areas. It is because the spatial information in the LIC kernel is extracted under the supervision of the anisotropic texture analysis. The anisotropic texture analysis identifies the extents and orientations of textures in edges, and only spatial information in specific orientations is utilized in these areas. Hence, for the “Japan city flooding” data set, the LIC method produces better performance. Blue ellipses show that the DCK-based MS-ELM can detect small dotlike changes well. It is because the kernel correlation in DCK has great identification of small changes. Since both the two methods consider the spatial information, they both provide strong noise immunity. The results in Fig. 9(a3)–(d3) and (a7)–(d7) further show that the LIC kernel method is good at resisting noise and preserving edge locations whereas the DCK-based MS-ELM is good at resisting noise and identifying small dotlike changed areas, which is in accordance with the conclusions in Section III-D.

In the part, objective evaluations of the change detection results are given. The kappa coefficient ( $\kappa$ ), false alarm rate, missed alarm rate, and total error rate are given in Table III [35]. Values in Table III are the average results obtained by running the system over ten different realizations of the training sets. Obviously, the method based on the DCK and MS-ELM produces quite good results. First, it can be seen from Table III that the DCK function produces relatively lower false alarm rates, lower overall error rates, and higher kappa coefficients compared with the RK function, which means that the DCK function has stronger noise immunity. Then, it can be found that the ELM produces a few higher or almost equal precisions compared with the SVM. Moreover, larger kappa coefficients and smaller false alarm rates of the MS-ELM over ELM are illustrated, which prove further the validity of its multistage structure and the worth of incorporating the spatial-neighborhood information. At last, the higher kappa coefficients and lower missed alarm rates of DCK-based MS-ELM over the LIC kernel on “Hidden tanks” and “Sweden air-vehicle” data sets show the great identification of small changes of the DCK function. In a word, the objective results prove further the great discriminating ability, strong noise immunity, and good identification of changed areas of the DCK-based MS-ELM method for severely contaminated SAR image change detection.

The running time is shown in Table IV for illustrating the efficiency of the seven methods. The thresholding methods with the log-ratio and MS-ITCD images are fast because of their efficient classifiers. Obviously, the ELM is faster than SVM because it has fewer constraints and simpler implementation [24]. The LIC kernel and DCK-based MS-ELM methods consume much more time because of their iterative procedures. Note that the heaviest burden of the LIC kernel method comes from its complex anisotropic texture analysis. Even though the

TABLE III  
COMPARISON OF CHANGE DETECTION PERFORMANCE

Data sets	Algorithms	Total error rate	False alarm rate	Missed alarm rate	Kappa coefficient
Hidden tanks	Log-ratio	3.264%	3.204%	0.060%	0.3043
	MS-ITCD	1.962%	1.915%	0.046%	0.5689
	LIC kernel	2.161%	1.380%	0.781%	0.5221
	RK+SVM	2.360%	2.160%	0.200%	0.4923
	DCK+SVM	0.694%	0.672%	<b>0.022%</b>	0.7970
	DCK+ELM	0.611%	0.581%	0.030%	0.816
	DCK+MS-ELM	<b>0.340%</b>	<b>0.250%</b>	0.09%	<b>0.841</b>
	Log-ratio	7.146%	7.025%	<b>0.1212%</b>	0.5130
Japan city flooding	MS-ITCD	5.377%	4.900%	0.4770%	0.5668
	LIC kernel	<b>2.111%</b>	<b>1.831%</b>	0.280%	<b>0.8636</b>
	RK+SVM	8.441%	8.222%	0.219%	0.4415
	DCK+SVM	4.880%	4.140%	0.740%	0.5548
	DCK+ELM	4.461%	3.660%	0.801%	0.5739
	DCK+MS-ELM	3.801%	2.161%	1.640%	0.6297
	Log-ratio	3.414%	2.221%	1.193%	0.1323
	MS-ITCD	2.703%	2.257%	0.446%	0.2746
Sweden air-vehicle	LIC kernel	0.510%	<b>0.120%</b>	0.390%	0.6060
	RK+SVM	1.550%	1.277%	0.273%	0.3320
	DCK+SVM	1.127%	0.820%	0.307%	0.5389
	DCK+ELM	0.401%	0.221%	<b>0.180%</b>	0.6973
	DCK+MS-ELM	<b>0.390%</b>	0.181%	0.209%	<b>0.7105</b>
	Log-ratio	11.58%	10.50%	1.078%	0.5016
	MS-ITCD	10.84%	10.04%	<b>0.798%</b>	0.5304
	LIC kernel	6.164%	3.653%	2.511%	0.6825
Indonesia flooding	RK+SVM	8.203%	5.640%	2.563%	0.6456
	DCK+SVM	6.013%	5.050%	0.963%	0.6947
	DCK+ELM	6.312%	5.130%	1.182%	0.6772
	DCK+MS-ELM	<b>5.020%</b>	<b>3.170%</b>	1.850%	<b>0.7116</b>

TABLE IV  
COMPARISON OF RUNNING TIME

Running time	Log-ratio	MS-ITCD	LIC kernel	RK+SVM	DCK+SVM	DCK+ELM	DCK+MS-ELM
Hidden tanks	<b>0.447s</b>	1.124s	18.21s	2.723s	2.762s	2.195s	14.63s
Japan city flooding	<b>0.506s</b>	1.035s	27.86s	3.011s	3.220s	2.732s	19.02s
Sweden air-vehicle	<b>0.289s</b>	0.842s	11.93s	2.561s	2.667s	1.805s	6.988s
Indonesia flooding	<b>0.513s</b>	1.143s	29.96s	3.039s	3.354s	2.961s	22.03s

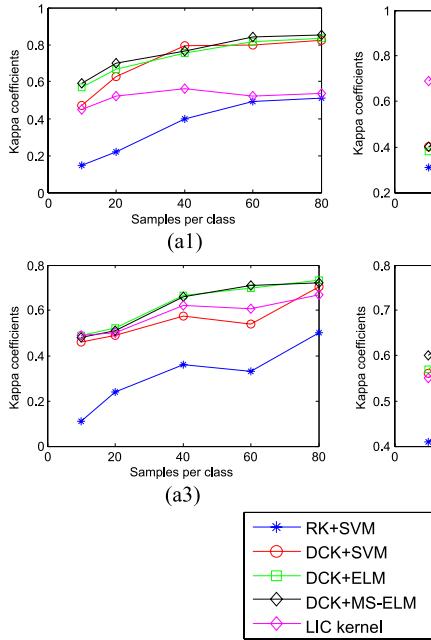


Fig. 10. Performance of the kernel methods with varied number of training samples. (a1) “Hidden tanks” images. (a2) “Japan city flooding” images. (a3) “Sweden air-vehicle” images. (a4) “Indonesia flooding” images.

multistage structure increases the computation burden of the MS-ELM, the fast training speed of ELM makes MS-ELM still acceptable.

It is well known that the number of training samples has great influence on the performance of a supervised kernel method. Therefore, the performances of the five kernel methods with varied number of training samples are given. Since, in reality, it is hard to gain much prior knowledge of the images, a few training samples are used in our experiments. Change detection performances with 10, 20, 40, 60, and 80 training samples per class are given in Fig. 10. Generally, the five kernel methods work better with the number of training samples increasing. However, it does not mean that the more training samples we use, the better the performances are, since not only the number of training samples but also their positions affect the performances. It is obtained intuitively from Fig. 10 that 60 training samples per class could produce rather good performances in our experiments. In addition, spatial-neighborhood information is utilized in the proposed and LIC kernel methods. Here, the performances of the DCK-based MS-ELM method and the LIC kernel method with varied neighborhood sizes are given in Fig. 11 ( $N = 3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9$ ). Even though the two methods are rather robust against the neighborhood size, too large neighborhoods may blur and decrease their performances.

## V. CONCLUSION

This paper has proposed a supervised kernel method for SAR image change detection. The method consists of a DCK function and an MS-ELM classifier. The DCK function is constructed first with a kernel correlation computed between coupled bitemporal image patches at all relative shifts and then with an algebraic subtraction between two pixels at different locations. The high-order spatial-neighborhood information utilized provides strong noise immunity for the DCK function,

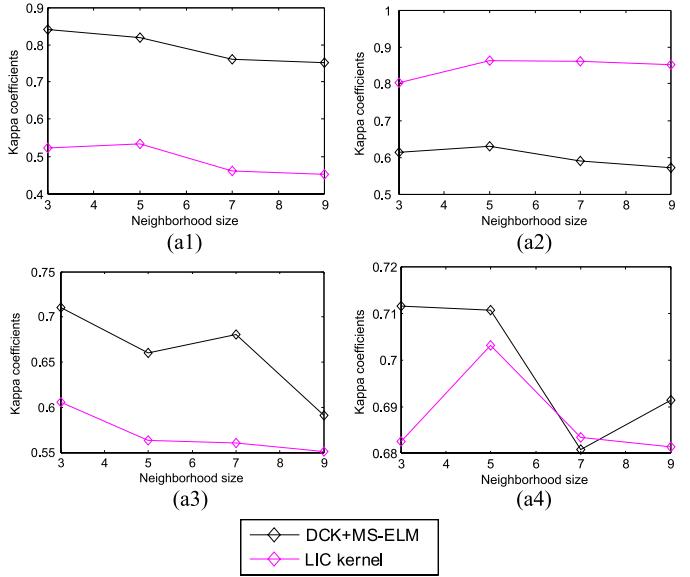


Fig. 11. Performance of the proposed kernel method with varied neighborhood sizes. (a1) “Hidden tanks” images. (a2) “Japan city flooding” images. (a3) “Sweden air-vehicle” images. (a4) “Indonesia flooding” images.

and the kernel correlation provides it with good identification of changed areas, even small dotlike changes. To further improve the change detection performance, an MS-ELM is proposed. In MS-ELM, the kernel functions that compose the hidden nodes and the weights between the hidden and output nodes are updated stage by stage with the output spatial-neighborhood information. In this way, the noise immunity and regional smoothness are improved further. The DCK function and MS-ELM are combined to get the change detection results. The effectiveness of the proposed method has been verified for real SAR image change detection, including the graphical exhibition of the DCK and RK values and the more precise change detection results of MS-ELM. The high training speed of ELM makes MS-ELM efficient. The strong noise immunity, powerful identification of changed areas, and good regional smoothness prove the potential of the DCK-based MS-ELM method for SAR image change detection.

In essence, as a kernel method, the proposed DCK-based MS-ELM has great capability in dealing with the high-dimensional multipolarization, Multiview, and multiband SAR image change detection which is becoming a hotspot nowadays. Moreover, in (13), the subtraction operation is utilized. Whether the ratio operation is adaptable is the problem that we will consider in the future.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments.

## REFERENCES

- [1] A. Singh, “Digital change detection techniques using remotely sensed data,” *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.
- [2] S. Martinis, A. Twele, and S. Voigt, “Unsupervised extraction of flood-induced backscatter changes in SAR data using Markov image modeling on irregular graphs,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 251–263, Jan. 2011.

- [3] D. C. Zanotta and V. Haertel, "Gradual land cover change detection based on multitemporal fraction images," *Pattern Recognit.*, vol. 45, pp. 2927–2937, Feb. 2012.
- [4] C. Pratola, F. D. Frate, G. Schiavon, and D. Solimini, "Toward fully automatic detection of changes in suburban areas from VHR SAR images by combining multiple neural-network models," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2055–2066, Apr. 2013.
- [5] A. Vidal and M. R. Moreno, "Change detection of isolated housing using a new hybrid approach based on object classification with optical and TerraSAR-X data," *Int. J. Remote Sens.*, vol. 32, no. 24, pp. 9621–9635, Aug. 2011.
- [6] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Norwood, MA, USA: Artech House, 1998.
- [7] M. Gong, Y. Cao, and Q. Wu, "A neighborhood-based ratio approach for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 307–311, Mar. 2012.
- [8] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and C. Zoppetti, "Non-parametric change detection in multitemporal SAR images based on mean-shift clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2022–2031, Apr. 2013.
- [9] J. Ma, M. Gong, and Z. Zhou, "Wavelet fusion on ratio images for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1122–1126, Nov. 2012.
- [10] T. Celik and K. K. Ma, "Multitemporal image change detection using undecimated discrete wavelet transform and active contours," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 706–716, Feb. 2011.
- [11] B. Hou, Q. Wei, Y. Zheng, and S. Wang, "Unsupervised change detection in SAR image based on gauss-log ratio image fusion and compressed projection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3297–3317, Aug. 2014.
- [12] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, Apr. 2005.
- [13] F. Chatelain, J. Y. Tourneret, and J. Ingla, "Change detection in multisensor SAR images using bivariate Gamma distributions," *IEEE Trans. Image Process.*, vol. 17, no. 3, pp. 249–258, Mar. 2008.
- [14] F. Wang, Y. Wu, Q. Zhang, P. Zhang, M. Li, and Y. Lu, "Unsupervised change detection on SAR images using triplet Markov field model," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 697–701, Jul. 2013.
- [15] H. Li, M. Li, P. Zhang, W. Song, L. An, and Y. Wu, "SAR image change detection based on hybrid conditional random field," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 910–914, Apr. 2015.
- [16] J. Ingla and G. Mervier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, May 2007.
- [17] X. Zhang, J. Chen, and H. Meng, "A novel SAR image change detection based on graph-cut and generalized Gaussian model," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 14–18, Jan. 2013.
- [18] J. Zheng and H. You, "A new model-independent method for change detection in multitemporal SAR images based on Radon transform and Jeffrey divergence," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 91–95, Jan. 2013.
- [19] G. Quin, B. Pinel-Puyssegur, J. M. Nicolas, and P. Loreaux, "MOMOSA: An automatic change detection method for SAR time series," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5349–5363, Sep. 2014.
- [20] K. R. Muller, S. Mika, G. Ratsh, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–202, Mar. 2001.
- [21] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. L. Rojo-Alvarez, and M. Martinez-Ramon, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, Jun. 2008.
- [22] L. Jia *et al.*, "SAR image change detection based on iterative label-information composite kernel supervised by anisotropic texture," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3960–3973, Jul. 2015.
- [23] L. Jia, M. Li, Y. Wu, P. Zhang, and H. Chen, "Semisupervised SAR image change detection using a cluster-neighborhood kernel," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 8, pp. 1443–1447, Aug. 2014.
- [24] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 4, pp. 513–529, Apr. 2012.
- [25] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [26] G. Huang, G. B. Huang, S. Song, and K. You, "Trends in extreme learning machine: A review," *Neural Netw.*, vol. 61, pp. 32–48, Jan. 2015.
- [27] G. B. Huang and L. Chen, "Enhanced random search based incremental extreme learning machine," *Neurocomputing*, vol. 71, no. 6, pp. 3460–3468, 2008.
- [28] G. B. Huang, L. Chen and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Trans. Neur. Net.*, vol. 17, no. 4, pp. 879–892, Jul. 2006.
- [29] Z. Bai, G. B. Huang, D. Wang, H. Wang, and M. Brando Westover, "Sparse extreme machine for classification," *IEEE Trans. Cybern.*, vol. 44, no. 10, pp. 1858–1870, Oct. 2014.
- [30] H. J. Rong, G. B. Huang, N. Sundararajan, and P. Saratchandran, "Online sequential fuzzy extreme learning machine for function approximation and classification problems," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 4, pp. 1067–1072, Aug. 2009.
- [31] W. W. Zong, G. B. Huang, and Y. Q. Chen, "Weighted extreme learning machine for imbalance learning," *Neurocomputing*, vol. 101, pp. 229–242, Feb. 2013.
- [32] G. B. Huang, S. Song, and C. Wu, "Semi-supervised and unsupervised extreme learning machines," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2405–2417, Mar. 2014.
- [33] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.
- [34] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," Dept. Comput. Sci., Nat. Taiwan Univ., Taipei, Taiwan, Tech. Rep., 2003.
- [35] G. Banko, "A review of assessing the accuracy of classifications of remotely sensed data and of methods including remote sensing data in forest inventory," Int. Inst. Appl. Syst. Anal., Laxenburg, Austria, Interim Rep., Nov. 1998.



**Lu Jia** was born in 1988. She received the B.S. degree in electronic engineering from the China University of Petroleum, Qingdao, China in 2011 and the M.S. degree in signal and information processing from Xidian University, Xi'an, China, in 2014, where she is currently working toward the Ph.D. degree in signal and information processing.

Her main research interest is remote sensing image change detection.



**Ming Li** (M'03) was born in 1965. He received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees in signal processing from Xidian University, Xi'an, China, in 1987, 1990, and 2007, respectively.

In 1987, he joined the Department of Electronic Engineering, Xidian University. He is currently a Professor in the National Laboratory of Radar Signal Processing, Xidian University. His research interests include adaptive signal processing, detection theory, ultrawideband, and SAR image processing.



**Peng Zhang** (M'14) received the B.S. degree in electronic and information engineering and the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 2006, 2009, and 2012, respectively.

He is currently with the National Laboratory of Radar Signal Processing, Xidian University. His main research interests are synthetic aperture radar image analysis and interpretation and statistical learning theory.



**Yan Wu** received the B.S. degree in information processing and the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1987, 1998, and 2003, respectively.

From 2003 to 2005, she was a Postdoctoral Fellow with the National Laboratory of Radar Signal Processing. Since 2005, she has been a Professor with the School of Electronic Engineering, Xidian University. She has published more than 60 technical papers. Her broad research interests are remote sensing image analysis and interpretation, data fusion of multisensor images, synthetic aperture radar autotarget recognition, and statistical learning theory and application.