A Benchmark Evaluation of Similarity Measures for Multi-temporal SAR Image Change Detection

Shiyong Cui, Gottfried Schwarz, and Mihai Datcu, *Fellow, IEEE*

Abstract—SAR image change detection is playing an important role in various Earth Observation (EO) applications. There exist a large number of different methods that have been proposed to address this issue. However, due to the fact that several kinds of changes with diverse characteristics can arise in SAR images, there is no consensus on the their performances because most methods have been evaluated using different datasets, probably facing several kinds of changes, but without an in-depth analysis of the characteristics of SAR image changes. Therefore, two important problems arise. The first is what kind of change each approach can detect. The second is how much they can detect a kind of change. Although the importance to model any kind of changes has been realized, there is no principled methodology to carry out the analysis due to the difficulty in modeling various kinds of changes. In this paper, we propose a benchmark methodology to reach this goal by simulating selected kinds of changes in addition to using real data with changes. Six kinds of SAR changes for eight typical image categories are simulated, i.e., reflectivity changes, first order, second order, and higher order statistical changes, linear and nonlinear changes. Based on this methodology for change simulation, a comprehensive evaluation of information similarity measures is carried out. An explicit conclusion we have drawn from the evaluation is that the various methods behave very differently for all kinds of changes. We hope that this study will promote the advancement of this topic.

Index Terms—Change detection, change simulation, information similarity measure, Synthetic Aperture Radar (SAR).

I. Introduction

IN Earth Observation (EO), with the availability of high resolution SAR images acquired by modern spaceborne systems, such as TerraSAR-X and TanDEM-X [1], multi-temporal SAR change detection is playing an important role in various applications such as environmental monitoring, disaster assessment, and land cover dynamics. Accordingly, many methods have been proposed and developed to address this task. We first give a brief review of related methods.

Most approaches fall into two categories: supervised [2] and unsupervised [3] approaches. In supervised change detection, reliable training samples based on prior knowledge of the area are collected and used to train a classifier, which will be used after training to classify each pixel as being changed or unchanged. On the contrary, in unsupervised change detection, the first step is to generate a change index by comparing extracted features or even the two images using some similarity metrics, and then to threshold or label the change index to produce a final binary change detection map consisting of changed and unchanged pixels. In the sense of sample selection, unsupervised approaches are more preferable to supervised ones in practice as training samples are not always available. Therefore, in this paper, we consider only (pixel-based) unsupervised SAR change detection.

Most unsupervised change detection methods comprise three steps: preprocessing such as despeckling and co-registration, image comparison to generate a change index, and change index analysis to generate a final binary change map. Preprocessing includes image registration, radiometric corrections, despeckling, etc. One widely used image comparison technique for SAR images is the log-ratio operator [4]–[6] which is particularly suited to SAR change detection due to the presence of multiplicative noise. Recently, methods based on information measures have shown promising performance for multi-temporal change detection. They assess image similarity by quantifying the dependence or distance between two random variables associated with two images. One prominent work in [7] proposed a method for multi-temporal SAR change detection based on the evolution of the local statistics, which was extended to object-based change detection in [8]. A similar method has been extended to the wavelet domain [9]. In [10], several information similarity measures including distance to independence, mutual information, cluster reward, Woods criterion, and correlation ratio, were compared for change detection, among which mutual information has been demonstrated to be rather efficient. Taking advantage of mutual information, a pixel-based approach comparing localized mutual information was proposed in [11]. Intuitively, if two pixels share a lot of information, it is reasonable to assume no change at their location. Based on this idea, another information measure for change detection derived from mutual information was introduced in [12], namely mixed information, which unifies mutual information and variational information by a parameter. Furthermore, stochastic kernels including both Kullback-Leibler divergence and mutual information were used in [13] as features in a support vector machine for SAR change detection. Based on the estimation of a bivariate Gamma distribution, mutual information was applied to SAR change detection in [14]. A region-based local mutual information change indicator was proposed by [15] to perform a change analysis of urbanization processes from multi-temporal panchromatic SPOT 5 images. Through a two-scale implementation, mutual information can be split into two terms to be linked to a change detection part and a registration part [16].

After obtaining a change index, many techniques have been proposed to analyze it in order to derive a final binary change

The authors are with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Münchener Straße 20, 82234 Weßling, Germany. E-mail: shiyong.cui@dlr.de, gottfried.schwarz@dlr.de, mihai.datcu@dlr.de.

map. The most intuitive method is to threshold the change index. In [17], a method was proposed for thresholding the change index under the assumption that both changed and unchanged pixels follow a generalized Gaussian distribution. Therefore, the distribution of the change index is a mixture of two generalized Gaussian distributions, which can be solved efficiently through the Expectation-Maximization (EM) algorithm. The same framework was used in [6] in the wavelet instead of the spatial domain. However, these methods do not consider contextual information. To overcome this drawback, a method was published in [18] to incorporate contextual information by a Markov random field model.

Although a large number of methods as reviewed above have been developed, there is no consensus about their performances because most of them were evaluated using different datasets, probably facing different kinds of changes, but without an in-depth analysis of the characteristics of the changes. Thus, it is not fair to claim one method to be superior or inferior to another method because they are not evaluated using the same kind of changes. In practice, several kinds of changes with different characteristics can arise in SAR images. A number of examples of various kinds of changes are given in Section II where we describe each kind of change. Here, two important problems arise. The first is what kind of changes each approach can detect. The second is to what extent one can detect different kinds of changes. To solve these two problems, we have to identify the different kinds of changes and develop a method to represent their characteristics. Although the importance to model different kinds of changes has been realized, there is no principled methodology to carry out the analysis due to the difficulty in modeling different kinds of changes. Another problem is the lack of standard benchmark approaches and datasets.

In this paper, we limit to the same SAR imaging geometry and propose a benchmark methodology to solve these problems by simulating six different kinds of changes. First, we try to define each kind of change based on our experience in browsing a large amount of TerraSAR-X images. Then, six kinds of SAR changes for eight typical image categories are simulated, i.e., reflectivity changes, first order, second order, and higher order statistical changes, as well as linear and nonlinear changes. Based on this methodology for change simulation, a comprehensive evaluation of information similarity measures is carried out. To the best of our knowledge, this is the first endeavor to present the problems in a systematic way and to solve them in a benchmark framework.

This rest of this paper is organized as follows. The benchmark methodology for SAR image change simulation is presented in Section II. In Section III, information similarity measures for SAR image change detection are briefly reviewed. A performance evaluation of information similarity measures using both synthetic data and real data is presented in Section IV and V. Finally, a conclusion is given in Section VI.

II. A BENCHMARK METHODOLOGY FOR SAR IMAGE CHANGE SIMULATION

In this section, we present a benchmark methodology for SAR image change simulation of six kinds of changes. This categorization of changes is mainly based on our experiences in browsing a large amount of TerraSAR-X images.

A. Simulation of Reflectivity Changes

Reflectivity changes are the main changes investigated in the literature, and happen quite often in the case of a flooding. As a rough surface becomes smoother after a flooding, its reflection is changed from diffuse reflection to specular reflection away from the observer as shown in Fig. 1(a). Thus, the pixel brightness values in SAR images are usually reduced significantly, which can be verified from the data distributions of the pixel values as shown in Fig. 1(b). Six real examples of reflectivity changes selected from a pair of calibrated TerraSAR-X images acquired before and after the big tsunami in Japan in 2011 are shown in Fig. 2.
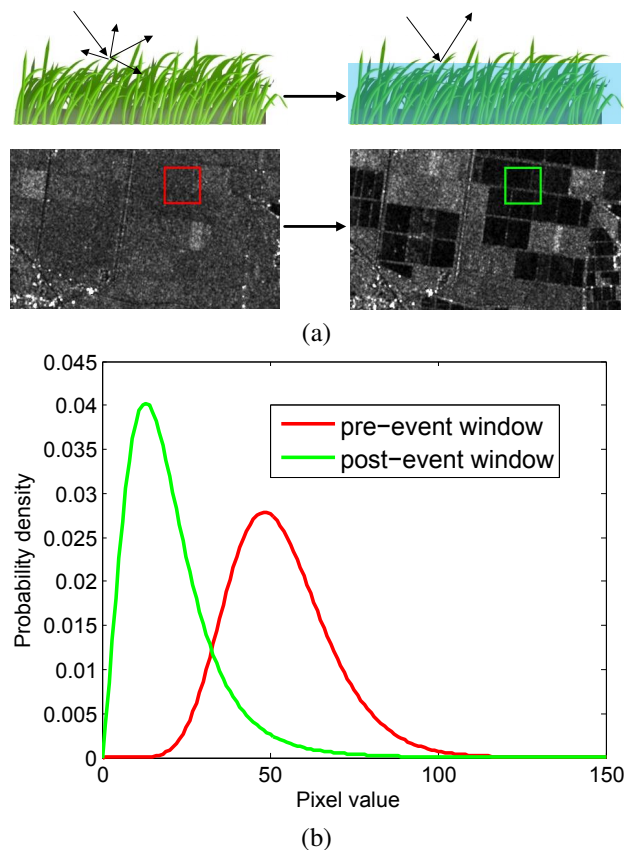

(a)

(b)

Fig. 1. An example of reflectivity changes in the case of a flooding: (a) Color image taken from http://www.clipartbest.com and TerraSAR-X images before and after the flooding; (b) data distributions of the pixels in the red and green windows before and after the flooding.

Let us consider two co-registered SAR images $X(i,j)$ and $Y(i,j)$, $1 \leq i \leq M, 1 \leq j \leq N$ acquired over the same area at different times. Suppose $x \in \{x_1, x_2, ..x_n\}$ and $y \in \{y_1, y_2, ...y_n\}$ are the pixel values in two corresponding windows w_x and w_y of image X and Y.

Definition 1: We define a set of pixels $S = \{i \mid |x_i - y_i| \gg T\}$, where T is a large value. The pixel values from the two windows falling into the set S are denoted by x_s and y_s. The cardinality of S is denoted by $|S|$. If $|S| \approx n$ and
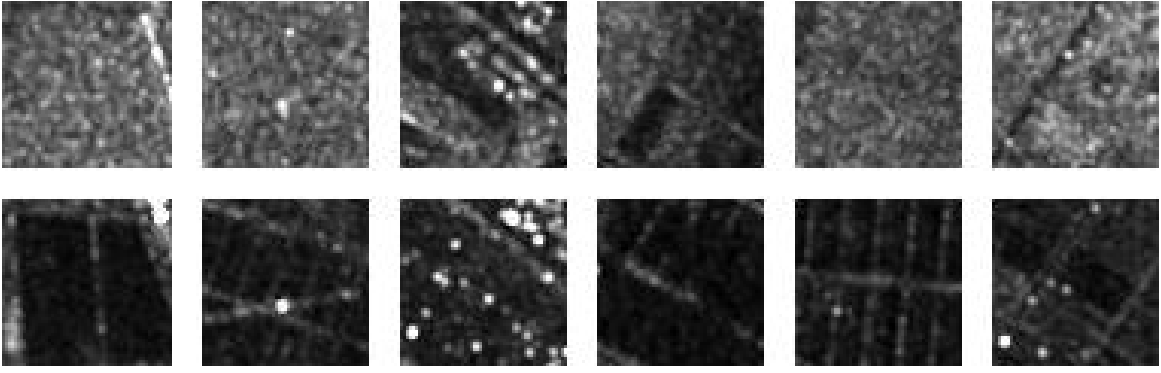
Fig. 2. Six typical examples of reflectivity changes selected from TerraSAR-X images acquired before (upper row) and after the tsunami (lower row) in Japan, 2011.

$\sigma(x_s) > \sigma(y_s)$, then we consider there is a reflectivity change associated with these two windows.

Based on this definition, most pixels are changed significantly and a post-event image becomes more homogeneous compared to its pre-event image. In order to simulate this kind of change, we first estimate the Probability Density Function (PDF) $p_X(x)$ using the pixel values of a selected window. Then, we change the parameters governing $P_X(x)$ and simulate reflectivity changes via a probability integral transform [19] based on the new Cumulative Distribution Function (CDF) $F'_X(x)$. In this paper, we choose the $\mathcal{G}^0$ model proposed in [20] to model the pixel distribution because it outperforms many other models [21]. The $\mathcal{G}^0$ model of a SAR amplitude[1] image is defined as

$$p_A(x) = \frac{2L^L \Gamma(L-\alpha)}{\gamma^\alpha \Gamma(L)\Gamma(-\alpha)} \frac{x^{2L-1}}{(\gamma + Lx^2)^{L-\alpha}}, \quad L, \gamma > 0, \alpha < 0 \tag{1}$$

where x is the SAR amplitude value, L is the number of looks, γ is a scale parameter, α is a shape parameter, and $\Gamma(.)$ is the Gamma function. The ith order raw moment is given by

$$E(x^i) = \left(\frac{\gamma}{L}\right)^i \frac{\Gamma(L+i)\Gamma(-\alpha-i)}{\Gamma(L)\Gamma(-\alpha)}. \tag{2}$$

The corresponding CDF for SAR amplitude data is given by

$$F_A(x) = F_{2L,-2\alpha}\left(-\frac{\alpha x^2}{\gamma}\right), \tag{3}$$

where F_{v_1,v_2} is the CDF of an F distribution [22] with two parameters v_1 and v_2.

We can observe from (2) that the expectation values are associated with the parameter γ. The effect of different γ parameters is demonstrated in Fig. 3. We can modify this parameter by $\gamma' = \alpha \times \gamma$ in order to simulate reflectivity changes of different degree. Thus, the degree of reflectivity change can be fully controlled by the parameter α, which is demonstrated in Fig. 4. Its influence on change detection accuracy will be evaluated in Section IV.
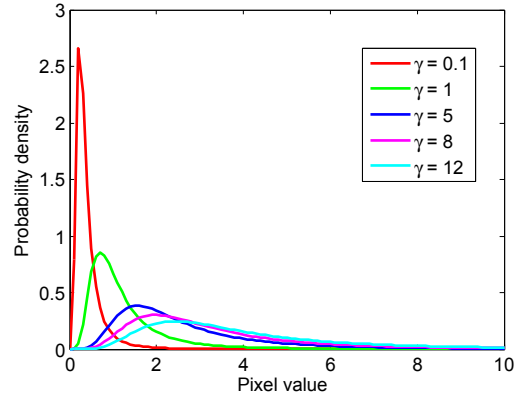
[1]The distribution of SAR intensity can be derived by the transform relation from amplitudes $p_I(x) = 2 \times \sqrt{x} \times p_A(\sqrt{x})$.

Fig. 3. $\mathcal{G}^0$ distribution with different γ parameters but the same $L = 2.9$ and $\alpha = -1.2$.

B. Simulation of Statistical Changes

Statistical changes, for instance, due to crop growing, are also quite common in SAR images, and are more difficult to detect. One prominent method in detecting statistical changes was proposed in [7] based on the evolution of local statistics which are estimated by a one-dimensional Edgeworth series expansion. The degree of evolution of the local statistics is measured by Kullback-Leibler divergence. It has shown advantages over classical SAR change detection methods, such as the mean ratio detector. To better understand the performances of this kind of statistical change detectors, we need to differentiate between changes of different orders. In this section, we present our methodology to simulate changes in first, second and higher order statistics.

1) First Order Statistical Changes: First order statistical changes are related with the differences in first order distributions. We first give a definition of first order statistical change and then give some examples.

Definition 2: We define the first order statistical distributions of the pixel values in two corresponding windows as $p_X(x)$ and $p_Y(y)$ and a similarity measure $D(.)$ defined for two PDFs. If there is a large discrepancy between these two distributions $D(p_X(x), p_Y(y)) \gg T$, we consider it as a first order statistical change associated with these two windows.

Three typical examples selected from a pair of TerraSAR-X images acquired before and after the tsunami in Sendai in 2011
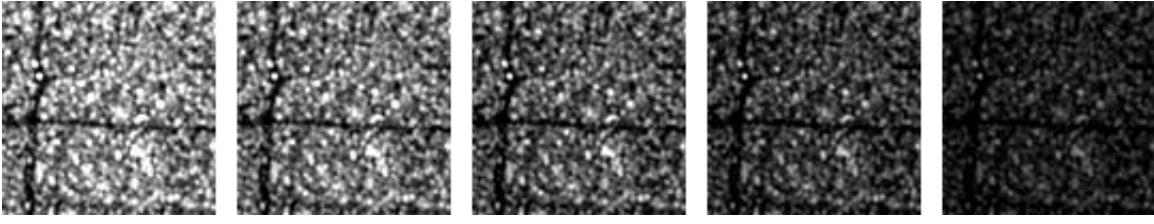
Fig. 4. The influence of parameter α on reflectivity changes. From left to right $\alpha = (1.0, 0.8, 0.6, 0.4, 0.2)$.
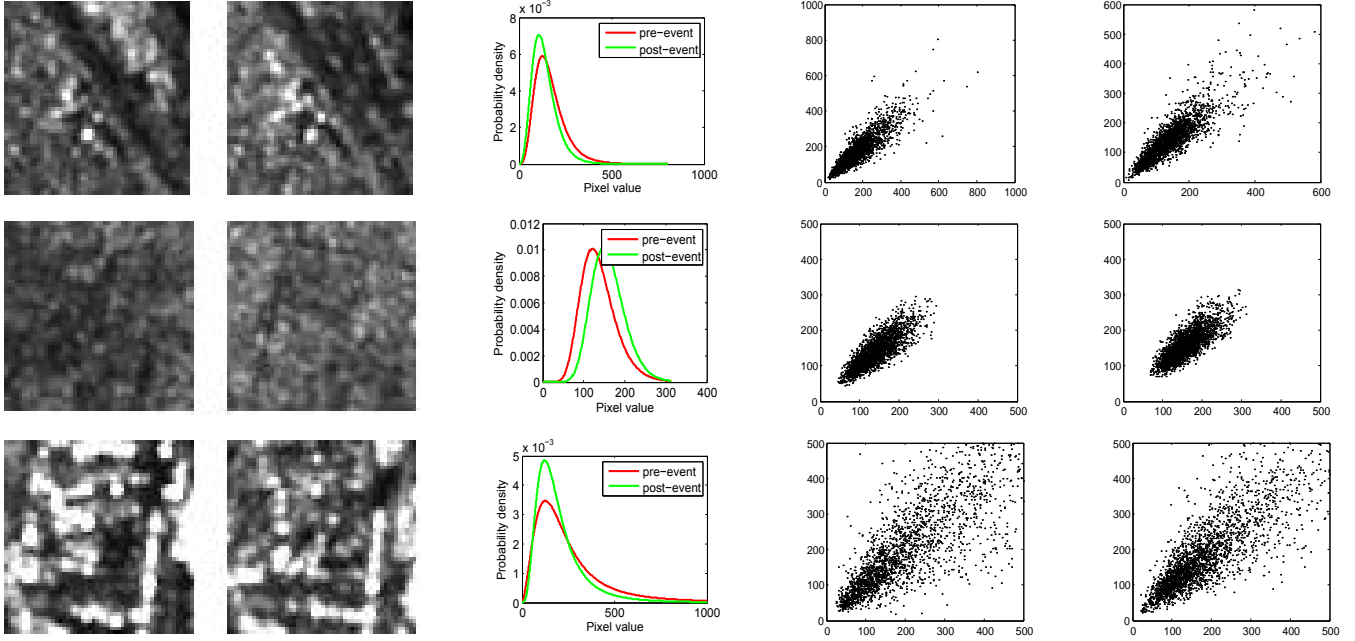

Fig. 5. Three row-wise examples of first order statistical changes. The first two images in each row are the pre- and post-event images, The histograms in the middle are the first order distributions of the two corresponding pre- and post-event images. The last two scatter plots in each row are the second order distributions of the two pre- and post-event images. From these three examples, we can observe that the first order distributions are different while the second order distributions are very similar. We consider this kind of changes associated with the first order distribution as first order statistical changes.

are shown in Fig. 5. We can see there are obvious changes in the first order distributions from the histograms (shown in the middle), while the second order distributions remain similar.

There are many different methods to modify a first order statistical distribution. In this paper, a first order statistical change is simulated by a probabilistic transform of the pixel values. Recall that if a random variable X has a cumulative distribution function (CDF) $F(x)$, the transformed random variable $U = F(X)$ will follow a uniform distribution defined on the interval $[0, 1]$. By applying the inverse CDF of any other distribution G to the uniform random variable U, we can get a desired random variable, whose distribution is exactly $G(x)$. In principle, any valid SAR model for an input SAR image and the resulting output image can be used for simulation. In our case, we assume that the input image follows a $\mathcal{G}^0$ distribution, which has been demonstrated for high resolution SAR images in [21]. We transform the pixels using the CDF of the $\mathcal{G}^0$ model given by (3) and then apply the inverse of a Gaussian CDF but keep the mean and multiply the pre-event standard deviation σ per window by α. The resulting pixels follow a Gaussian distribution with the same mean and a standard deviation of $\alpha \times \sigma$. This is demonstrated in Fig. 6. The degree of first order statistical changes can be partially
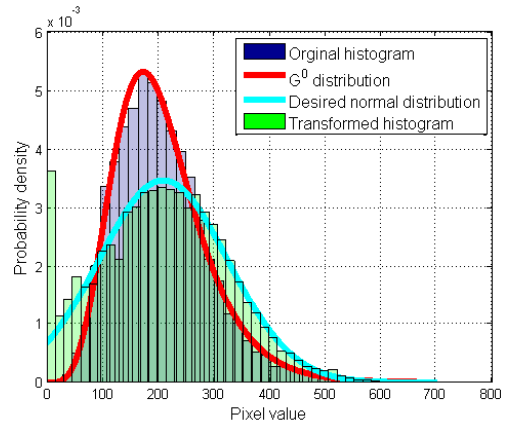

Fig. 6. Demonstration of a change simulation in first order statistics. The violet bars are the histogram of a patch before transformation, which is assumed to follow a $\mathcal{G}^0$ distribution corresponding to the red curve. After transforming the pixels to a normal distribution, the histogram of green bars is fit using a normal distribution (light blue curve). It is worth to note that the mean value is preserved while the variance is changed.

controlled by the parameter α and its effect on simulated changes is demonstrated in Fig. 7. Its influence on change
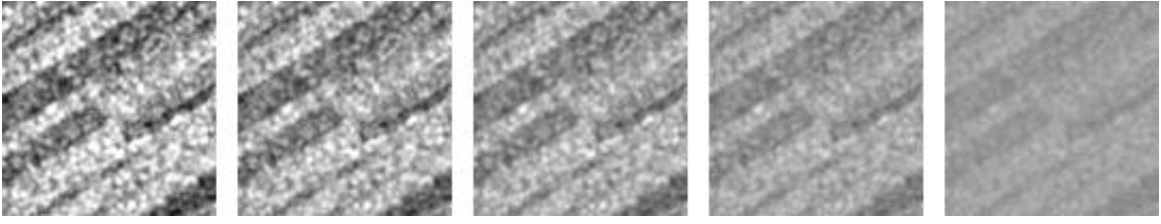
Fig. 7. The influence of parameter α on first order statistical changes. From left to right $\alpha = (1.0, 0.8, 0.6, 0.4, 0.2)$.
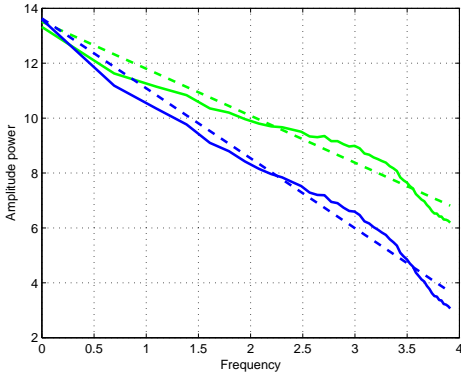

Fig. 8. Demonstration of a change simulation in second order statistics by modifying the slope of the power spectrum of an image patch. The blue dashed curve is the plot of amplitude power as a function of frequency on a log-log scale. The blue solid line is the least squares straight line fit of the blue dashed curve. Its slope is -1.7121. By multiplying the amplitude of each frequency by $f^{-\Delta\alpha}, \Delta\alpha = 0.8$, the slope of the power spectrum is changed to -2.5481, which corresponds to the green solid line. The green dashed curve is the plot of amplitude power after multiplication as a function of frequency on a log-log scale.

detection accuracy will be evaluated in Section IV.

2) Second Order Statistical Changes: Although first order statistics provide considerable information about the probability distribution of all pixels, they can not describe the structure of an image because its pixels are assumed to be independent and identically distributed (i.i.d.). Therefore, we step further to simulate changes in second order statistics, which consider the relation between pairs of pixels.

Definition 3: We define the second order statistical distributions of the pixel values in two corresponding windows as $p_{X,X'}(x,x')$ and $p_{Y,Y'}(y,y')$, where (x,x') and (y,y') are two pairs of pixel values from the two corresponding windows, and a similarity measure $D(.)$ defined for two PDFs. If there is a large discrepancy between these two distributions $D(p_{X,X'}(x,x'), p_{Y,Y'}(y,y')) \gg T$, we consider there is a second order statistical change associated with these two windows.

Three examples of second order statistical changes selected from a pair of TerraSAR-X images acquired before and after the tsunami in Sendai in 2011 are shown in Fig. 9. We can see there are obvious changes in the second order distributions from the last two scatter plots shown in each row, while the first order distributions shown in the middle are very similar.

Based on its definition, second order statistical changes indicate a change in the relation between pairs of pixels, which can be characterized by their auto-correlation. Based on the

Wiener-Khintchine theorem, stating that the auto-correlation function and the power spectrum form a Fourier transform pair, changes in second order statistics can be simulated by modifying the slope of the power spectrum. The power spectrum of an $N \times N$ image is defined by

$$S(u,v) = \frac{|F(u,v)|^2}{N^2}, \qquad (4)$$

where $F(u,v)$ is the Fourier transform of an image. By transforming the frequency $u = f\cos\theta$ and $v = f\sin\theta$ to polar coordinates and averaging over all orientations, the amplitude power as a function of frequency on a log-log scale lies approximately on a straight line, as shown by the blue solid line, which represents the well-known power law $P = 1/f^\alpha$ in image analysis [23]. The slope of this line α, -1.7121 in this case, is a second order statistic that can be modified for simulation. Any other desired slope $\alpha + \Delta\alpha$ can be obtained by multiplying the amplitude power of each frequency by $f^{-\Delta\alpha}$. In other words, second order statistical changes can be simulated through a digital, finite impulse response filter with the following frequency response

$$H(u,v) = \begin{cases} 1 & \text{if } u = v = 0 \\ \frac{1}{\left(\sqrt{u^2+v^2}\right)^{\Delta\alpha}} & \text{otherwise} \end{cases} \qquad (5)$$

To see the influence of $\Delta\alpha$, we use a series of different values $\Delta\alpha = (-1.2, -0.8, -0.2, 0.2, 0.8, 1.2)$ for simulation. The simulated results are shown in Fig. 10. It seems that the changes in slope correspond to band pass filtering. With increasing $\Delta\alpha$, the images become smoother. The influence on change detection accuracy will be evaluated in Section IV.

3) Higher Order Statistical Changes: We consider further the ability of change detection methods in detecting higher order changes. Higher order statistical changes can be considered as texture changes as well. For this kind of changes, the pre- and post-event images have similar appearances but different higher order statistics.

Definition 4: We define the higher order statistical distributions of the pixel values in two corresponding windows as $p_{\mathbf{X}}(\mathbf{x})$ and $p_{\mathbf{Y}}(\mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ are two corresponding groups of pixel values from the two windows, and a similarity measure $D(.)$ defined for two PDFs. If there is a large discrepancy between these two distributions $D(p_{\mathbf{X}}(\mathbf{x}), p_{\mathbf{Y}}(\mathbf{y})) \gg T$, we consider there is a higher order statistical change associated with these two windows.

Higher order statistical changes can be simulated through texture synthesis. In the literature, Markov random fields [24] have been successfully used for texture synthesis. They
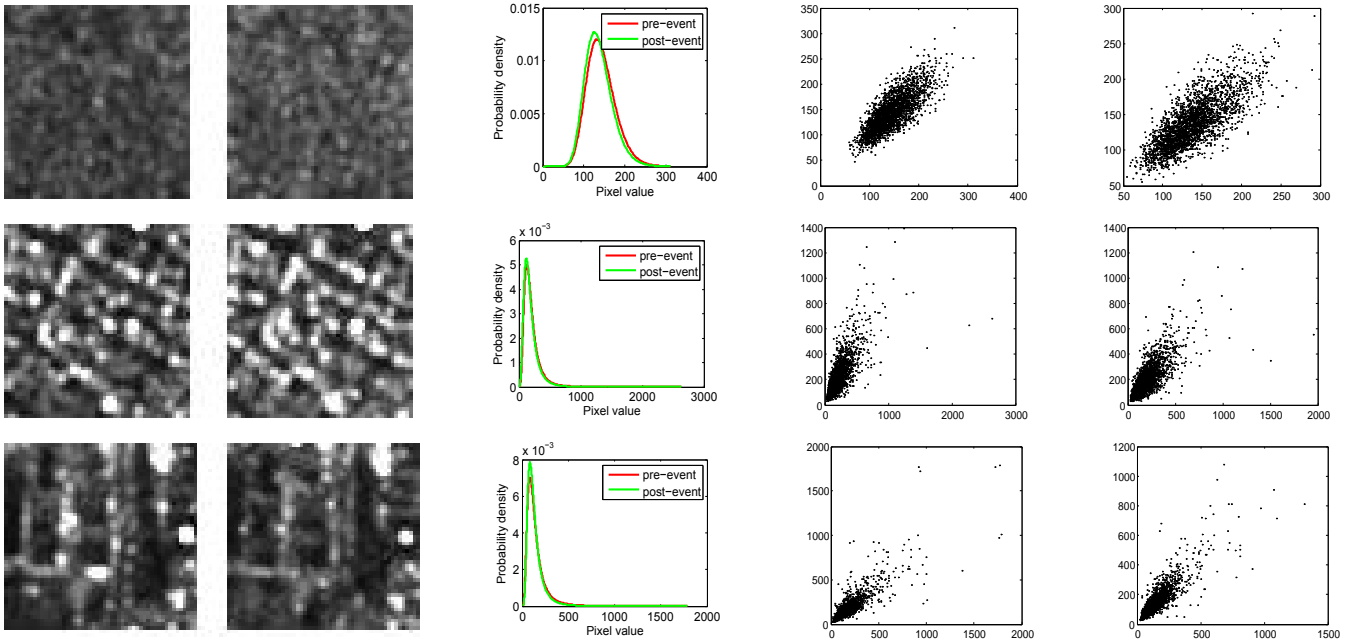
Fig. 9. Three row-wise examples of second order statistical changes and their first order and second order data distributions. The first two images in each row are pre- and post-event images. The histogram in the middle are the first order distributions of the two pre- and post-event images. The last two scatter plots in each row are the second order distributions of the two images. From these three examples, we can observe that the second order distributions are different while the first order distributions are very similar. Thus, we consider this kind of changes in second order distribution as second order statistical changes.
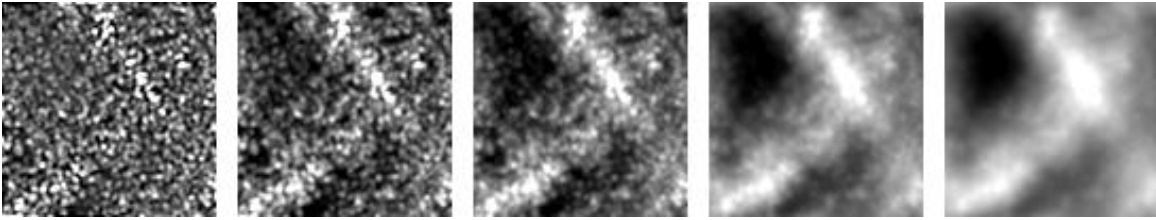

Fig. 10. Effect of $\Delta\alpha = (-0.8, -0.2, 0.2, 0.8, 1.2)$ (from left to right) in second order change simulation. With a negative $\Delta\alpha$ value, the high frequency content will be emphasized. On the contrary, with a positive $\Delta\alpha$ value, the image will be smoothed.

generate similar texture patterns using parameters learned from input images. However, due to their computational complexity, we choose another patch-based approach [25] for texture simulation that is nonparametric and has low computational complexity.

The main idea of this algorithm is to stitch together texture patches with a size of $s \times s$ pixels selected from an input image and allow overlap between two neighboring patches which are selected based on their similarity in the overlapping part. The novelty of this algorithm is that it allows neighboring patches to match smoothly at the border. It achieves texture synthesis in two iterative steps until the entire image to be synthesized is covered. The first step is to choose randomly a small patch from the input image to initialize the simulation and then to select from all available patches another patch with most similarity to the previous patch within the overlapping part. The second step is to calculate the error surface and to determine the minimum error cut between the overlapping patches such that they match smoothly at the boundary. A minimum error cut can be solved by dynamic programming algorithms, like Dijkstra's algorithm. This is demonstrated in Fig. 11. The image after simulation has a similar texture as the

input image, the images are difficult to discriminate visually but they are indeed different in higher order statistics, as shown in Fig. 12.

C. Simulation of Linear and Nonlinear Changes

In addition to statistical changes, we consider another two categories of linear and nonlinear changes. Some examples of linear and nonlinear changes are shown in Fig. 13 and 14. We can see from the data distributions shown in these two figures that they have obvious linear and nonlinear relations. First, we give a definition of these two kinds of changes. Suppose $x = \{x_i, x_2, ..., x_n\}$ and $y = \{y_i, y_2, ..., y_n\}$ are the pixel values in two corresponding windows of image X and Y.

Definition 5: For $\forall x$, if the relation $y \approx f(x) = ax + b$ holds, we consider there is a linear change between these two patches.

Definition 6: Similarly, for $\forall x$, if the relation $y \approx g(x)$ holds with $g(x)$ being a nonlinear function, we consider there is a nonlinear change between these two patches.

Changes due to linear and nonlinear transformations can be easily simulated through linear and nonlinear functions to
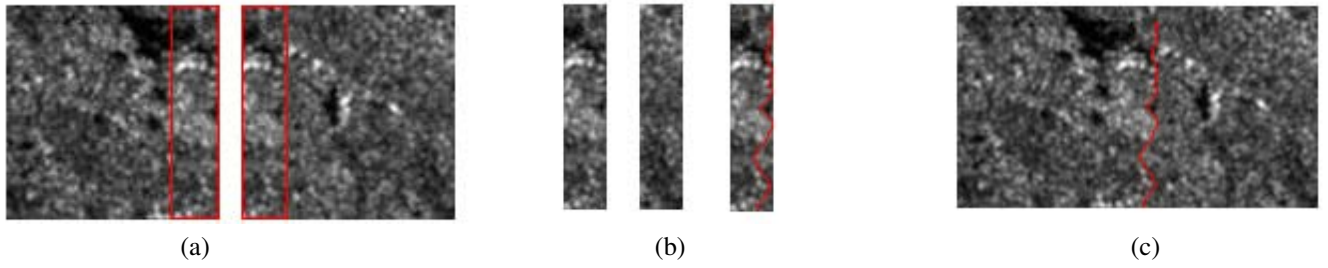
(a)   (b)   (c)

Fig. 11. Texture synthesis by image quilting for texture change simulation. Two similar patches (a) from the intersection part are selected. In this case, the patch size is 100×100 pixels and the overlap size is 20×20 pixels. If we stitch together the two patches without further processing, we would get a straight line at the border. To reduce this effect, the error surface is computed within the overlapping part, which is the squared difference between two corresponding pixels. Dynamic programming is applied to find the minimum error bound (b), where the two patches match best (c).
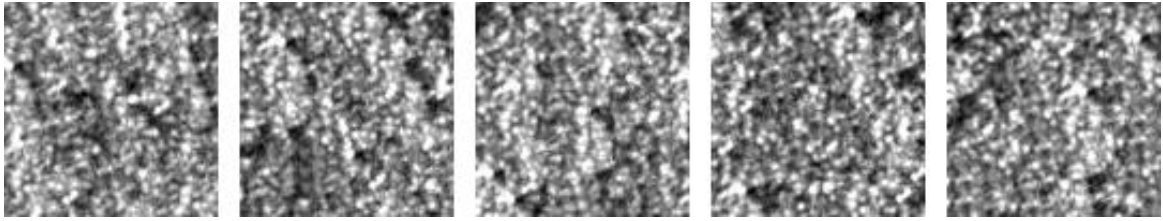

Fig. 12. The influence of parameter s on higher order statistical changes. From left to right $s = (10, 12, 14, 16, 18)$.
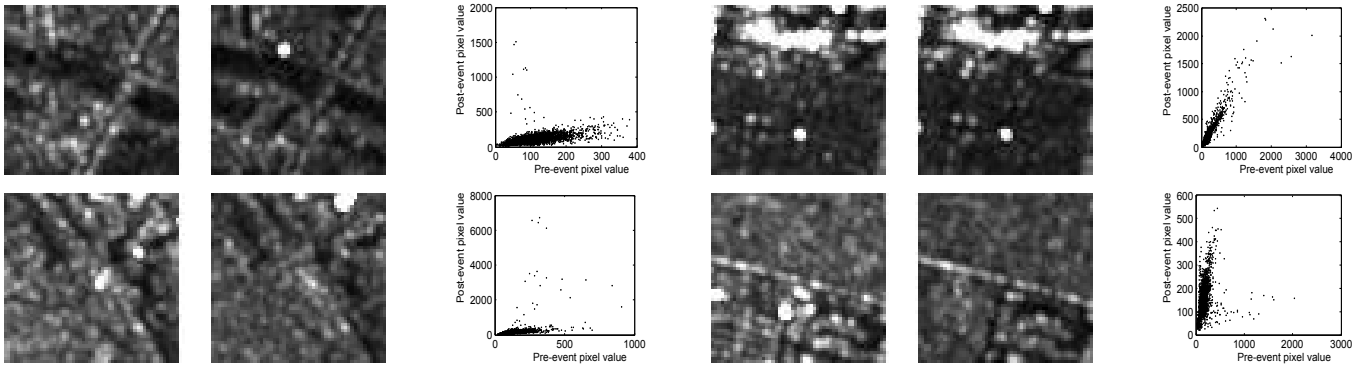

Fig. 13. Four examples of linear changes and their joint data scatter plots. It can be seen that there is a linear relation between pairs of corresponding patches.
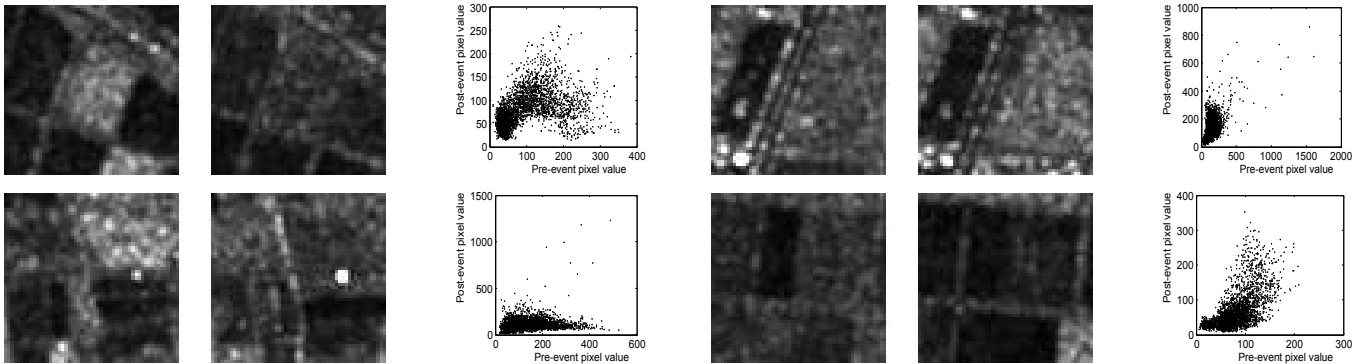

Fig. 14. Four examples of nonlinear changes and their joint data scatter plots. It can be seen that there is a nonlinear relation between pairs of corresponding patches.

transform the pixel values. For linear transforms, the simulation can be performed by a linear function $y = a(x - x_{min}) + x_{min}$, where x_{min} is the minimum value of the input patch for simulation. The parameter a controls the degree of linear change. Its effect is demonstrated in Fig. 15.

In contrast, there are many different nonlinear functions that can be applied. In this paper, we select a quadratic function $y = ax^2 + bx + c$ to achieve a nonlinear transform for nonlinear
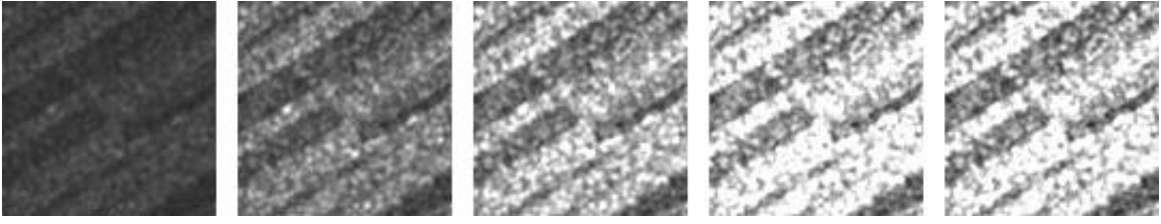
Fig. 15.   The influence of parameter α on linear changes. From left to right $a = (0.2, 0.6, 1.0, 1.4, 1.8)$. The image in the middle is the original image.

change simulation.

III. INFORMATION SIMILARITY MEASURES FOR CHANGE DETECTION

In this section, information similarity measures are briefly reviewed in the context of multi-temporal SAR image change detection. In the following, we define $p_{X,Y}(x, y), p_X(x)$ and $p_Y(y)$ as the joint and marginal probability density functions of the pixel values of two corresponding sliding windows in two images.

A. Information Similarity Measures

1) Mutual Information: Mutual information [26], as a distance measure to independence, is defined as

$$I(X,Y) = \iint p_{X,Y}(x, y) \log \frac{p_{X,Y}(x, y)}{p_X(x)p_Y(y)} dx dy. \quad (6)$$

Compared with correlation, mutual information is a general measure to quantify statistical dependence. Mutual information is always positive and is zero if and only if the two random variables X and Y are independent $p_{X,Y}(x, y) = p_X(x)p_Y(y)$. On the contrary, if they are similar, their mutual information is higher.

2) Variational Information: Variational information, defined by (7), was proposed to compare clusters [27] by measuring the amount of information lost and gained in changing from one cluster to another

$$V(X,Y) = \iint p_{X,Y}(x, y) \log \frac{p_{X,Y}(x, y)^2}{p_X(x)p_Y(y)} dx dy. \quad (7)$$

While mutual information quantifies the common information between X and Y, variational information quantifies the different information transmitted through X and Y. The relationship between these information measures is shown by the Venn diagram in Fig. 16, where $H(X)$ is the entropy and $H(Y|X)$ is the conditional entropy. In contrast to mutual information, a large value of variational information denotes a high degree of change.

3) Mixed Information: A new information measure, namely mixed information, was proposed in [12] which unifies mutual and variational information by a parameter α to trade off between mutual information and variational information. It is defined as

$$I_\alpha(X,Y) = \iint p_{X,Y}(x, y) \log \frac{p_{X,Y}(x, y)^{1+\alpha}}{p_X(x)p_Y(y)} dx dy. \quad (8)$$

As α may vary between 0 and 1, I_α can be viewed as a mixture of mutual and variational information measures. In particular, when $\alpha = I(X,Y)/H(X,Y)$, the mixed information becomes zero.
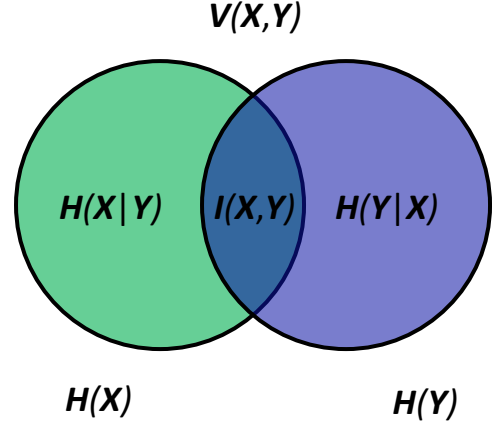

Fig. 16.   Venn diagram of information measures

4) Kullback-Leibler Divergence: As an effective distance measure between two probability density functions, Kullback-Leibler divergence [28] is defined as

$$KL(X|Y) = \int p_X(x) \log \left(\frac{p_X(x)}{p_Y(x)} \right) dx. \quad (9)$$

If the two probability density functions $p_X(x)$ and $p_Y(x)$ are close to each other, the Kullback-Leibler divergence is small. In contrast, it is larger if there is a great deviation between the two probability density functions.

B. SAR Image Change Detection Based on Information Similarity Measures

In order to apply these similarity measures to change detection, we first need a method to estimate the involved marginal and joint probability density functions. Then, we need to use these probability density functions to compute the value of a similarity measure. In general, change detection based on information similarity measures is performed locally in a sliding window by assessing the similarity between two windows. However, one critical issue is that there should exist an analytical expression of the information similarity measure, otherwise, certain approximations [29] or a Monte Carlo simulation [30] should be applied for approximation, like the work in [7], where an Edgeworth series was used to approximate the divergence. A Monte Carlo simulation relies on sampling the estimated distribution to obtain a large amount of samples, which are then used to approximate a similarity measure. Although it can approximate an information measure precisely as long as the number of samples is sufficient, it is prohibitively slow to apply to change detection especially with

a sliding window. Unfortunately, there are almost not closed-form expressions for all the information similarity measures presented in Section III-A for most distributions. Therefore, we resort to a local average of the information measures as used in [31], except for Kullback-Leibler divergence in our case. Since Kullback-Leibler divergence involves only the first order distributions, we separate it from the other three similarity measures and present it separately in the following.

1) Estimation of Local Similarity Measures: For density estimation, we use Kernel Density Estimation (KDE) for mutual information, variational information and mixed information. Suppose $x = \{x_1, x_2, ...x_n\}$ and $y = \{y_1, y_2, ...y_n\}$ are the pixel values in two corresponding windows w_x and w_y of image X and Y. The marginal and the joint probability density functions $p_X(x), p_Y(y)$ and $p_{X,Y}(x, y)$ of the pixels in the sliding window from the two images are estimated using kernel density estimation. The kernel density estimator of the unknown distribution is defined as

$$\hat{p}_X(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - x_i}{h}), \tag{10}$$

where h is a smoothing parameter (also called bandwidth), and $K(.)$ is the kernel function which should be a valid probability density function. As can be seen, the estimator $\hat{p}_X(x)$ is a weighted mixture of a series of kernel functions located at each data point. Because the smoothing parameter plays an important role in the estimation quality, the method presented in [32] for computing the best bandwidth is used. The kernel function we used is a Gaussian kernel

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}). \tag{11}$$

With the above density estimates, the local mutual information, variational information and mixed information as change indicators can be estimated (as an average of point-wise local information measures over the pixels in the local window) as follows

$$I(x, y) = \sum_{i=1}^{n} \log \frac{p_{X,Y}(x_i, y_i)}{p_X(x_i)p_Y(y_i)}, \tag{12}$$

$$V(x, y) = -\sum_{i=1}^{n} \log \frac{p_{X,Y}(x_i, y_i)^2}{p_X(x_i)p_Y(y_i)}, \tag{13}$$

$$I_\alpha(x, y) = \sum_{i=1}^{n} \log \frac{p_{X,Y}(x_i, y_i)^{1+\alpha}}{p_X(x_i)p_Y(y_i)}. \tag{14}$$

As mutual information measures the common information, high mutual information indicates a high probability of change. The opposite holds for the variational information. Mixed information is a combination of mutual information and variational information through the parameter α. As proposed by Gueguen et al. in [31], the α parameter in the mixed information is selected by a brute force search such that the accuracy is optimized.

2) Estimation of Kullback-Leibler Divergence: In contrast to mutual information, variational information, and mixed information, there is an analytical formula of Kullback-Leibler divergence for most parametric SAR image models. Thus, we choose three parametric distributions, i.e., the Gamma distribution, the one-sided generalized Gamma GΓD, and the $\mathcal{G}^0$ distribution. The probability density function of the $\mathcal{G}^0$ distribution is given in (1). The probability density functions of the Gamma distribution and GΓD and their parameters are given as follows

$$p_{\Gamma D}(x) = \frac{2}{\Gamma(L)}(\frac{L}{\mu})^L x^{2L-1} \exp\left(-\frac{Lx^2}{\mu}\right), \quad L, \mu > 0,$$

$$p_{G\Gamma D}(x) = \frac{\beta x^{\beta\lambda-1}}{\alpha^{\beta\lambda}\Gamma(\lambda)} \exp\left(-(\frac{x}{\alpha})^\beta\right), \quad \alpha, \beta, \lambda > 0, \tag{15}$$

The parameters μ and L in $p_{\Gamma D}(x)$ are the scale parameter and the number of looks. The parameters β, λ and α in $p_{G\Gamma D}(x)$ are power, shape and scale parameters. The method we used for estimating the parameters of these three distributions is the method of log-cumulants (MoLC) [33]. The detailed equations for parameter estimation can be found in [21].

The closed-form expressions of Kullback-Leibler divergence for the parametric Gamma, the generalized Gamma, and the $\mathcal{G}^0$ distribution are given by (16), (17), and (18) respectively, where $p_1(x)$ and $p_2(x)$ are the corresponding parametric probability density functions estimated for the two windows. The functions $\Psi(x), \Gamma(x)$ and $B(x, y)$ are the digamma function, the gamma function and the beta function. Due to its asymmetry, the symmetric version of Kullback-Leibler divergence $KLD(X, Y) = KL(X|Y)/2 + KL(Y|X)/2$ is used for assessing similarity. They are abbreviated as KLD_ΓD, KLD_GΓD, and KLD_$\mathcal{G}^0$D. The integral in the last term of KLD_$\mathcal{G}^0$D can be computed using Gauss-Kronrod quadrature [34].

IV. EVALUATION ON SYNTHETIC DATA

A. Simulation and Evaluation Setup

To evaluate and assess comprehensively the performances of information similarity measures for change detection, synthetic data are generated by simulating changes using TerraSAR-X images selected from eight categories, i.e., agriculture, grass land, strongly reflecting buildings, dense buildings, crops, forest, industrial area, and mountains. These images are sub-scenes of Geocoded Ellipsoid Corrected (GEC) products acquired in Stripmap mode. Their ground range resolution and azimuth resolution are 1.9406 m and 3.2999 m and their number of looks is 1.0. The incidence angle is 35.20º (ascending orbit). The patches of the eight typical categories shown in the first column in Fig. 17 are selected and used for simulation.

The assumed distribution in simulating reflectivity and first order statistical changes is a $\mathcal{G}^0$ distribution. Its parameters are estimated using the Method-of-Log-Cumulants (MoLC). The equations for parameter estimation can be found in [21]. The quadratic function for nonlinear simulation is $a(x - $

$$KLD_{\Gamma D}\Big(p_1(x; L_1, \mu_1) || p_2(x; L_2, \mu_2)\Big) = - L_1 + L_2 \ln \frac{\mu_2 L_1}{\mu_1 L_2} + \ln \frac{\Gamma(L_2)}{\Gamma(L_1)} + \frac{\mu_1 L_2}{\mu_2} + (L_1 - L_2) \Psi(L_1) \qquad (16)$$

$$KLD_{\mathrm{G}\Gamma D}\Big(p_1(x; \alpha_1, \beta_1, \lambda_1) || p_2(x; \alpha_2, \beta_2, \lambda_2)\Big) = - \lambda_1 - \lambda_2 + (\beta_1\lambda_1 - \beta_2\lambda_2)\Big(\log \frac{\alpha_1}{\alpha_2} + \frac{\Psi(0, \lambda_1)}{\beta_1} - \frac{\Psi(0, \lambda_2)}{\beta_2}\Big)$$
$$+ \Big(\frac{\alpha_1}{\alpha_2}\Big)^{\beta_2} \frac{\Gamma(\lambda_1 + \beta_2/\beta_1)}{\Gamma\lambda_1} + \Big(\frac{\alpha_2}{\alpha_1}\Big)^{\beta_1} \frac{\Gamma(\lambda_2 + \beta_1/\beta_2)}{\Gamma\lambda_2} \qquad (17)$$

$$KL_{\mathcal{G}^0}(p_1(x; L_1, \alpha_1, \gamma_1) || p_2(x; L_2, \alpha_2, \gamma_2)) = \ln \frac{B(L_2, -\alpha_2)}{B(L_1, -\alpha_1)} + L_1 \ln \frac{L_1}{\gamma_1} + L_2 \ln \frac{L_2}{\gamma_2} + (L_1 - L_2)[\ln \frac{\gamma_1}{L_1} + \Psi(L_1) + \Psi(-\alpha_1)]$$
$$- (L_1 - \alpha_1)[\Psi(L_1 - \alpha_1) - \Psi(-\alpha_1)] + \frac{L_2 - \alpha_2}{B(L_1, -\alpha_1)} \int_0^\infty \frac{x^{L_1 - 1}}{(1 + x)^{L_1 - \alpha_1}} \ln(1 + \frac{L_2 \gamma_1}{L_1 \gamma_2} x) dx$$
$$(18)$$

$x_{mean})^2/(x_{min} - x_{max}) + x_{max}$, where x_{min}, x_{max}, and x_{mean} are the minimum, maximum and mean value of each patch. To test the robustness of the information similarity measures to noise, Gaussian white noise ($\sigma = 4$) is added to the synthesized images for all six simulated cases. The parameters used for all simulations are summarized in Table I. As density estimation depends heavily on the sample size, we choose different window sizes (9×9, 11×11, 13×13, 15×15, 17×17, 19×19, 21×21, and 23×23 pixels) for evaluation. The simulated images are shown in Fig. 17.

To evaluate the accuracy of a change map independent of any thresholding algorithm, the receiver operating characteristic (ROC) curve is used and the area under the ROC curve (AUC) is computed as a performance measure. A ROC curve can be considered as the evolution of the true positive rate (TPR) as a function of the false alarm rate (FAR). TPR is defined as the fraction of correctly detected changes and FAR is the fraction of correctly detected no changes. The area under the ROC curve (AUC) is a good performance measure. The larger the area under ROC curve, the better the performance. As AUC is an overall performance measure, thresholding or labeling of the change index should be also applied to generate a binary change detection map such that TPR and FAR can also be used for evaluation and comparison. Although many thresholding methods have been proposed, none of them is perfect. Therefore, an optimal threshold corresponding to the nearest point to $(0.0, 1.0)$ on the ROC curve is selected, which gives the best performance, as shown in Fig. 18.

As our baseline for evaluation and comparison, the method based on an Edgeworth series approximation in [7] is also included in the following experiments. All the information measures, i.e., Kullback-Leibler divergence estimated respectively by Edgeworth series (KLD_EW), Gamma distribution (KLD_ΓD), GΓD (KLD_GΓD), and $\mathcal{G}^0$ distribution (KLD_$\mathcal{G}^0$D) as well as mutual information (Mi_Info), variational information (Vi_Info), and mixed information (Mix_Info) estimated by KDE, are applied to generate change indices. In the following, we first present an overall evaluation of similarity measures by computing the average accuracy over
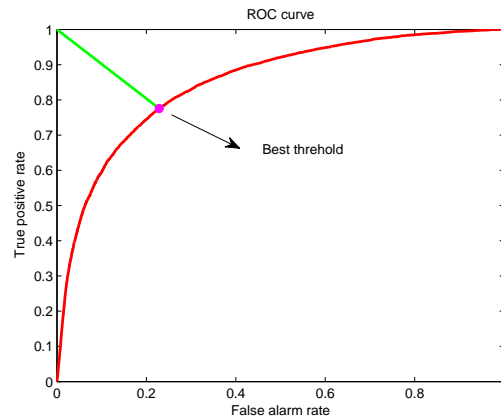

Fig. 18. ROC curve of mixed information performed on first order statistical changes in agriculture using a 9×9 window. We select the optimal threshold corresponding to the closest point to $(0.0, 1.0)$.

all classes. Then, we evaluate the influence of the parameters shown in Table I on change detection accuracy.

B. Results on Synthetic Data

All the information similarity measures are applied to each kind of simulated changes. The accuracy in terms of AUC is summarized in Fig. 19. The detailed results are presented in the supplementary material. In general, all the similarity measures do not have much difference for different window sizes. For reflectivity changes as shown Fig. 19(a), we can clearly see that both Kullback-Leibler divergence and mixed information perform quite well. On the contrary, variational information ranks far behind and mutual information comes last. The only difference among mutual information, variational information and mixed information is the parameter α. The impact of α on the accuracy for reflectivity changes in the first class (agriculture) is plotted in Fig. 20(a). We can see the reason why mixed information performs better than mutual information and variational information. This indicates a necessity to search for the best α parameter in the definition of mixed information when we apply it to SAR change detection.
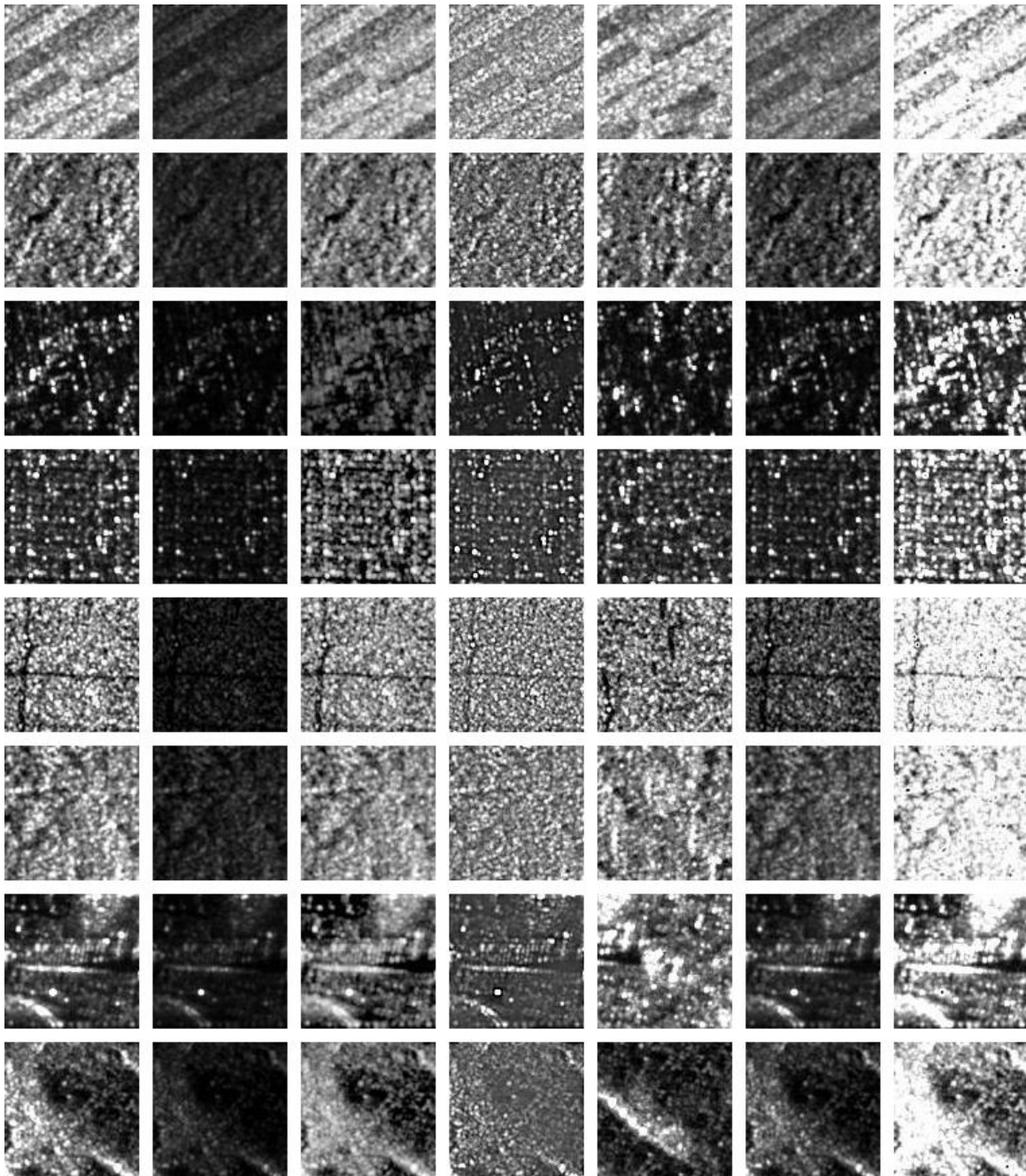
Fig. 17.   Examples of change simulation for the eight selected classes. Each row is a class and from top to bottom contains agriculture, grass, bright buildings, dense buildings, crops, forest, industrial areas, and mountains. The first column is the patch for simulation which has been selected from a large image. The following columns contain the simulated changes, which are reflectivity changes, first order statistical changes, second order statistical changes, higher order statistical changes, linear and nonlinear changes.

TABLE I
VALUES OF PARAMETERS USED FOR CHANGE SIMULATION.

Change index	Parameter	Values
Reflectivity change	α	$0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$
First order statistical change	α	$0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$
Second order statistical change	$\Delta\alpha$	$-1.2, -0.8, -0.2, 0.0, 0.2, 0.8, 1.2$
Higher order statistical change	S	$8, 10, 12, 14, 16, 18, 20$
Linear change	a	$0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8$
Nonlinear change	a	$2.0, 4.0, 6.0, 8.0, 10.0$

For first order statistical changes shown in Fig. 19(b), we can observe an obvious diversity in their performances. Similar to reflectivity changes, both variational and mutual information have an accuracy of less than 75%. However,
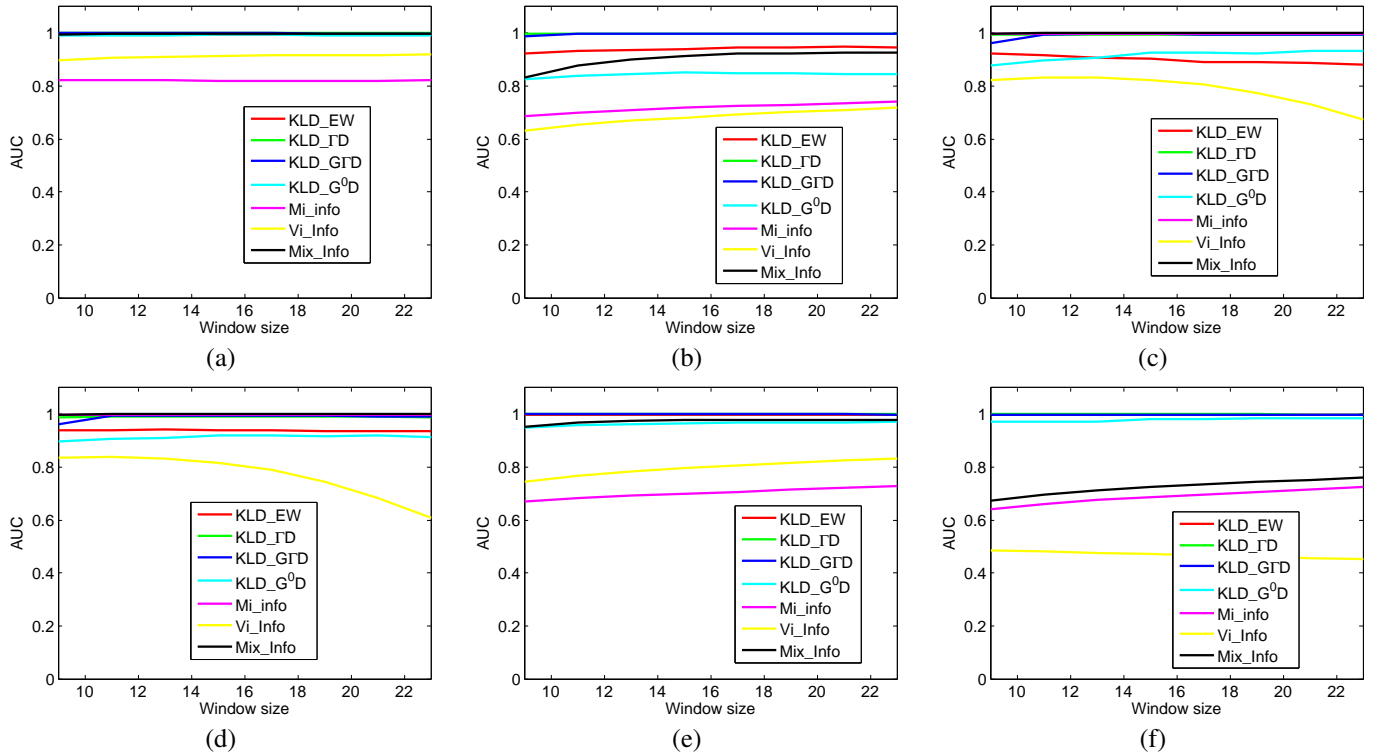
Fig. 19. Accuracy comparison of similarity measures for six kinds of simulated changes: (a) reflectivity changes; (b) first order statistical changes; (c) second order statistical changes; (d) higher order statistical changes; (e) linear changes; (f) nonlinear changes.
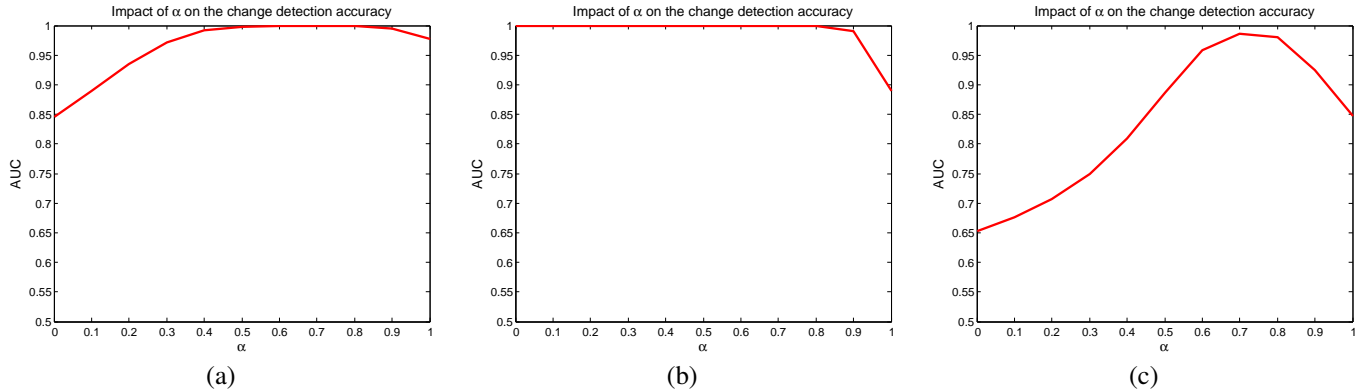

Fig. 20. Impact of α on the performance of mixed information measure for intensity changes, second order statistical changes, and linear changes in the case of agriculture.

mixed information shows a much better accuracy (of around 92.5%) than mutual information and variational information. It is surprising that KLD_ΓD and KLD_GΓD rank on top and they have a similar accuracy of more than 99%. The potential reason is that the images selected for simulation have fully developed speckle. KL_EW has about 6% lower accuracy than KLD_ΓD and KLD_GΓD. The underlying reason is the Edgeworth series expansion because it assumes that the underlying distribution is close to a Gaussian distribution. Otherwise, the fitting accuracy will be lower. However, for SAR images, the underlying distribution is rarely a Gaussian. KLD_$\mathcal{G}^0$D has a lower accuracy than KL_EW. The reason can be the numerical integral in (18).

When we move to second order statistical changes as shown

in Fig. 19(c), mutual information and mixed information perform much better than those for first order statistical changes. The reason can be seen in Fig. 20(b). As before, KLD_ΓD and KLD_GΓD have a very good performance which is similar to mutual information and mixed information. Another notable observation is that the accuracy of variational information decreases as the window size increases. KL_EW and KLD_$\mathcal{G}^0$D have lower accuracies of around 90%. For higher order statistical changes as shown in Fig. 19(d), all similarity measures have similar behaviors as for second order statistical changes.

For linear changes shown in Fig. 19(e), both variational and mutual information have an inferior performance than in the the case of reflectivity changes. However, mixed information
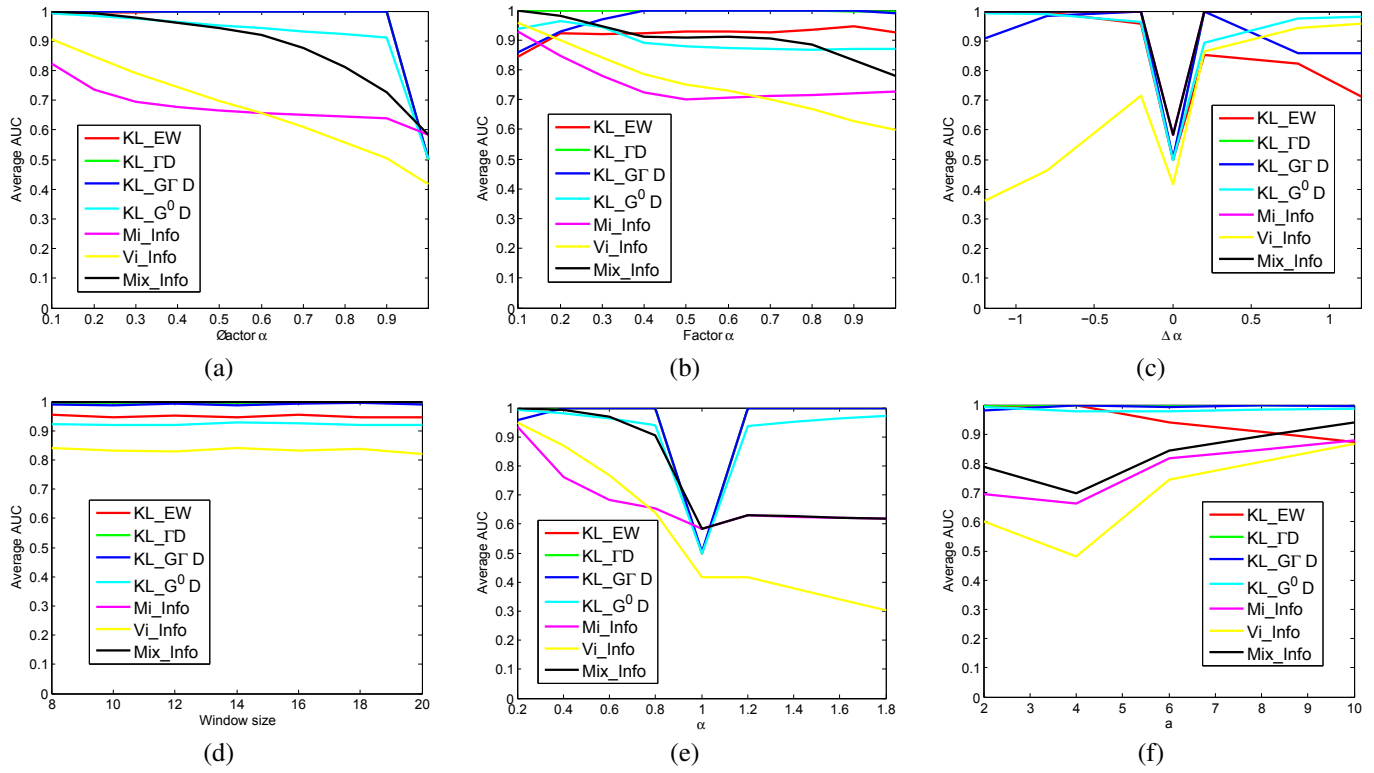
Fig. 21. Influence of the parameters shown in Table I on the change detection accuracy: (a) reflectivity changes; (b) first order statistical changes; (c) second order statistical changes; (d) higher order statistical changes; (e) linear changes; (f) nonlinear changes.

performs much better. This can be explained by Fig. 20(c). KLD_ΓD, KLD_GΓD, and KL_EW are on top with accuracies of more than 99%. Mixed information and KLD_$\mathcal{G}^0$D have similar performances which are about 2% lower than that of KL_EW. For nonlinear changes shown in Fig. 19(f), Kullback-Leibler divergence shows an obvious advantage over mutual information, variational information and mixed information. Mixed information has the best accuracy among these three similarity measures, but it is still much lower than the accuracy obtained by Kullback-Leibler divergence.

Since each kind of change simulation involves a parameter, its influence on change detection accuracy is presented in Fig. 21. For reflectivity changes, the parameter α fully controls the degree of change. Therefore, as expected, the accuracy decreases as α increases from 0.1 to 1.0, as shown in Fig. 21(a). For first order statistical changes, we change the pixel probability density distribution to a Gaussian distribution by varying the standard deviation. Thus, the parameter α does not fully control the degree of change. Its influence is shown in Fig. 21(b). For second order statistical changes, the effect of $\Delta\alpha$ is shown in Fig. 21(c). We can see that the accuracy decreases as $\Delta\alpha$ becomes close to zero. The influence of the patch size used in higher order statistical change simulation is shown in Fig. 21(d). It is clear that it does not have much influence on the accuracy. Similar to the parameter $\Delta\alpha$ in second order statistical changes, the effect of varying parameter a in linear changes is shown in Fig. 21(e). Finally, the influence of parameter a on nonlinear changes is shown in Fig. 21(f). In addition, we can observe that all the similarity
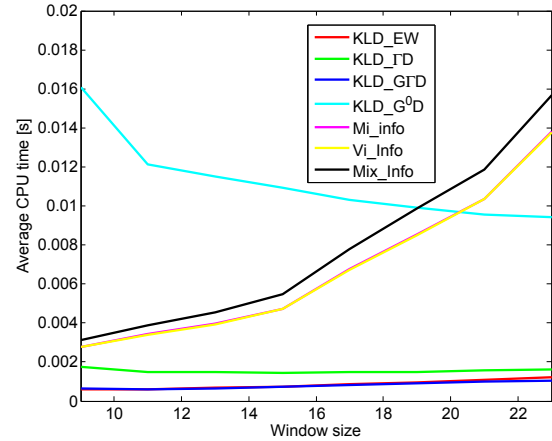

Fig. 22. Computational complexity comparison of all similarity measures in terms of average CPU time.

measures behave consistently as shown in Fig. 19.

The computational complexities of all similarity measures versus window size are presented in Fig. 22. KL_EW and KLD_GΓD have the lowest computational complexity because these two methods can be computed analytically without any other numerical optimization. Although KLD_ΓD can be computed analytically by (16), we have to rely on a numerical optimization method for parameter estimation, which is, in our case, a Levenberg-Marquardt method [35]. Thus, it has a bit higher computational complexity than KL_EW and KLD_GΓD. In the case of KLD_$\mathcal{G}^0$D, the Gauss-Kronrod

quadrature algorithm for computing the integral in (18) slows down the method significantly, in addition to the numerical method for parameter estimation. However, a notable advantage of Kullback-Leibler divergence is that the computational complexity does not increase with window size because the parameter estimation relies only on log-cumulants of the samples, which is different from KDE being used for the other three similarity measures. Mutual information and variational information have similar computational complexities. Mixed information is slightly slower than mutual information and variational information because a number of change index maps have to be computed for a series of parameter α. The computational time of these three similarity measures increases rapidly with window size because more and more samples are available for density estimation. The most time-consuming part for these three methods is kernel density estimation, which is implemented in a MATLAB MEX function using C language.

To sum up, different methods perform quite differently for different kinds of changes. Kullback-Leibler divergence generally performs quite well for SAR image change detection provided that a good SAR image model is chosen. In most of our cases, variational information has an inferior performance. Mutual information is better to be applied for detecting second and higher order statistical changes. It is necessary to search for the best α parameter when we apply mixed information to SAR change detection, and gives a better accuracy than mutual information.

V. EVALUATION ON REAL SAR DATA SETS

In this section, all the similarity measures are evaluated using real data of TerraSAR-X images.

A. Selected Datasets

We selected five datasets of TerraSAR-X images corresponding to reflectivity changes and statistical changes. The first four datasets are selected, as shown in the first two rows in Fig. 23, from two radiometrically enhanced TerraSAR-X images acquired in Stripmap mode prior to (on Oct. 20, 2010) and after (on May 6, 2011) the Sendai earthquake in Japan. Their pixel spacing is about 2.5 m. The sizes of these four image pairs are respectively 549×560, 613×641, 590×687, and 689×734 pixels. As the two TerraSAR-X images acquired before and after the disaster have nearly the same imaging parameters, we can reasonably assume that there is only a linear geometrical translation between the images. To achieve precise registration, ten strong point scatterers from each image were manually selected to determine the translation along both azimuth and range direction. To make sure that the translation is precise, the normalized cross-correlation is computed to check the translation and the residual pixel shift is less than one pixel. The reference data shown in the third row were produced through careful manual interpretation of optical images. In optical images, these changes are more visible. Thus, referring to optical images makes the reference data reliable. Due to the earthquake, a tsunami occurred, which led to a devastating flooding, as can be seen from the images. In the first two datasets, the agricultural fields were severely flooded and the image intensities changed dramatically; these are appropriate scenarios for performance assessment in detecting reflectivity changes. The third and fourth datasets contain both reflectivity and statistical changes, which were used for assessing statistical changes. These four datasets are abbreviated as "Sendai1", "Sendai2", "Sendai3", and "Sendai4" in the following descriptions and figures. The fifth dataset is selected from another TerraSAR-X image covering the Vâlcea county in Romania. The images have a pixel spacing of 2.5 m and an incidence angle of around 36º. The image registration was performed as described above. In this scenario, some underwater grass and vegetation in the three lakes changed the statistical characteristics of the images. This dataset is referred to as "Vâlcea" in the following sections and figures. The experimental setup for evaluation follows the evaluation of the synthetic data. In the following sections, change index maps are shown using color for highlighting.

B. Results on Real Datasets

The average AUCs of all the similarity measures versus window size on the five datasets are shown in the first row in Fig. 24. The detailed results of AUCs, TPRs, and FARs are presented in the supplementary material. Different from the evaluation on synthetic data, Mix_Info, Mi_Info and Vi_Info have similar performances as indicated by the ROC curves shown in Fig. 25, although Mix_Info performs slightly better. The impact of α for the first three datasets is shown in Fig. 26, where we can see that the impact of α is negligible. The best change index maps of the five datasets by Mix_Info, Mi_Info and Vi_Info are shown in Fig. 25. By comparing the performances on the first two datasets and other three datasets, we can see they are much better at detecting statistical changes than reflectivity changes.

In contrast to these three information measures, Kullback-Leibler divergence performs better for the first two datasets, which mainly consist of reflectivity changes. For the remaining three datasets, especially the third and the fifth one, Kullback-Leibler divergence performs worse with an accuracy of around 80% for the third dataset and 70% for the fifth dataset. This result is consistent with the conclusion drawn from the evaluation of synthetic data that Kullback-Leibler performs well for reflectivity changes and both mutual information and mixed information perform well in detecting changes in second and higher order statistics.

As for the parametric models used in computing the Kullback-Leibler divergence, all the three models, i.e., Gamma distribution, GΓD, and $\mathcal{G}^0$ distribution, have a similar performance. For example, in the case of the second and the fourth dataset, the best change index maps (in terms of AUC) of KLD_EW, KLD_ΓD, KLD_GΓD, and KLD_$\mathcal{G}^0$D are shown in Fig. 27. However, as can be seen from the first column in Fig. 27, there are many false detections, which can be confirmed by the TPRs and FARs presented in the supplementary material. The reason is that KLD_EW relies on an Edgeworth series expansion, which is based on the assumption that the pixel amplitude distribution is close to a Gaussian, which is not
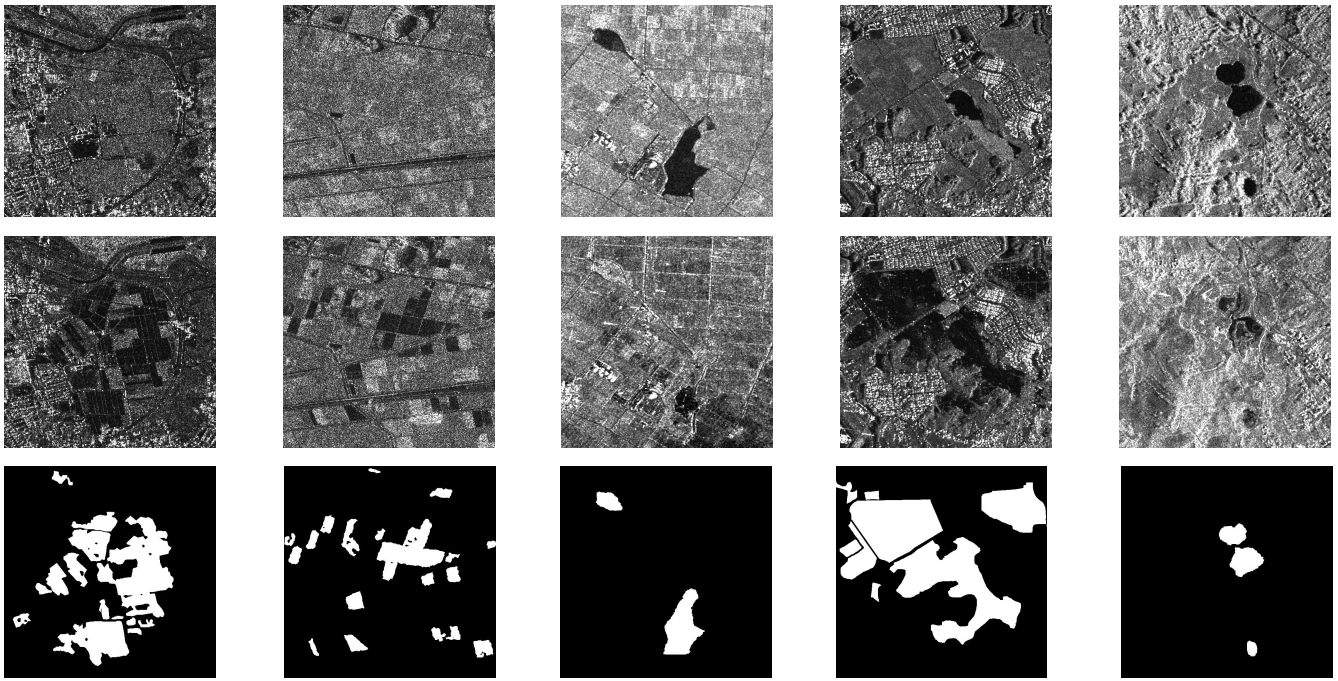
Fig. 23. TerraSAR-X datasets (from left to right: "Sendai1", "Sendai2", "Sendai3", "Sendai4", and "Vâlcea") being used for evaluation. The first row are the images before changes and the second row are the images after changes. The third row are the reference data generated by manual interpretation.
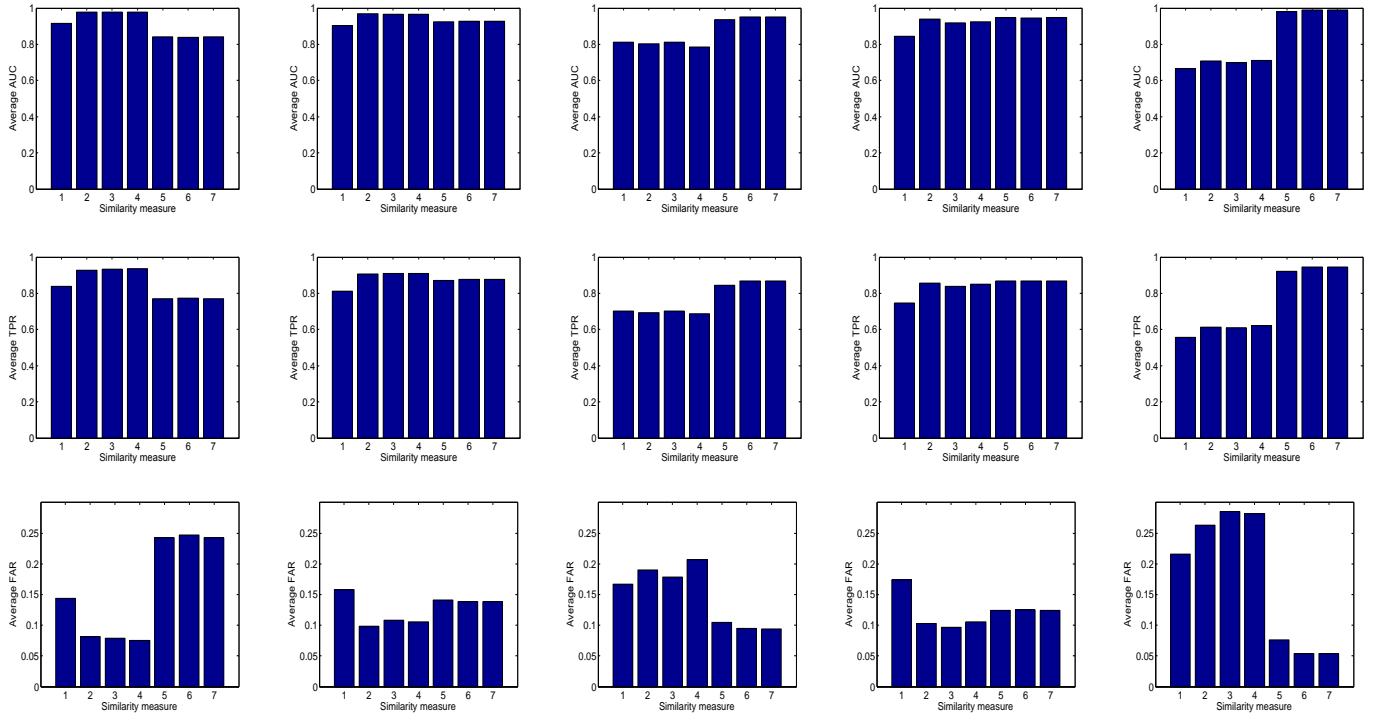

Fig. 24. The average AUCs, TPRs, and FARs of the five datasets (from left to right: "Sendai1", "Sendai2", "Sendai3", "Sendai4", and "Vâlcea"). Each column shows the average AUCs, TPSs, and FARs for each dataset. The horizontal axis of each bar plot denotes the similarity measure. Columns 1 to 7 of each bar plot represent KL_EW, KLD_ΓD, KLD_GΓD, KLD_$\mathcal{G}^0$D, Mi_Info, Vi_Info, and Mix_Info.

the case for SAR images. Furthermore, a fluctuating Hermite polynomial is used for series expansion, which could introduce false detections. Please note that there are some singular points in the change index maps of KLD_$\mathcal{G}^0$D shown in the fourth columns. For those pixels, no solutions can be found. Local

interpolation is applied to fill the missing values. However, the change index maps obtained by both KLD_ΓD and KLD_GΓD have almost no numerically invalid values, which results in better convergence in solving the equations of the Levenberg-Marquardt algorithm.
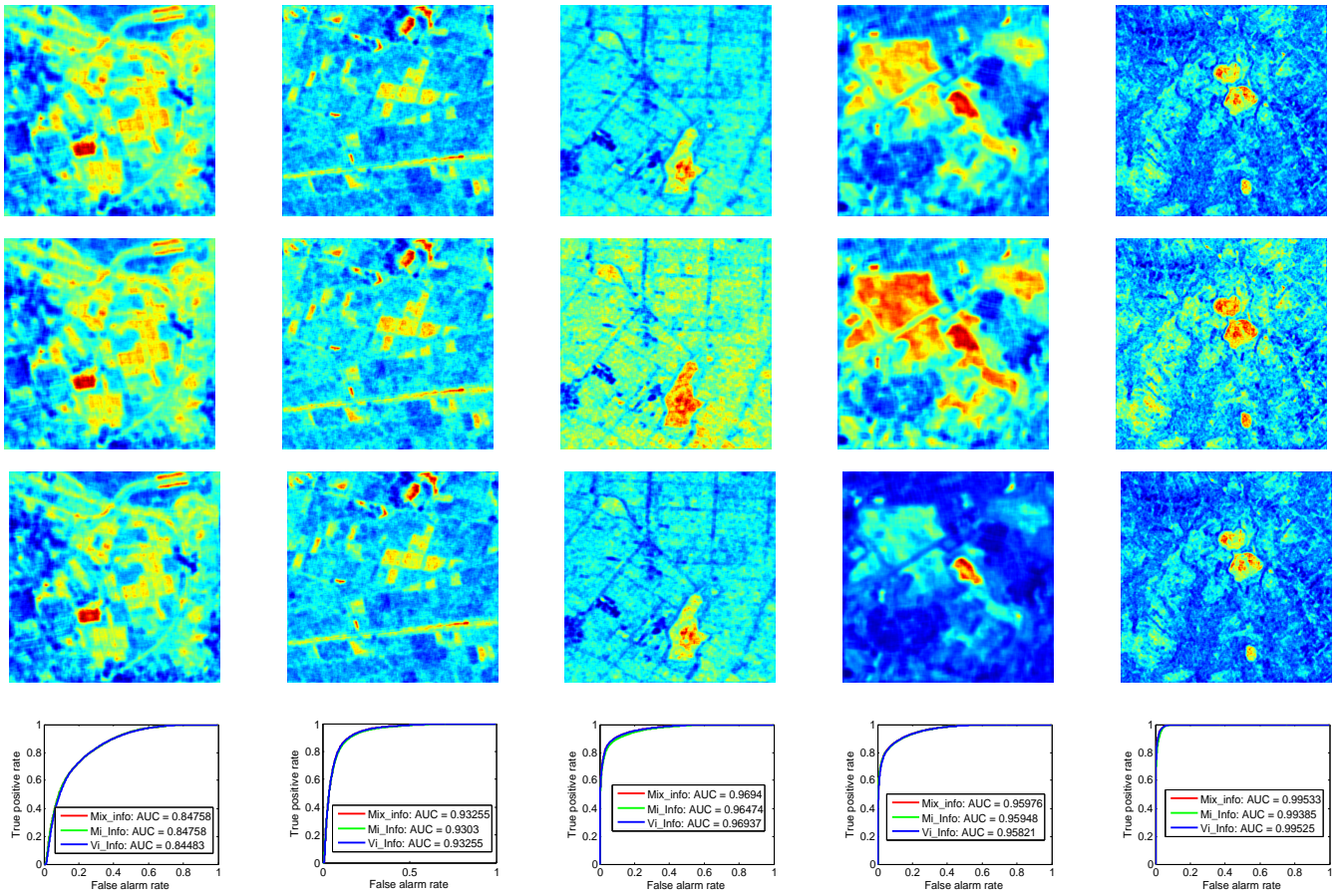
Fig. 25. The best change index map (in terms of AUC) and ROC curves of the five datasets (from left to right: "Sendai1", "Sendai2", "Sendai3", "Sendai4", and "Vâlcea") obtained by Mix_Info, Mi_Info and Vi_Info (row by row form top to bottom). We can observe that the accuracies of all these three similarity measures are similar. Therefore, the influence of the parameter α is negligible for the best cases. Note that the change index maps in the second row are inverted such that changed pixels have large values.
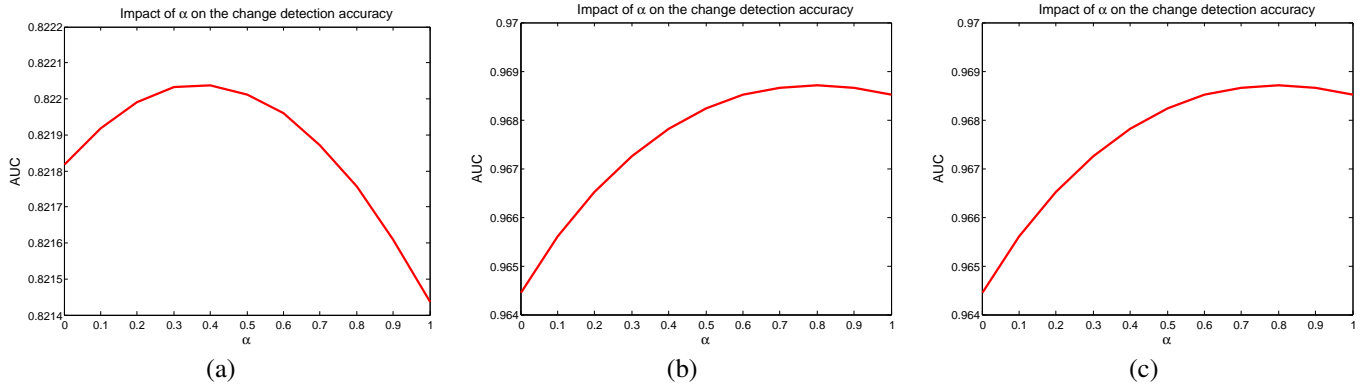

Fig. 26. Impact of α on the performance of the mixed information measure for real datasets: (a) the first dataset "Sendai1" (α has a similar impact on the fourth dataset "Sendai4"); (b) the second dataset "Sendai2"; (c) the third dataset "Sendai3" (α has a similar impact on the fifth dataset "Vâlcea").

From the TPRs and FARs shown in Fig. 24, we can see that Kullback-Leibler divergences estimated by KLD_ΓD, KLD_GΓD, and KLD_$\mathcal{G}^0$D perform much better than Mix_Info, Mi_Info and Vi_Info for the first two datasets and achieve similar performance for the fourth dataset. In contrast, Mix_Info, Mi_Info and Vi_Info have better accuracy than KLD_ΓD, KLD_GΓD, and KLD_$\mathcal{G}^0$D for the third and the fifth data sets, which is consistent with the conclusion drawn previously from the evaluation on synthetic data. This can

be seen from the change index maps generated by Kullback-Leibler divergence for the third and the fifth datasets shown in Fig. 28. Comparing the change index maps in Fig. 28 with the ones shown in Fig. 25, we can see that the change index maps generated by Kullback-Leibler divergence for the third and the fifth datasets are inferior.

In addition, it seems that KLD_EW has a better performance when using a smaller window size and degrades as the window size increases. For the second and fifth datasets, KLD_EW
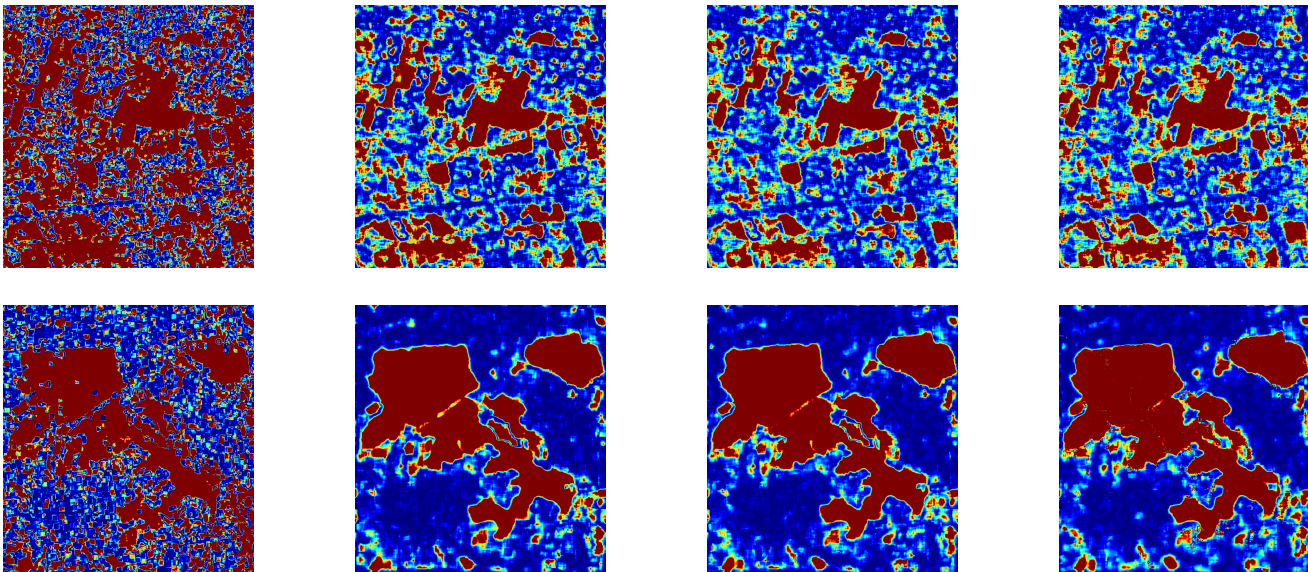
Fig. 27.   Change index maps of the second dataset "Sendai2" (first row) and the fourth dataset "Sendai4" (second row) datasets obtained by KLD_EW, KLD_ΓD, KLD_GΓD, and KLD_$\mathcal{G}^0$D (from left to right).
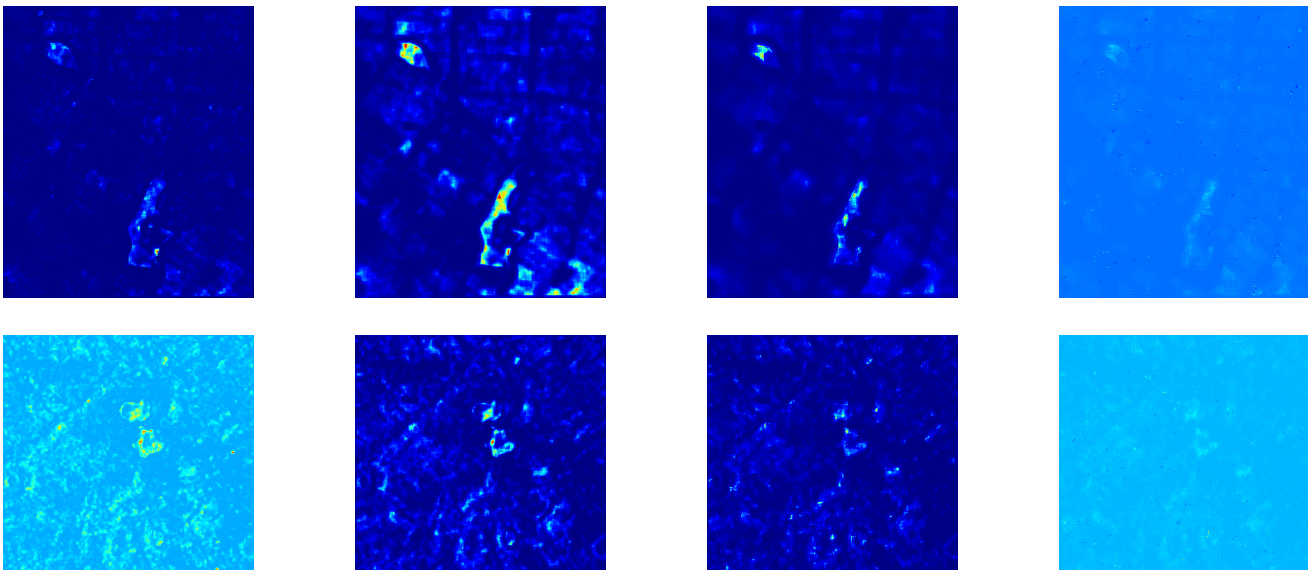

Fig. 28.   The best change index maps for the third dataset "Sendai3" (first row) and the fifth dataset "Vâlcea" (second row) datasets obtained by KLD_EW, KLD_ΓD, KLD_GΓD, and KLD_$\mathcal{G}^0$D (from left to right).

achieves the best performance with a window size of 9×9 pixels. Similarly, it achieves its maximum accuracy for the first and third data sets with a window size of 11×11 pixels. Beyond that point, the accuracy decreases when the window size is expanded. The other three information measures achieve the best accuracy using a window size of around 15×15 pixels.

VI. Conclusion

Although a large number of different methods have been developed for SAR image change detection, there is no consensus about their performances because most methods have been evaluated using different datasets, facing probably different kinds of changes. Thus, it is not fair to claim one method is superior or inferior to another method because they are not evaluated using the same kind of changes. In practice, several different kinds of changes with different characteristics can arise. To solve these two problems, we have to identify different kinds of changes and develop a method to represent their characteristics. Although the importance to model different kinds of changes has been realized, there is no principled methodology to carry out the analysis due to the difficulty in modeling different kinds of changes. Another problem is the lack of standard benchmark approaches and datasets. We believe this is a severe problem, which should be

solved by joint efforts of the entire community.

In this paper, we propose a methodology that can partially solve these problems. We propose a principled methodology for evaluating change detection methods by simulating different kinds of changes. Six kinds of SAR changes for eight typical categories are simulated, i.e., reflectivity changes, first order, second order, and higher order statistical changes, linear and nonlinear changes. Based on this methodology, we try to find out what kind of changes each approach can detect and to what extent the different kinds of changes can be detected. Based on this methodology for change simulation, a comprehensive evaluation of information similarity measures is carried out. An explicit conclusion we have drawn from the evaluation is that different methods behave very differently for different kinds of changes. Kullback-Leibler divergence generally performs quite well for SAR image change detection provided that a good SAR image model is chosen. Mutual information is more appropriate to be applied for detecting second and higher order statistical changes. It is necessary to search for the best α parameter when we apply mixed information to SAR change detection, which often gives a better accuracy than mutual information. We hope this study will promote the advancement of this topic.

REFERENCES

[1] [Online], "TerraSAR-X," Available: http://en.wikipedia.org/wiki/TerraSAR-X.

[2] L. Bruzzone and D. F. Prieto, "An Adaptive Semiparametric and Context-based Approach to Unsupervised Change Detection in Multi-temporal Remote-sensing Images," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 452–466, Apr. 2002.

[3] L. Bruzzone and D. F. Prieto, "Automatic Analysis of the Difference Image for Unsupervised Change Detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[4] E. J. M. Rignot and J. J. van Zyl, "Change Detection Techniques for ERS-1 SAR Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 4, pp. 896–906, Jul. 1993.

[5] 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.

[6] T. Celik, "A Bayesian approach to unsupervised multiscale change detection in synthetic aperture radar images," *Signal Process.*, vol. 90, no. 5, pp. 1471–1485, May 2010.

[7] J. Inglada and G. Mercier, "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.

[8] F. Bovolo and L. Bruzzone, "An Adaptive Technique based on Similarity Measures for Change Detection in Very High Resolution SAR Images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Boston, Massachusetts, USA, Jul. 2008, vol. 3, pp. III–158 – III–161.

[9] S. Cui and M. Datcu, "Statistical Wavelet Subband Modeling for Multi-Temporal SAR Change Detection," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.*, vol. 5, no. 4, pp. 1095–1109, Aug. 2012.

[10] V. Alberga, "Similarity Measures of Remotely Sensed Multi-sensor Images for Change Detection Applications," *Remote Sensing*, vol. 1, no. 3, pp. 122–143, 2009.

[11] A. Winter, H. Maitre, N. Cambou, and E. Legrand, "Entropy and multiscale analysis: a new feature extraction algorithm for aerial images," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Munich, Germany, Apr. 1997, vol. 4, pp. 2765–2768.

[12] L. Gueguen and M. Datcu, "Mixed Information Measure: Application to Change Detection in Earth Observation," in *Proc. 5th International Workshop on the Analysis of Multi-temporal Remote Sensing Images, MultiTemp*, Connecticut, USA, Jun. 2009.

[13] G. Mercier, S. Derrode, W. Pieczynski, J.-M. Nicolas, A. Joannic-Chardin, and J. Inglada, "Copula-based Stochastic Kernels for Abrupt Change Detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.(IGARSS)*, Denver, Colorado, USA, Aug. 2006, pp. 204–207.

[14] F. Chatelain, J. Y. Tourneret, J. Inglada, and A. Ferrari, "Bivariate Gamma Distributions for Image Registration and Change Detection," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1796–1806, Jul. 2007.

[15] L. Gueguen, M. Pesaresi, D. Ehrlich, and Linlin Lu, "Urbanization analysis by mutual information based change detection between SPOT 5 panchromatic images," in *Proc. 6th International Workshop on the Analysis of Multi-temporal Remote Sensing Images (MultiTemp)*, Trento, Italy, Jul. 2011, pp. 157–160.

[16] G. Mercier and J. Inglada, "Change detection with misregistration errors and heterogeneous data through the Orfeo toolbox," Tech. Rep., Institut TELECOM, TELECOM Bretagne, 2008.

[17] Y. Bazi, L. Bruzzone, and F. Melgani, "Image thresholding based on the EM algorithm and the generalized Gaussian distribution," *Pattern Recognition*, vol. 40, no. 2, pp. 619 – 634, 2007.

[18] Y. Bazi, F. Melgani, L. Bruzzone, and G. Vernazza, "A genetic expectation-maximization method for unsupervised change detection in multitemporal SAR imagery," *International Journal of Remote Sensing*, vol. 30, no. 24, pp. 6591–6610, 2009.

[19] J. Angus, "The Probability Integral Transform and Related Results," *SIAM Review*, vol. 36, no. 4, pp. 652–654, 1994.

[20] A. C. Frery, H.-J. Muller, C. C. F. Yanasse, and S. J. S. Sant'Anna, "A Model for Extremely Heterogeneous Clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, May 1997.

[21] S. Cui, G. Schwarz, and M. Datcu, "A Comparative Study of Statistical Models for Multilook SAR Images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 175–1756, Oct. 2014.

[22] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Vol. 1 (Wiley Series in Probability and Statistics)*, Wiley-Interscience.

[23] A. Torralba and A. Oliva, "Statistics of natural image categories," *Network: Computation in Neural Systems*, pp. 391–412, Aug. 2003.

[24] R. Paget and I. D. Longstaff, "TextureSsynthesis via a Noncausal Nonparametric Multiscale Markov Random Field," *IEEE Trans. Image Process.*, vol. 7, no. 6, pp. 925–931, Jun. 1998.

[25] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *Proceedings of the 28th annual conference on computer graphics and interactive techniques (SIGGRAPH)*, New York, NY, USA, 2001, pp. 341–346, ACM.

[26] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, Jul. & Oct. 1948.

[27] M. Meila, "Comparing Clusterings by the Variation of Information," in *Learning Theory and Kernel Machines*, Bernhard Schoelkopf and Manfred K. Warmuth, Eds., vol. 2777 of *Lecture Notes in Computer Science*, pp. 173–187. Springer Berlin Heidelberg, 2003.

[28] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[29] J. Goldberger, S. Gordon, and H. Greenspan, "An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures," in *Proc. Ninth IEEE International Conference on Computer Vision (ICCV)*, Nice, France, Oct 2003, pp. 487–493.

[30] J. R. Hershey and P. A. Olsen, "Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, HawaII, USA, Apr. 2007, vol. 4, pp. IV–317–IV–320.

[31] L. Gueguen, P. Soille, and M. Pesaresi, "Change Detection Based on Information Measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4503–4515, Nov. 2011.

[32] H. Shimazaki and S. Shinomoto, "Kernel bandwidth optimization in spike rate estimation," *Journal of Computational Neuroscience*, vol. 29, pp. 171–182, 2010.

[33] J. M. Nicolas, "Introduction to second kind statistics: Application of log-moments and log-cumulants to analysis of radar images," *Traitement du Signal*, vol. 19, no. 3, pp. 139–167, 2002.

[34] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, New York, NY, USA, 3rd edition, 2007.

[35] C. Kanzow, N. Yamashita, and M. Fukushima, "Levenberg-Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints," *J. Comput. Appl. Math.*, vol. 172, no. 12, pp. 375–397, Dec. 2004.