INTERNATIONAL JOURNAL OF REMOTE SENSING, 2016
VOL. 37, NO. 14, 3232–3254
http://dx.doi.org/10.1080/01431161.2016.1196838

Taylor & Francis
Taylor & Francis Group

Unsupervised change detection in SAR images using curvelet and L1-norm based soft segmentation

Fang Li[a], Faming Fang[b] and Guixu Zhang[b]

[a]Department of Mathematics, East China Normal University, Shanghai, China; [b]Department of Computer Science, East China Normal University, Shanghai, China

ABSTRACT

In this article, we propose a novel unsupervised change detection method for synthetic aperture radar (SAR) images. First, we generate a difference image as a weighted average of a log-ratio image and a mean-ratio image, which has the advantage of enhancing the information of changed regions and restraining the information of unchanged background regions simultaneously. Second, we propose a variational soft segmentation model based on non-differentiable curvelet regularization and L1-norm fidelity. Numerically, by using the split Bregman technique for curvelet regularization term and reformulating the L1-norm fidelity as weighted L2-norm fidelity, we get an effective algorithm in which each sub-problem has a closed-form solution. The numerical experiments and comparisons with several existing methods show that the proposed method is promising, with not only high robustness to non-Gaussian noise or outliers but also high change detection accuracy. Moreover, the proposed method is good at detecting fine-structured change areas. Especially, it outperforms other methods in preserving edge continuity and detecting curve-shaped changed areas.

ARTICLE HISTORY
Received 3 September 2015
Accepted 27 May 2016

1. Introduction

Image change detection aims to detect the changed areas in images of the same scene taken at different time instances (Radke et al. 2005). In recent years, there has been a growing interest in developing change detection techniques for the analysis of multitemporal remote-sensing images due to its wide range of applications in environmental monitoring, agricultural surveys, urban studies, forest monitoring, etc. (Bruzzone and Prieto 2000). In particular, change detection in synthetic aperture radar (SAR) images is an attractive topic since SAR sensors are independent of atmospheric and sunlight conditions. However, SAR images change detection is more difficult than optical ones due to the presence of speckle noise.

The existing change detection methods can be categorized as either supervised or unsupervised methods. The former is based on supervised classification and requires some ground truth data to train the classifier, while the latter performs change

CONTACT Guixu Zhang gxzhang@cs.ecnu.edu.cn Department of Computer Science, East China Normal University, Shanghai, China

detection directly on the two multitemporal images without any other additional information (Celik 2009). Although the supervised approach exhibits some advantages over the unsupervised one, the generation of an appropriate ground truth is usually a difficult and expensive task (Bruzzone and Prieto 2000). Therefore, the designing of effective unsupervised change detection methods is fundamental in many application areas when the ground truth is unavailable. In this article, we focus on unsupervised methods.

In general, there are mainly three steps in unsupervised change detection for remote-sensing images. The first step is pre-processing including noise reduction, co-registration, and geometric correction (Leprince et al. 2007). Regarding SAR, the registration in general is done automatically. An advanced filtering approach is proposed in Schmitt, Wendleder, and Hinz (2015) for the joint pre-processing of multiple input images, which is shown to improve the radiometric accuracy while preserving the geometric resolution of SAR images. In this article, we assume the first step has been carried out on the multitemporal images before applying any change detection method.

The second step is generating difference image between the multitemporal images. In the difference image, it is required that the changed areas and unchanged areas are significantly different. Basically, there are two methods to produce difference images (Rignot and van Zyl 1993; Bujor et al. 2004). One is differencing, in which the difference image is produced by subtracting the intensity of the two multitemporal images pixel by pixel. The other is ratioing, in which the difference image is generated by applying ratio operator on the two multitemporal images pixel by pixel. In the case of SAR images, due to the multiplicative nature of speckle noise, the difference image is usually taken as the logarithmic or a mean scale of the ratio image, which are called log-ratio and mean-ratio, respectively (Inglada and Mercier 2007). Note that a detailed explanation of log-ratio and mean-ratio will be given in the following context. Both log-ratio and mean-ratio difference image are robust to calibration error and speckle noise. Based on log-ratio and mean-ratio, some other kind of difference images are proposed for SAR images. A wavelet fusion technique is introduced to generate the difference image as a fusion of the basic log-ratio and mean-ratio images (Gong, Zhou, and Ma 2012). A robust difference image is constructed by the weighted average of pixel similarity and patch similarity measured by symmetrized log-ratio (Zhang, Chen, and Meng 2013). In Zheng et al. (2014), a new difference image is generated by combining the differencing image processed by mean filter and log-ratio image processed by median filter.

In the third step, changes are identified by analysing the difference image with some unsupervised classification method. In this step, changed areas and unchanged areas are classified. The classification method can be applied either on the difference image directly or on the features extracted from the difference image.

The following methods are applied on the difference image directly. In Bruzzone and Prieto (2000), two automatic thresholding methods based on the Bayes theory are proposed, which are expectation-maximization (EM) based thresholding and Markov random fields (MRF) based thresholding. In Zhang, Chen, and Meng (2013), an SAR image change detection method based on graph-cut and generalized Gaussian model is proposed. In Zheng et al. (2014), the classical *k*-means clustering method with two classes is applied to the combined difference image. The variants of fuzzy c-means (FCM) clustering method incorporating local information (Krinidis and Chatzis 2010) are

adopted in Gong, Zhou, and Ma (2012). In Gong et al. (2014), a fuzzy clustering with a modified MRF energy function is proposed for change detection.

By contrast, some other methods include a feature selection process and apply classification on the features instead of on the original difference image. In Bovolo and Bruzzone (2005), undecimated discrete wavelet transform (UDWT) is used to extract multiresolution features from the log-ratio difference image and the final change detection map according to a scale-driven fusion strategy. The principal component analysis (PCA) technique has been used in Celik (2009) to extract the features from the overlapping patches of the difference image, and then *k*-means is used to clustering the features into two classes. In Celik and Ma (2010), dual-tree complex wavelet transform (DT-CWT) has been applied on the two multitemporal images to get the multiresolution difference image, and binary change detection has been conducted on each subband and then they are fused as the final change map. In Celik and Ma (2011), UDWT is applied on original difference image to extract multiresolution difference images, and then a two-phase vectored-valued Chan–Vese model (Chan and Vese 2001) is used to segment the vector-valued difference image. An unsupervised change detection from multichannel SAR data by Markovian data fusion is proposed in Moser and Serpico (2009).

The above-mentioned methods provide impressive change detection results with relatively high detection accuracy. However, their performance is still limited in handling non-Gaussian noise, outliers, and detecting fine-structured areas. To overcome this limitation, in this article, we propose a new segmentation method based on curvelet transform and L1-norm minimization.

Our motivation comes from the following observations. First, curvelet transform is very efficient in encoding images with both smooth regions and edges, which is widely used in image processing. Indeed, curvelet-based approaches for change detection are considered in Schmitt, Wessel, and Roth (2009, 2014), which shows that curvelet transform is a powerful tool to describe fine-structured change areas. Hence we choose curvelet regularization in the proposed segmentation method. It is notable that our method is very different from the methods in Schmitt, Wessel, and Roth (2009, 2014) since no segmentation model is involved in the latter. Second, L1-norm fidelity is proved to be more robust to non-Gaussian noise and outliers than its L2-norm counterpart, which is widely used in image restoration problem (Chan and Esedoglu 2005; Guo, Li, and Ng 2009; Jung, Kang, and Kang 2014). This motivates us to use L1-norm in the proposed segmentation model to fitting the difference image, which contains non-Gaussian noise and outliers.

Since the performance of image change detection depends on both the quality of the difference image and the accuracy of the classification method, we build our method by considering a new difference image and a new classification method. The contribution of this work is two-fold. First, we produce a difference image as a weighted average of the log-ratio image and the mean-ratio image, which can enhance the information of changed regions and at the same time restrain the information of unchanged background regions. Second, we propose a new variational soft segmentation model based on curvelet regularization and L1-norm fidelity. In the numerical aspect, to overcome the non-smoothness of the energy functional, split Bregman technique (Goldstein and Osher 2009) is used for curvelet regularization term, and meanwhile the L1-norm fidelity is

reformulated as weighted L2-norm fidelity. Then we derive an effective algorithm in which each sub-problem has closed-form solution.

2. Curvelet transform and Chan–Vese model

Before the presentation of our technique, in this section we describe some elements useful for our purpose.

2.1. *Curvelet transform*

The curvelet transform (Candes and Donoho 2004; Candes et al. 2006) was developed to overcome inherent limitations of traditional multiscale representations such as wavelets. The curvelet transform is a multiscale pyramid with many directions and positions at each length scale and needle-shaped elements at fine scales. Curvelets are indexed by three parameters: the scale 2^{-j}, $j = 0,1,\ldots$; the equispaced sequence of rotation angle θ_ℓ with $\ell = 0,1,\ldots$ such that $0 \le \theta_\ell < 2\pi$; the spatial translation parameters k_1, k_2 are integers. Assume $\varphi_j(\boldsymbol{x})$ is the 'mother' curvelet; see Candes et al. (2006) for detailed construction of φ_j. With these notations, the curvelets as a function of $\boldsymbol{x} = (x_1,x_2)$ at scale 2^{-j}, orientation θ_ℓ, and position $x_k^{(j,\ell)} = \mathbf{R}_{\theta_l}^{-1}(k_1 2^{-j}, k_2 2^{-l})$ is given by

$$\varphi_{j,\ell,k}(\boldsymbol{x}) = \varphi_j\left(\mathbf{R}_{\theta_\ell}\left(\boldsymbol{x} - x_k^{(j,\ell)}\right)\right), \tag{1}$$

where $\mathbf{R}_\theta$ is the rotation

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{2}$$

A curvelet coefficient $c_{j,\ell,k}$ is then simply the inner product between an element $f \in \mathbf{L}^2(\cdot^2)$ (R^n denotes the n-dimensional Euclidean space and $\mathbf{L}^2(\cdot^2)$ denotes the space of square-integrable functions.) and a curvelet $\varphi_{j,\ell,k}$:

$$c_{j,\ell,k} = \langle f, \varphi_{j,\ell,k}\rangle = \int_{\bullet R^2} f(x)\overline{\varphi_{j,\ell,k}(\boldsymbol{x})}\mathrm{d}\boldsymbol{x} \tag{3}$$

Here $\langle\rangle$ denotes inner product of two functions, which is defined by the last formula in (3) and $\overline{\varphi_{j,\ell,k}(\boldsymbol{x})}$ denotes the conjugate of $\varphi_{j,\ell,k}(\boldsymbol{x})$. The family of curvelet functions $\{\varphi_{j,\ell,k}(\boldsymbol{x})\}$ forms a tight frame of $\mathbf{L}^2(\bullet^2)$ such that an arbitrary function $f \in \mathbf{L}^2(\bullet^2)$ can be expanded by a series of curvelets, that is, we have the reconstruction formula:

$$f = \sum_{j,\ell,k}\langle f, \varphi_{j,\ell,k}\rangle\varphi_{j,\ell,k}. \tag{4}$$

For simplicity, we denote the curvelet transform as operator C, that is, Cf represents all the curvelet coefficients $\{c_{j,\ell,k}\}$ of function f. We denote C^{T} as the conjugate operator of C, which is the curvelet reconstruction operator. By the perfect construction property of curvelet transform, the following equality holds:

$$\mathcal{C}^{\mathrm{T}}\mathcal{C} = I, \tag{5}$$

where I denotes the identity transform.

The advantage of curvelet transform is that it is very efficient in encoding images with both smooth regions and edges. To make use of this advantage, curvelet-based approaches for change detection are considered in Schmitt, Wessel, and Roth (2009, 2014). In this article, we will introduce curvelet regularization in our segmentation model for change detection in order to well detect the fine-structured change areas. Note that the software package CurveLab implements the curvelet transform and is available at http://www.curvelet.org.

2.2. *Chan–Vese model*

One of the most well known variational segmentation model is the Chan–Vese model (Chan and Vese 2001). Let us recall the following two-phase Chan–Vese model in level set formulation:

$$\min_{\phi,c_1,c_2}\left\{\begin{array}{l}\int_\Omega |\nabla H(\phi)|\mathrm{d}\boldsymbol{x}+\mu\int_\Omega H(\phi)\mathrm{d}\boldsymbol{x}+\lambda_1\int_\Omega |\mathbf{I}-c_1|^2H(\phi)\mathrm{d}\boldsymbol{x}\\ +\lambda_2\int_\Omega |\mathbf{I}-c_2|^2(1-H(\phi))\mathrm{d}\boldsymbol{x}\end{array}\right\}, \tag{6}$$

where $\mathbf{I}:\Omega\subset \cdot^2\rightarrow \cdot$ is the given image, c_1 and c_2 are constants, $\phi:\Omega\rightarrow \bullet$ is the level set function, $H(\phi)$ is the Heaviside function satisfying

$$H(\phi)=\left\{\begin{array}{ll}0, & \text{if } \phi\le 0\\ 1, & \textit{otherwise}\end{array}\right.. \tag{7}$$

λ_1, λ_2, and μ are fixed positive parameters. The final segmentation curve is given by $\phi = 0$. The two classes are $\phi > 0$ and $\phi < 0$. The first term in the energy (6) is regularization term, which requires that the segmentation boundary should be as short as possible. The second term requires that the area of the region $\phi > 0$ should be as small as possible. The last two terms are fidelity terms, which require that the image should be approximated by constants c_1 and c_2 in each region. In other words, the Chan–Vese model seeks a piecewise constant approximation of the image with smooth boundary. Numerically, negative Gradient Descent method is used to solve the associated Euler–Lagrange equation for level set function ϕ. Since this equation is highly nonlinear, the numerical scheme is not easy to implement and converges slowly due to the small size of the time step, see Chan and Vese (2001) for more details.

The soft version of the Chan–Vese model (Chan, Esedoglu, and Nikolova 2006; Bresson et al. 2007) is

$$\min_{0\le\mathbf{u}\le 1,c_1,c_2}\left\{\begin{array}{l}\int_\Omega |\nabla\mathbf{u}|\mathrm{d}\boldsymbol{x}+\lambda_1\int_\Omega |\mathbf{I}-c_1|^2\mathbf{u}\mathrm{d}\boldsymbol{x}\\ +\lambda_2\int_\Omega |\mathbf{I}-c_2|^2(1-\mathbf{u})\mathrm{d}\boldsymbol{x}\end{array}\right\}, \tag{8}$$

in which a soft membership function $\mathbf{u}$ ranged in [0,1] is used to replace the binary Heaviside function $H(\phi)$ in the Chan–Vese model. Note that the second term in the ordinal Chan–Vese model is not considered here since its function is negligible in many cases. It can be explained as that the probability of $\boldsymbol{x}$ belonging to one class is $\mathbf{u}(\boldsymbol{x})$ and the probability in another class is $1 - \mathbf{u}(\boldsymbol{x})$. The advantage of the soft version over the

original Chan–Vese model is that the soft version (8) is convex with respect to **u** such that some efficient numerical method can be applied, while the Chan–Vese model is non-convex with respect to ϕ.

Both the Chan–Vese model (6) and its soft version (8) give impressive segmentation results for nearly piecewise constant images, which has been widely used in image segmentation. Since the involved L2-norm fidelity is based on the Gaussian distribution, (6) and (8) are robust to Gaussian noise. However, both models are not robust to non-Gaussian noise or outliers. In SAR images change detection, the difference image suffers from large extent of non-Gaussian noise, such that both (6) and (8) cannot achieve good enough segmentation results. On the other hand, the total variation regularization in (8) causes staircase effect in image processing problems (Li et al. 2007), which may affect the accuracy in image segmentation. In this work, we propose a variant of soft Chan–Vese model for image change detection to overcome the drawbacks and increase the segmentation accuracy.

3. The proposed method

In this section, we propose our method for change detection with two steps. In the first step, we generate a difference image as a combination of log-ratio image and mean-ratio image. In the second step, we propose an unsupervised soft segmentation method based on curvelet regularization and L1-norm fidelity.

3.1. *Generation of the difference image*

Given two multitemporal SAR images $\mathbf{X}_1,\mathbf{X}_2:\Omega\rightarrow[0,255]$, where $\Omega \subset \bullet^2$ is a rectangular area. Note that if the given data in not in the range [0,255], a rescaling step is needed. As mentioned in Section 1, SAR images are polluted by multiplicative speckle noise. With the logarithm operator, the multiplicative noise can be transformed into additive noise such that the following log-ratio difference image is widely used:

$$\left|\ln\frac{\mathbf{X}_2}{\mathbf{X}_1}\right| = |\ln\mathbf{X}_2 - \ln\mathbf{X}_1|. \tag{9}$$

To avoid the intensity to be zero, as in Zheng et al. (2014), we define

$$\mathbf{D}_{\mathrm{l}} = \left|\ln\frac{\mathbf{X}_2+1}{\mathbf{X}_1+1}\right|. \tag{10}$$

Because of the multiplicative nature of speckle noise, the classical approach consists in using the ratio of the local means in the neighbourhood of each pair of homologous pixels. Here 'local means' represents the result by applying mean filter on the image. The mean-ratio difference image is defined as (Inglada and Mercier 2007)

$$\mathbf{D}_{\mathrm{m}} = 1 - \min\left(\frac{\mathbf{M}_1}{\mathbf{M}_2},\frac{\mathbf{M}_2}{\mathbf{M}_1}\right), \tag{11}$$

where $\mathbf{M}_1$, $\mathbf{M}_2$ denote the local mean value of $\mathbf{X}_1$, $\boldsymbol{X}_2$, respectively. This difference image assumes that a change in the scene will appear as a modification of the local mean value

of the image. This measure is also robust to speckle noise. See Inglada and Mercier (2007) for more details.

Both of the two difference images have their own advantages and disadvantages (Rignot and van Zyl 1993; Gong, Zhou, and Ma 2012). The advantage of the log-ratio difference image is that the unchanged areas seem flat. However, the changed areas are weakened by logarithm operator such that it may detect fewer changed pixels than actually changed ones. In contrast, the advantage of mean-ratio difference image is that the changed areas seem to include almost all the real changed pixels. However, some unchanged areas may also be mistakenly detected as changed areas since the ratio operator may emphasize the differences of low intensities in the multitemporal images (e.g. 200/20 = 10, 20/2 = 10). To combine the advantages of log-ratio and mean-ratio and decrease their disadvantages, in this article, we propose to combine them by weighted average:

$$\mathbf{D} = 0.4\mathbf{D}_{\mathrm{m}} + 0.6\mathbf{D}_{\mathrm{l}}. \tag{12}$$

The weights 0.4 and 0.6 are chosen by trial and error. Experimental results show that this combined difference image works well in our proposed method. Note that we have rescaled $\mathbf{D}_1$ by a factor of 2 since the range of $\mathbf{D}_1$ is a little larger than $\mathbf{D}_{\mathrm{m}}$. Note that in Zheng et al. (2014) the combined difference image is also defined by a weighted average of two kinds of difference images, which are obtained by applying median filter on $\mathbf{D}_1$ and applying median filter on $\mathbf{D}_S = |\mathbf{X}_2 - \mathbf{X}_1|$. It is obvious that the proposed combined difference image is different from that in Zheng et al. (2014).

3.2. *Unsupervised soft segmentation*

As mentioned in Section2, the advantage of the soft Chan–Vese model (8) over the original Chan–Vese model (6) is the convexity with respect to $\mathbf{u}$, which gives more choices in numerical schemes. However, due to the speckle noise in SAR images, the difference image contains a large amount of non-Gaussian noise and outliers. In addition, the staircase effect of total variation regularization term causes the loss of segmentation accuracy. Hence model (8) is not good enough in dealing with segmentation of SAR difference image.

To overcome the above-mentioned two drawbacks of the soft Chan–Vese model, we propose a variant based on (8) as follows, named curvelet L1 model (C_L1):

$$\min_{0\le\mathbf{u}\le1,c_1,c_2}\left\{\begin{array}{l}\displaystyle\int_\Omega|\mathcal{C}\mathbf{u}|\mathrm{d}\boldsymbol{x}+\lambda_1\int_\Omega|\mathbf{I}-c_1|\mathbf{u}\mathrm{d}\boldsymbol{x}\\ \displaystyle+\lambda_1\lambda_2\int_\Omega|\mathbf{I}-c_2|(1-\mathbf{u})\mathrm{d}\boldsymbol{x}\end{array}\right\}, \tag{13}$$

where $\mathbf{u}$ is the soft membership function, $\mathbf{I} = \mathbf{D}$ is the difference image (12), $\lambda_1\lambda_2$ and λ_2 are fixed positive parameters, $\mathcal{C}$ denotes the curvelet transform as described in Section 2, and $\mathcal{C}\mathbf{u}$ denotes the curvelet transform coefficients of $\mathbf{u}$. Note that in the fidelity terms, we set the coefficients to be λ_1 and $\lambda_1\ \lambda_2$, respectively, in order to simplify the computation when deriving the algorithm in the next section, Section 3.3.1.

There are two novelties in the proposed model. On the one hand, curvelet regularization is used to replace the total variation regularization in (8), which gives nearly

optimal sparse representation of images with smooth regions and edges. On the other hand, L1-norm fidelity is used instead of L2-norm to measure the distance of the image intensity with the class centres c_1 and c_2. It is motivated by the fact that L1-norm is more robust to non-Gaussian noise and outliers, which is widely used in image restoration problem (Chan and Esedoglu 2005; Guo, Li, and Ng 2009; Jung, Kang, and Kang 2014).

3.3. *Numerical algorithm*

In this section, we introduce an effective algorithm to solve the proposed model based on the popular alternating split Bregman (ASB) method (Goldstein and Osher 2009). First we give a brief introduction on the ASB method. Assume that H and $|G|$ are convex functions and H is differentiable (here $|\cdot|$ is absolute value operator). Let us consider the problem:

$$\min_{\mathbf{u},\mathbf{d}} |\mathbf{d}| + H(\mathbf{u}) \, s.t. \, G(\mathbf{u}) = \mathbf{d}. \tag{14}$$

Let us introduce another variable $\mathbf{b}$ and define

$$F(\mathbf{u},\mathbf{d},\mathbf{b}) = |\mathbf{d}| + H(\mathbf{u}) + \frac{\mu}{2}\left\|G(\mathbf{u}) - \mathbf{d} + \mathbf{b}^k\right\|_2^2, \tag{15}$$

where μ is a positive parameter. In fact, $F(\mathbf{u},\mathbf{b},\mathbf{d})$ is equivalent to the augmented Lagrangian of the original constrained problem (14). Then the ASB algorithm for this problem is given by the following iteration scheme:

$$\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} F\left(\mathbf{u}, \mathbf{d}^k, \mathbf{b}^k\right),$$

$$\mathbf{d}^{k+1} = \arg\min_{\mathbf{d}} F\left(\mathbf{u}^{k+1}, \mathbf{d}, \mathbf{b}^k\right),$$

$$\mathbf{b}^{k+1} = \mathbf{b}^k + G\left(\mathbf{u}^{k+1}\right) - \mathbf{d}^{k+1}.$$

It has been proved that ASB is equivalent to the classical alternating direction method of multipliers (ADMM) and the convergence is proved in Boyd et al. (2011) under weak conditions.

In the proposed C_L1 model (13), the functional is non-differentiable with respect to variables $\mathbf{u}$, c_1, and c_2. As mentioned above, ASB can be used to deal with the non-differentiable terms by introducing new variables. In this way, three additional variables are needed corresponding to all the three L1-norm involved terms. Actually, we have a simpler approach to handle the two non-differentiable L1-norm fidelity terms. We rewrite them as weighted L2-norm terms and get the following approximate problem:

$$\min_{\mathbf{u},c_1,c_2} \left\{ \begin{array}{l} \int_\Omega |\mathcal{C}\mathbf{u}| \mathrm{d}\boldsymbol{x} + \lambda_1 \int_\Omega \mathbf{w}_1 |\mathbf{I} - c_1|^2 \mathbf{u} \mathrm{d}\boldsymbol{x} \\ +\lambda_1\lambda_2 \int_\Omega \mathbf{w}_2 |\mathbf{I} - c_2|^2 (1-\mathbf{u}) \mathrm{d}\boldsymbol{x} + \chi_U(\mathbf{u}) \end{array} \right\}, \tag{16}$$

where $\chi_U(\mathbf{u})$ is the indicator function of

$$U = \{\mathbf{u} : 0 \le \mathbf{u} \le 1\} \tag{17}$$

and the weights are defined as

$$\mathbf{w}_1 = \frac{1}{|\mathbf{I} - \bar{c}|_1}, \mathbf{w}_2 = \frac{1}{|\mathbf{I} - \bar{c}_2|} \tag{18}$$

in which $\bar{c}_1$ and $\bar{c}_2$ are estimated values of c_1 and c_2, respectively. In an iterative scheme, $\bar{c}_1$ and $\bar{c}_2$ can be chosen as the values of the last iteration.

Following ASB, for the non-differentiable curvelet regularization term, we introduce an auxiliary variable $\mathbf{d}$ and reformulate the above problem as a constrained problem:

$$\begin{aligned} &\min_{\mathbf{u},c_1,c_2} \left\{ \begin{array}{l} \int_\Omega |\mathbf{d}| d\boldsymbol{x} + \lambda_1 \int_\Omega \mathbf{w}_1 |\mathbf{I} - c_1|^2 \mathbf{u} d\boldsymbol{x} \\ +\lambda_1\lambda_2 \int_\Omega \mathbf{w}_2 |\mathbf{I} - c_2|^2 (1 - \mathbf{u}) d\boldsymbol{x} + \chi_U(\mathbf{u}) \end{array} \right\} \\ &\text{s.t.} \quad \mathcal{C}\mathbf{u} = \mathbf{d}. \end{aligned} \tag{19}$$

The corresponding augmented Lagrangian is equivalent to

$$\mathcal{L}(\mathbf{u}, c_1, c_2, \mathbf{d}, \mathbf{b}) = \left\{ \begin{array}{l} \int_\Omega |\mathbf{d}| d\boldsymbol{x} + \lambda_1 \int_\Omega \mathbf{w}_1 |\mathbf{I} - c_1|^2 \mathbf{u} d\boldsymbol{x} \\ +\lambda_1\lambda_2 \int_\Omega \mathbf{w}_2 |\mathbf{I} - c_2|^2 (1 - \mathbf{u}) d\boldsymbol{x} \\ +\frac{\mu}{2} \int_\Omega |\mathcal{C}\mathbf{u} - \mathbf{d} + b|^2 d\boldsymbol{x} + \chi_U(\mathbf{u}) \end{array} \right\}, \tag{20}$$

where $\mathbf{b}$ is the dual variable. By utilizing ASB technique and alternating minimization method, we get the following iteration scheme:

$$c_1^{k+1} = \arg\min_{c_1} \mathcal{L}\left(\mathbf{u}^k, c_1, c_2^k, \mathbf{d}^k, \mathbf{b}^k\right), \tag{21}$$

$$c_2^{k+1} = \arg\min_{c_2} \mathcal{L}\left(\mathbf{u}^k, c_1^{k+1}, c_2, \mathbf{d}^k, \mathbf{b}^k\right), \tag{22}$$

$$\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \mathcal{L}\left(\mathbf{u}, c_1^{k+1}, c_2^{k+1}, \mathbf{d}^k, \mathbf{b}^k\right), \tag{23}$$

$$\mathbf{d}^{k+1} = \arg\min_{\mathbf{d}} \mathcal{L}\left(\mathbf{u}^{k+1}, c_1^{k+1}, c_2^{k+1}, \mathbf{d}, \mathbf{b}^k\right), \tag{24}$$

$$\mathbf{b}^{k+1} = \mathbf{b}^k + C\mathbf{u}^{k+1} - \mathbf{d}^{k+1}. \tag{25}$$

In the following, we solve each sub-problem (9)–(12) in detail. For simplicity, we omit the iteration index in the derivation.

3.3.1. *Solving c_1 and c_2*

Since the sub-problems (9) and (10) for c_1 and c_2 are separable, we consider them together, that is

$$\min_{c_1,c_2} \int_\Omega \mathbf{w}_1 |\mathbf{I} - c_1|^2 \mathbf{u} d\boldsymbol{x} + \lambda_2 \int_\Omega \mathbf{w}_2 |\mathbf{I} - c_2|^2 (1 - \mathbf{u}) d\boldsymbol{x}. \tag{26}$$

The first-order optimality conditions for c_1 and c_2 are

$$\int_{\Omega} \mathbf{w}_1(\mathbf{I} - c_1)\mathbf{u}d\boldsymbol{x} = 0 \tag{27}$$

and

$$\int_{\Omega} \mathbf{w}_2(\mathbf{I} - c_2)\mathbf{u}d\boldsymbol{x} = 0, \tag{28}$$

respectively. Hence it is easy to get the closed-form solutions of c_1 and c_2 as follows:

$$c_1 = \frac{\int_{\Omega} \mathbf{w}_1\mathbf{I}\mathbf{u}d\boldsymbol{x}}{\int_{\Omega} \mathbf{w}_1\mathbf{u}d\boldsymbol{x}}, c_2 = \frac{\int_{\Omega} \mathbf{w}_2\mathbf{I}(1-\mathbf{u})d\boldsymbol{x}}{\int_{\Omega} \mathbf{w}_2(1-\mathbf{u})d\boldsymbol{x}}. \tag{29}$$

3.3.2. *Solving* u

The sub-problem (11) for **u** is

$$\min_{0\le\mathbf{u}\le 1} \lambda_1 \int_{\Omega} \mathbf{r}\mathbf{u}d\boldsymbol{x} + \frac{\mu}{2}\int_{\Omega} |\mathcal{C}\mathbf{u} - \mathbf{d} + \mathbf{b}|^2 d\boldsymbol{x}, \tag{30}$$

where $\mathbf{r} := \mathbf{w}_1|I - c_1|^2 - \lambda_2\mathbf{w}_2|\mathbf{I} - c_2|^2$. Since both the functional and the constraint are convex, we can solve this problem exactly in two steps. First, we calculate the first order optimality condition for **u** without considering the constraint, which is

$$\lambda_1\mathbf{r} + \mu\mathcal{C}^{\mathrm{T}}(\mathcal{C}\mathbf{u} - \mathbf{d} + \mathbf{b}) = 0, \tag{31}$$

where C^{T} is the conjugate of C. Using the perfect construction property of curvelet in (5), we can easily deduce from the above equality that the closed-form solution of **u** is

$$\mathbf{u} = \mathcal{C}^{\mathrm{T}}(\mathbf{d} - \mathbf{b}) - \theta\mathbf{r}, \tag{32}$$

where $\theta = \lambda_1/\mu$. Then the convex constraint $0 \le u \le 1$ is applied and we get the solution of constrained problem (30) as

$$\mathbf{u} = \min\big(\max\big(\mathcal{C}^{\mathrm{T}}(\mathbf{d} - \mathbf{b}) - \theta\mathbf{r}, 0\big), 1\big). \tag{33}$$

3.3.3. *Solving* d

The sub-problem (12) for **d** is

$$\min_{\mathbf{d}} \int_{\Omega} |\mathbf{d}|d\boldsymbol{x} + \frac{\mu}{2}\int_{\Omega} |\mathcal{C}\mathbf{u} - \mathbf{d} + \mathbf{b}|^2 d\boldsymbol{x}. \tag{34}$$

Let $\tau = 1/\mu$. It is well known that the closed-form solution of this sub-problem can be expressed using the soft shrinkage (Goldstein and Osher 2009; Li and Zeng 2014), that is,

$$\mathbf{d} = \mathcal{S}(\mathcal{C}\mathbf{u} + \mathbf{b}, \tau), \tag{35}$$

where S is the soft shrinkage operator pointwise defined as

$$\mathcal{S}(v,\tau)=\begin{cases} v-\tau, & \text{if } v>\tau \\ v+\tau, & \text{if } v<-\tau \\ 0 & \text{otherwise} \end{cases}. \tag{36}$$

Note that this sub-problem may also seem to be a denoising problem with given data $\mathcal{C}\mathbf{u}+\mathbf{b}$. The thresholding τ controls the smoothness of solution $\mathbf{d}$, bigger means smoother.

Finally, combining the updating formulas (8), (29), (33), and (35), according to the framework of ASB, the detailed C_L1 algorithm is summarized in Algorithm C_L1 as follows:

- Initialization: $\mathbf{u}^0=\mathbf{I}^0, \mathbf{d}^0=\mathbf{0}$, $\mathbf{b}^0=\mathbf{0}, \mathbf{w}_1^0=\mathbf{1}, \mathbf{w}_2^0=\mathbf{1}$.
- For k = 0,1,2,..., repeat until stopping criterion is reached:

$$c_1^{k+1}=\frac{\int_\Omega \mathbf{w}_1^k\mathbf{I}\mathbf{u}^k d\boldsymbol{x}}{\int_\Omega \mathbf{w}_1^k\mathbf{u}^k d\boldsymbol{x}},\quad c_2^{k+1}=\frac{\int_\Omega \mathbf{w}_2^k\mathbf{I}(1-\mathbf{u}^k) d\boldsymbol{x}}{\int_\Omega \mathbf{w}_2^k(1-\mathbf{u}^k) d\boldsymbol{x}};$$

$$\mathbf{w}_1^{k+1}=\frac{1}{|\mathbf{I}-c_1^{k+1}|},\quad \mathbf{w}_2^{k+1}=\frac{1}{|\mathbf{I}-c_2^{k+1}|};$$

$$\mathbf{r}^{k+1}=\mathbf{w}_1^{k+1}|\mathbf{I}-c_1^{k+1}|^2-\lambda_2\mathbf{w}_2^{k+1}|\mathbf{I}-c_2^{k+1}|^2;$$

$$\mathbf{u}^{k+1}=\min\left(\max\left(\mathcal{C}^{\mathrm{T}}\left(\mathbf{d}^k-\mathbf{b}^k\right)-\theta\mathbf{r}^{k+1},0\right),1\right);$$

$$\mathbf{d}^{k+1}=\mathcal{S}\left(\mathcal{C}\mathbf{u}^{k+1}+\mathbf{b}^k,\tau\right);$$

$$\mathbf{b}^{k+1}=\mathbf{b}^k+\mathcal{C}\mathbf{u}^{k+1}-\mathbf{d}^{k+1};$$

- Output: $\mathbf{u}^{k+1}$.

Note that the initial membership function is set as $\mathbf{I}^\circ = \mathbf{I}/\max(\mathbf{I})$ valued in [0,1]. Here $\mathbf{0}$ denotes matrix with all entries equal to 0 and 1 denotes matrix with all entries equal to 1. The stopping criterion is defined as the distance of class centres in the successive iterations are small enough, that is,

$$\left(c_1^{k+1}-c_1^k\right)^2+\left(c_2^{k+1}-c_2^k\right)^2<\epsilon, \tag{37}$$

where ε is a very small number.

3.4. *Partial convergence analysis*

Theoretically, by fixing $\mathbf{w}_1$ and $\mathbf{w}_2$, we can prove the partial convergence result following the method in Zhang (2010). Let $\mathbf{Z}^*=\left(\mathbf{u}^*,c_1^*,c_2^*,\mathbf{d}^*,\mathbf{b}^*\right)$ be the Karush–Kuhn–Tucher

(KKT (Boyd and Vandenberghe 2009)) point of problem (19). Then $\mathbf{Z}^*$ satisfies the following KKT conditions:

$$\int_{\Omega} \mathbf{w}_1 \left(\mathbf{I} - c_1^*\right) \mathbf{u}^* \mathrm{d}\boldsymbol{x} = 0, \tag{38}$$

$$\int_{\Omega} \mathbf{w}_2 \left(\mathbf{I} - c_2^*\right) (1 - \mathbf{u}^*) \mathrm{d}\boldsymbol{x} = 0, \tag{39}$$

$$\theta \mathbf{r}^* + \mathcal{C}^{\mathrm{T}} \mathbf{b}^* + \partial \chi_U(\mathbf{u}^*) \ni 0, \tag{40}$$

$$\partial |\mathbf{d}^*| - \mu \mathbf{b}^* \ni 0, \tag{41}$$

$$\mathcal{C}\mathbf{u}^* - \mathbf{d}^* = 0, \tag{42}$$

where $\mathbf{r}^* = \left(\mathbf{w}_1 \left(\mathbf{I} - c_1^*\right)^2 + \lambda_2 \left(\mathbf{I} - c_2^*\right)^2\right)$. Denote $\mathbf{Z}^k = \left(\mathbf{u}^k, c_1^k, c_2^k, \mathbf{d}^k, \mathbf{b}^k\right)$ and $\mathbf{X}^k = \left(\mathbf{u}^k, c_1^k, c_2^k\right)$. Then we have the following partial convergence theorem.

Theorem 1. *Let $\{\mathbf{Z}^k\}_{k=1}^{\infty}$ be the sequence generated by the algorithm C_L1 for fixed w_1 and w_2 that satisfies the condition*

$$\lim_{k \to \infty} \left(\mathbf{Z}^{k+1} - \mathbf{Z}^k\right) = 0. \tag{43}$$

Then any accumulation point of $\{\mathbf{Z}^k\}_{k=1}^{\infty}$ is a KKT point of constrained problem (19). Consequently, any accumulation point of $\{\mathbf{X}^k\}_{k=1}^{\infty}$ is a KKT point of (16).

See the proof of Theorem 1 in Appendix. Based on Theorem 1, the following claim results immediately. Whenever $\{\mathbf{Z}^k\}_{k=1}^{\infty}$ converges, it converges to a KKT point of (19). Note that since the minimization problem (19) is non-convex, the KKT conditions are only necessary conditions to (19).

4. Numerical results and analysis

To demonstrate the effectiveness of the proposed SAR images change detection method, in this section we test three data sets and compare the performance of the following methods:

- FCM – Fuzzy c-mean clustering applied on the difference image (12);
- PCA_K – PCA and *k*-means based method in Celik (2009);
- CV – Soft Chan–Vese model applied on the difference image (12);
- UDWT_CV – UDWT and Chan–Vese model based method in Celik and Ma (2011);
- CDI_K – Combined difference image and *k*-means based method in Zheng et al. (2014);
- FLICM – Fuzzy local information c-means clustering applied on the difference image (12) (Gong, Zhou, and Ma 2012);
- C_L1 – Curvelet and L1-norm based change detection method proposed in this article.

For quantitative analysis of change detection results, we use the following performance measures as in Rosin and Ioannidis (2003); Viera et al. (2005):

- FP – False positives, i.e. number of unchanged pixels detected as changed;
- FN – False negatives, i.e. number of changed pixels detected as unchanged;
- TP – True positives, i.e. number of changed pixels detected as changed;
- TN – True negatives, i.e. number of unchanged pixels detected as unchanged;
- FCC – Percentage correct classification defined as

$$\mathrm{FCC} := \frac{(\mathrm{TP}) + (\mathrm{TN})}{(\mathrm{TP}) + (\mathrm{FP}) + (\mathrm{TN}) + (\mathrm{FN})}, \tag{44}$$

which has a value in the range [0,1] and higher is better;

- Kappa coefficient (Viera et al. 2005) – a measure of agreement between two raters who each classify N items into C mutually exclusive categories, which is defined as

$$\kappa = \frac{P(\mathrm{a}) - P(\mathrm{e})}{1 - P(\mathrm{e})}, \tag{45}$$

where $P(\mathrm{a})$ is the relative observed agreement among raters, and $P(\mathrm{e})$ is the hypothetical probability of chance agreement. If the raters are in complete agreement then $\kappa = 1$. If there is no agreement among the raters other than what would be expected by chance, $\kappa = 0$. See Viera et al. (2005) for more details.

In the comparison, we use the five measures FP, FN, FP + FN, FCC, and kappa coefficient. Note that FP + FN is the total number of misclassified pixels.

The parameters of the proposed method C_L1 are set by trial and error. In generating the difference image, we use 3×3 window to calculate the local mean of images. The default parameters of algorithm C_L1 is set as $\lambda_2 = 1.3$, $\tau = 0.02$, $\theta = 0.1$, $\varepsilon = 10^{-10}$ In particular, for the Bern data (see Figure 2 below), we set $\lambda_2 = 1.1, \quad \tau = 0.015$. Note that the thresholding τ controls the smoothness of $\mathbf{d}$ and thus the smoothness of the segmentation, bigger means smoother. Here λ_2 relates to the scale of two clusters. Since the Bern data is not much noisy and the changed area is relatively smaller than the other two test data (see Figures 1 and 3 below), relatively smaller λ_2 and τ yield better result. For the other five methods that are selected for comparison, we follow the original papers to set the parameters. Note that PCA_K, CDI_K, and UDWT_CV give the binary change detection map directly. However, for FCM, FLICM, and C_L1, the final binary change detection map is obtained by thresholding the output membership function $\mathbf{u}$ by 0.5.

4.1. *The test data sets*

In the experiments, we test three data sets of SAR images, which are cited from Gong, Zhou, and Ma (2012). Note that the exact geographical information such as scale bars and geographical coordinates is not provided in that paper.

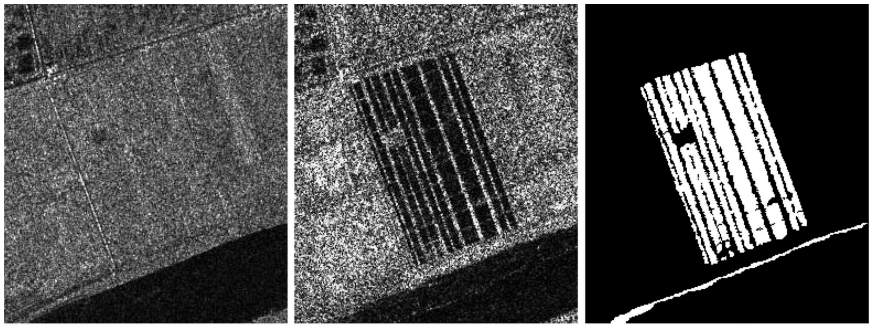


Figure 1. Multitemporal SAR images of the Yellow River Estuary. (*a*) Image acquired in June 2008; (*b*) image acquired in June 2009; (*c*) the ground truth of changed areas (white indicates changed areas and black indicates unchanged areas).

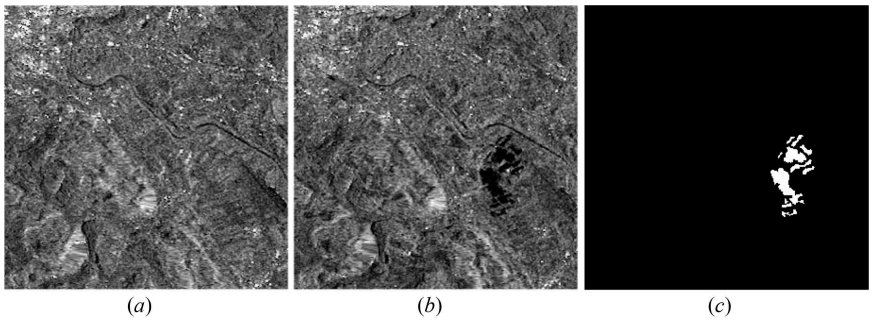


Figure 2. Multitemporal SAR images of the Bern city. (*a*) Image acquired in April 1999 before flooding; (*b*) image acquired in May 1999 after flooding; (*c*) the ground truth of changed areas.

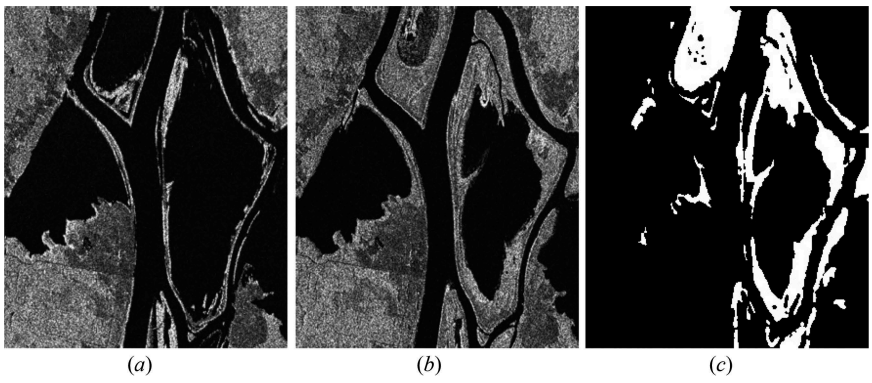


Figure 3. Multitemporal SAR images of the Ottawa city. (*a*) Image acquired in July 1997 during flooding; (*b*) image acquired in August 1997 after flooding; (*c*) the ground truth of changed areas.

The first data set displayed in Figure 1(*a*,*b*) is two SAR images acquired by Radarsat-2 at the regions of Yellow River Estuary in China in June 2008 and June 2009, respectively, with an area of 257 × 289 pixels. The two images Figure 1(*a*,*b*) are single-look image and

four-look image, respectively, such that the speckle noise in Figure 1(*a*) is much heavier than that in Figure 1(*b*). The huge difference level of noise makes change detection even harder. The ground truth (reference image) of change detection is shown in Figure 1(*c*) in which the changed areas are in white and unchanged areas are in black.

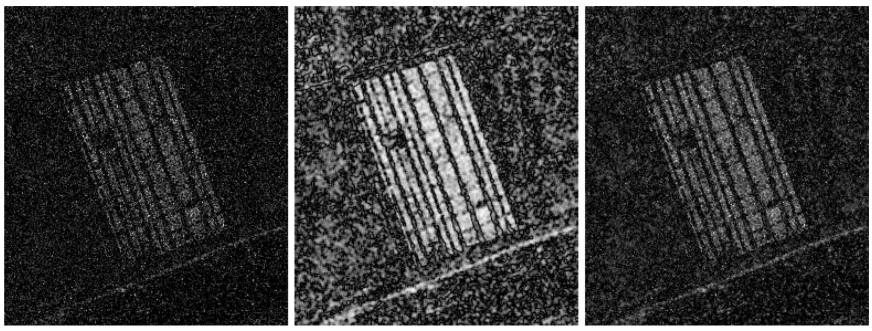
The second data set displayed in Figure 2(*a*,*b*) is two SAR images acquired by European Remote Sensing 2 satellite SAR sensor over an area near the city Bern, Switzerland, in April and May 1999 before and after flooding, respectively, with an area of 301 × 301 pixels. The Aare valley between Bern and Thun is selected as a test site for detecting flooded areas. The ground truth of changed areas is shown in Figure 2(*c*).

The third data set displayed in Figure 3(*a*,*b*) is two SAR images acquired by the Radarsat SAR sensor over the city of Ottawa, Canada, in July and August 1997 during and after summer flooding, respectively, with an area of 290 × 350 pixels. There are mainly two regions in the images, namely, land and water. The ground truth of changed areas is shown in Figure 3(*c*).

4.2. *Experimental results*

To illustrate the advantage of the proposed difference image, we display the log-ratio, mean-ratio, and the proposed weighted difference image in Figure 4(*a*–*c*), respectively. In the log-ratio difference image Figure 4(*a*), we observe that the unchanged areas are flatter than in the mean-ratio image Figure 4(*b*). On the other hand, the changed areas indicated by mean-ratio image seems better than log-ratio image in which the changed areas are much weakened by logarithm operator. By taking weighted average, the difference image in Figure 4(*c*) keeps a relatively flat unchanged area and an enhanced changed area, which makes it easier to classify unchanged and changed areas.

In Figures 5, 7 and 8 and Tables 1–3, we compare the performance of the proposed method C_L1 with five existing methods FCM, PCA_K, CV, UDWT_CV, CDI_K, and FLICM. Remark that FCM, CV, and FLICM are applied on the difference image (12). The bold-faced data in Tables 1–3 indicates the best performance.



(*a*) (*b*) (*c*)

Figure 4. The difference images of the Yellow River Estuary data. (*a*) log-ratio image; (*b*) mean-ratio image; (*c*) the difference image using the proposed method.

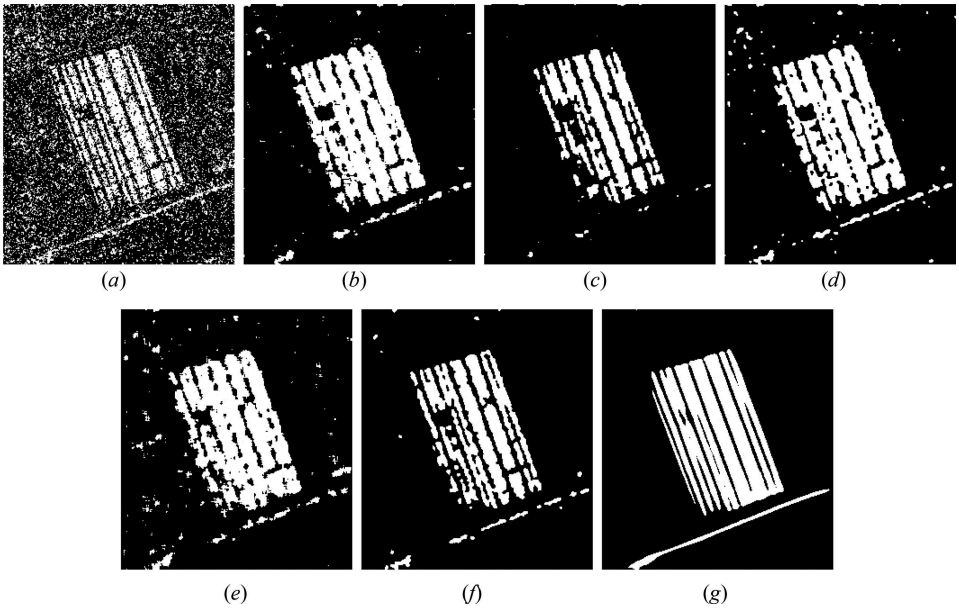


Figure 5. Change detection results of the Yellow River Estuary data obtained by different methods after (*a*) FCM; (*b*) PCA_K; (*c*) CV; (*d*) UDWT_CV; (*e*) CDI_K; (*f*) FLICM; (*g*) the proposed C_L1 method.

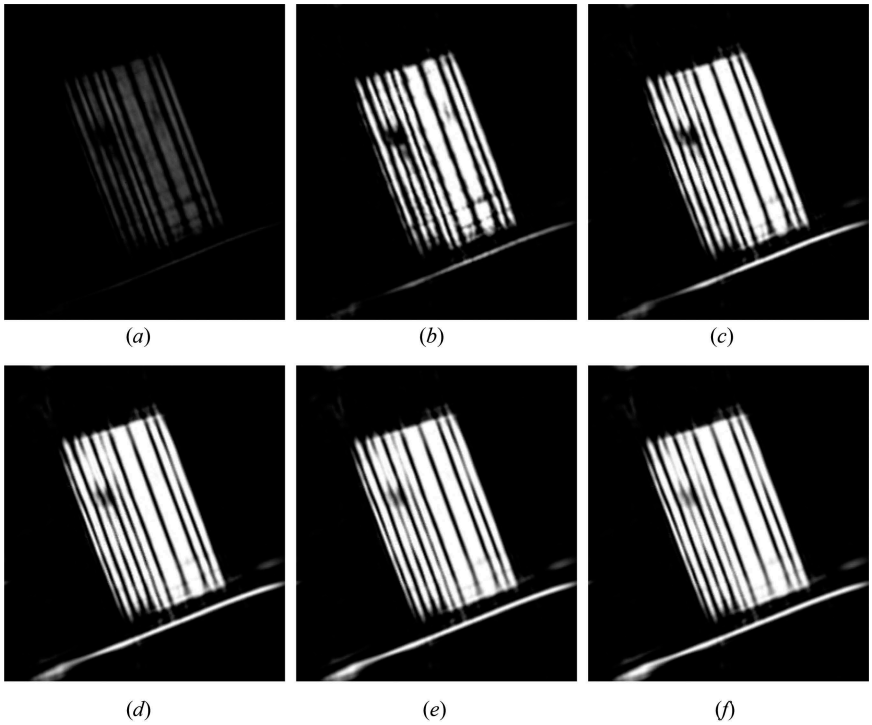


Figure 6. The soft membership function values **u** for the Yellow River Estuary data during the iteration process of Algorithm C_L1 after (*a*) 10 iterations; (*b*) 50 iterations; (*c*) 100 iterations; (*d*) 200 iterations; (*e*) 300 iterations; (*f*) 388 iterations.

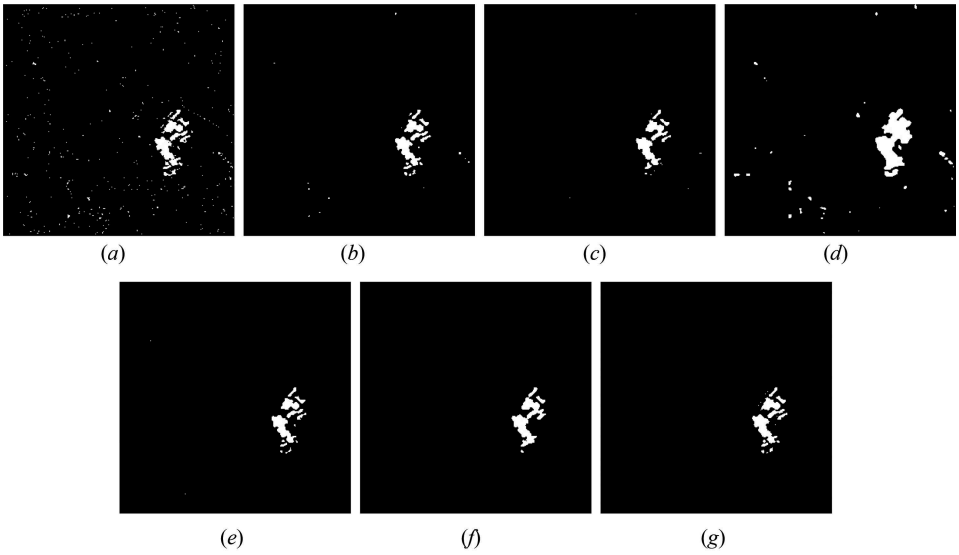


Figure 7. Change detection results of the Bern data by different methods after (*a*) FCM; (*b*) PCA_K; (*c*) CV; (*d*) UDWT_CV; (*e*) CDI_K; (*f*) FLICM; (*g*) the proposed C_L1 method.

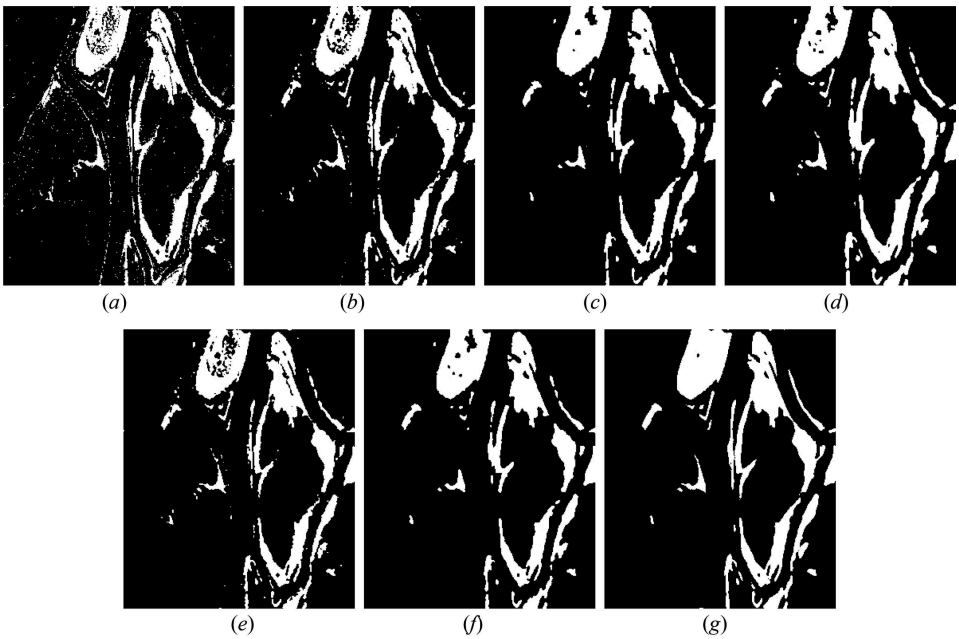


Figure 8. Change detection results of the Ottawa data by different methods after (*a*) FCM; (*b*) PCA_K; (*c*) CV; (*d*) UDWT_CV; (*e*) CDI_K; (*f*) FLICM; (*g*) the proposed C_L1 method.

Figure 5 and Table 1 show the results and performance measures on the Yellow River Estuary data, respectively. As mentioned before, the multitemporal images are suffered from different level of noise and the difference image is very noisy; see Figure 4. The detection of change is much more difficult in this case. The result of FCM in Figure 5(*a*)

Table 1. Summary of change detection performance of different methods for the Yellow River Estuary data set.

	FP	FN	FP + FN	PCC	Kappa coefficient
FCM	9822	3704	13,526	0.8179	0.4780
PCA_K	2963	1959	4922	0.9337	0.7827
CV	**310**	4559	4869	0.9344	0.7476
UDWT_CV	2561	1782	4343	0.9415	0.8070
CDI_K	3267	2516	5783	0.9221	0.7428
FLICM	577	3626	4203	0.9434	0.7905
C_L1	1657	**1143**	**2800**	**0.9623**	**0.8746**

Table 2. Summary of change detection performance of different methods for the Bern data set.

	FP	FN	FP + FN	PCC	Kappa coefficient
FCM	704	122	826	0.9909	0.7099
PCA_K	158	146	304	0.9966	0.8674
CV	87	219	306	0.9966	0.8758
UDWT_CV	1201	**9**	1210	0.9866	0.6485
CDI_K	**76**	209	285	0.9969	0.8675
FLICM	149	173	322	0.9964	0.8573
C_L1	108	165	**273**	**0.9970**	**0.8773**

Table 3. Summary of change detection performance of different methods for the Ottawa data set.

	FP	FN	FP + FN	PCC	Kappa coefficient
FCM	1485	1775	3260	0.9679	0.8785
PCA_K	583	1901	2484	0.9755	0.9049
CV	273	1810	2083	0.9795	0.9198
UDWT_CV	1196	1071	2267	0.9777	0.9164
CDI_K	425	2125	2550	0.9749	0.9014
FLICM	**257**	1653	1910	0.9812	0.9267
C_L1	772	**746**	**1518**	**0.9850**	**0.9439**

seems 'noisy' since many unchanged background area are detected as changed areas. The reason is that FCM clustering is completely based on the intensity and lacks spatial regularity. With some denoising steps or spatial regularity constraint, the results of PCA_K, CV, UDWT_CV, CDI_K, and FLICM in Figure 5(*b*–*f*) are much better than Figure 5(*a*). However, there are still many spots that appear in the background areas. The result of the proposed method in Figure 5(*g*) seems better than others since the unchanged background areas are very clean without any spots and the changed areas are better detected. The fine-structured changed areas are better detected by the proposed method. In particular, the edge continuity of the changed areas and curve-shaped changed areas are better preserved by the proposed method than others. For example, by observing carefully the narrow curve-shaped area in the bottom of the images, we find that Figure 5(*g*) detects a continuous changed area that is closer to the ground truth (Figure 1(*c*)), while the other methods have broken this area. Table 1 shows the performance measures corresponding to Figure 5. FCM performs worst, which has the largest number of misclassified pixels FP+FN (13526) and the lowest PCC (0.8179) and kappa coefficient (0.4780) measures among all. Our proposed method C_L1 performs best, which has the smallest number of misclassified pixels FP + FN (2800) and the highest PCC (0.9623) and kappa coefficient (0.8746) measures among all. For the other five methods, the misclassified pixels FP + FN are in the range 4203–5783, the PCC

measures are roughly in the range 0.9221–0.9434 and the kappa coefficients are in the range 0.7428–0.8070.

We remark that the edge continuity of C_L1 is benefited from curvelet regularization in the proposed model due to the fact that curvelet is nearly optimal in describing curves. Meanwhile, the robustness of the proposed method is enhanced by both the curvelet regularization and the L1-norm fidelity term.

To illustrate the behaviour of our proposed algorithm C_L1, we show the values of the soft membership function **u** for the Yellow River Estuary data during the iteration process in Figure 6. The algorithm stops at iteration 388. In Figure 6(*a*–*f*), we show the results of **u** for 10, 50, 100, 200, 300, and 388 iterations, respectively. From Figure 6 we observe that the proposed algorithm is very robust and converges. Note that thresholding Figure 6(*f*) by 0.5 gives the change detection map in Figure 5(*f*).

Figure 7 and Table 2 show the results and performance measures on the Bern data, respectively. The result of UDWT_CV in Figure 7(*d*) has many spots in the background areas and over-smooths the changed areas such that it has the largest number of misclassified pixels FP+FN (1210), lowest PCC (0.9866) and lowest kappa coefficient (0.6485) values. The result of FCM in Figure 6(*a*) also contains many little spots in the background, which has better performance than UDWT_CV with FP+FN 826, PCC 0.9909, and kappa coefficient 0.7099. The other five methods provide much better results and are competitive, which have misclassified pixels in the range 273–322, PCC measures in the range 0.9964–0.9970, and kappa coefficient in the range 0.8573–0.8773. Among them, PCA_K contains a few little spots in the background areas and FLICM leads to a little over-smoothing of the changed areas; see Figure 7(*c*,*f*), respectively. The result of the proposed C_L1 in Figure 7(*g*) is the best among all with the least misclassified pixels FP+FNN (273), the highest PCC (0.9970), and kappa coefficient (0.8773) measures. We can also observe that the fine-structured changed areas are better detected by the proposed methods than by others.

The change detection results and measures of the Ottawa data are displayed in Figure 8 and Table 3, respectively. The result of FCM in Figure 8(*a*) is somewhat noisy and leads to lowest quality measures among all (FP+FN = 3260, PCC = 0.9679, kappa coefficient = 0.8785). PCA_K and CDI_K give similar results in Figure 8(*b*,*e*), respectively, with similar FCC and kappa coefficient measures. CV, UDWT_CV, and FLICM lead to similar results as shown in Figure 8(*c*,*d*,*f*), respectively, which are smoother than the results of FCM, PCA_K, and CDI_K and have higher performance measures. Among all, the proposed C_L1 achieves the best result in Figure 8(*g*) with the best visual quality and highest performance measures (FP +FN = 1518, PCC = 0.9850, kappa coefficient = 0.9439). Again, we observe that C_L1 is good at detecting fine-structured change areas. In particular, it can preserve edges very well and can detect narrow curve-shaped changed areas very well without breaking them. See, for instance, the curve-shaped changed area in the right top part of the image.

In terms of computational time, the proposed method C_L1 and UDWT_CV are more time consuming than others since the segmentation part is more complicated in modelling and algorithm. We report the iterations and computational time in each test of our method. The Yellow River Estuary data: number of iterations = 388, time = 100.92 s; the Bern data: number of iterations = 101, time = 28.71 s; the Ottawa data: number of iterations = 164, time = 50.66 s. Actually, the computational time can be shortened if the stopping criterion is properly adjusted. For instance, we can set

maximum iteration as 100 or decrease ε in (37), which also gives good results. All the experiments are performed under Windows 8 and MATLAB R2012a with Intel Core i7-4500 CPU@1.80 GHz and 8 GB memory.

5. Concluding remarks

In this article, we have presented a novel change detection method for SAR images based on a weighted difference image and a new variational soft segmentation method in which curvelet regularization and L1-norm data term is used. Since curvelets are good at representing images with both smooth region and edges, the proposed method performs pretty well at preserving edge continuity and detecting curve-shaped structures. Compared with L2-norm fidelity term, which is widely used in *k*-means or FCM methods, L1-norm fidelity is more robust to non-Gaussian-type noise and outliers. Benefited from both the curvelet regularization and L1-norm fidelity, the proposed method is robust to the noise presented in the SAR difference image. Numerically, we build an effective iterative algorithm to solve the variational model based on the split Bregman technique. The experimental results show that the proposed method outperforms some existing methods and is quite promising in change detection.

In the future work, we will study the problem of automatic selection of parameters and speed up the proposed algorithm by including some parallel computing techniques. We will also consider generalization of the variational segmentation model to multiphase image segmentation problem, that is, the number of regions is larger than two.

Acknowledgments

This work is supported by the 973 Program (2011CB707104), the Science and Technology Commission of Shanghai Municipality (STCSM) (13dz2260400), the National Science Foundation of China (11271049, 61501188, 61372147), and the Science Foundation of Shanghai (15ZR1410200).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the 973 Program [grant number 2011CB707104], the Science and Technology Commission of Shanghai Municipality (STCSM) [grant number 13dz2260400]; the National Science Foundation of China [grant number 11271049], [grant number 61501188], [grant number 61372147]; and the Science Foundation of Shanghai [grant number 15ZR1410200].

References

Bovolo, F., and L. Bruzzone. 2005. "A Detail-Preserving Scale-Driven Approach to Change Detection in Multitemporal SAR Images." *IEEE Transactions on Geoscience and Remote Sensing* 43 (12): 2963–2972. doi:10.1109/TGRS.2005.857987.

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein. 2011. "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers." *Foundations and Trends in Machine Learning* 3 (1): 1–122. doi:10.1561/2200000016.

Boyd, S., and L. Vandenberghe. 2009. *Convex Optimization*. Cambridge: Cambridge University Press.

Bresson, X., S. Esedoğlu, P. Vandergheynst, J.-P. Thiran, and S. Osher. 2007. "Fast Global Minimization of the Active Contour/Snake Model." *Journal of Mathematical Imaging and Vision* 28 (2): 151–167. doi:10.1007/s10851-007-0002-0.

Bruzzone, L., and D. F. Prieto. 2000. "Automatic Analysis of the Difference Image for Unsupervised Change Detection." *IEEE Transactions on Geoscience and Remote Sensing* 38 (3): 1171–1182. doi:10.1109/36.843009.

Bujor, F., E. Trouvé, L. Valet, J.-M. Nicolas, and J.-P. Rudant. 2004. "Application of Log-Cumulants to the Detection of Spatiotemporal Discontinuities in Multitemporal SAR Images." *IEEE Transactions on Geoscience and Remote Sensing* 42 (10): 2073–2084. doi:10.1109/TGRS.2004.835304.

Candes, E., L. Demanet, D. Donoho, and L. Ying. 2006. "Fast Discrete Curvelet Transforms." *Multiscale Modeling & Simulation* 5 (3): 861–899. doi:10.1137/05064182X.

Candes, E. J., and D. L. Donoho. 2004. "New Tight Frames of Curvelets and Optimal Representations of Objects with Piecewise C2 Singularities." *Communications on Pure and Applied Mathematics* 57 (2): 219–266. doi:10.1002/cpa.10116.

Celik, T. 2009. "Unsupervised Change Detection in Satellite Images Using Principal Component Analysis And-Means Clustering." *IEEE Geoscience and Remote Sensing Letters* 6 (4): 772–776. doi:10.1109/LGRS.2009.2025059.

Celik, T., and K.-K. Ma. 2011. "Multitemporal Image Change Detection Using Undecimated Discrete Wavelet Transform and Active Contours." *IEEE Transactions on Geoscience and Remote Sensing* 49 (2): 706–716. doi:10.1109/TGRS.2010.2066979.

Celik, T., and K.-K. Ma. 2010. "Unsupervised Change Detection for Satellite Images Using Dual-Tree Complex Wavelet Transform." *IEEE Transactions on Geoscience and Remote Sensing* 48 (3): 1199–1210. doi:10.1109/TGRS.2009.2029095.

Chan, T. F., and S. Esedoglu. 2005. "Aspects of Total Variation Regularized L 1 Function Approximation." *SIAM Journal on Applied Mathematics* 65 (5): 1817–1837. doi:10.1137/040604297.

Chan, T. F., S. Esedoglu, and M. Nikolova. 2006. "Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models." *SIAM Journal on Applied Mathematics* 66 (5): 1632–1648. doi:10.1137/040615286.

Chan, T. F., and L. A. Vese. 2001. "Active Contours without Edges." *IEEE Transactions on Image Processing* 10 (2): 266–277. doi:10.1109/83.902291.

Goldstein, T., and S. Osher. 2009. "The Split Bregman Method for L1-Regularized Problems." *SIAM Journal on Imaging Sciences* 2 (2): 323–343. doi:10.1137/080725891.

Gong, M., L. Su, M. Jia, and W. Chen. 2014. "Fuzzy Clustering with a Modified MRF Energy Function for Change Detection in Synthetic Aperture Radar Images." *IEEE Transactions on Fuzzy Systems* 22 (1): 98–109. doi:10.1109/TFUZZ.2013.2249072.

Gong, M., Z. Zhou, and J. Ma. 2012. "Change Detection in Synthetic Aperture Radar Images Based on Image Fusion and Fuzzy Clustering." *IEEE Transactions on Image Processing* 21 (4): 2141–2151. doi:10.1109/TIP.2011.2170702.

Guo, X., F. Li, and M. K. Ng. 2009. "A Fast Âď$ 1-TV Algorithm for Image Restoration." *SIAM Journal on Scientific Computing* 31 (3): 2322–2341. doi:10.1137/080724435.

Inglada, J., and G. Mercier. 2007. "A New Statistical Similarity Measure for Change Detection in Multitemporal SAR Images and its Extension to Multiscale Change Analysis." *IEEE Transactions on Geoscience and Remote Sensing* 45 (5): 1432–1445. doi:10.1109/TGRS.2007.893568.

Jung, M., M. Kang, and M. Kang. 2014. "Variational Image Segmentation Models Involving Non-Smooth Data-Fidelity Terms." *Journal of Scientific Computing* 59 (2): 277–308. doi:10.1007/s10915-013-9766-0.

Krinidis, S., and V. Chatzis. 2010. "A Robust Fuzzy Local Information C-Means Clustering Algorithm." *IEEE Transactions on Image Processing* 19 (5): 1328–1337. doi:10.1109/TIP.2010.2040763.

Leprince, S., S. Barbot, F. Ayoub, and J.-P. Avouac. 2007. "Automatic and Precise Orthorectification, Coregistration, and Subpixel Correlation of Satellite Images, Application to Ground Deformation

Measurements." *IEEE Transactions on Geoscience and Remote Sensing* 45 (6): 1529–1558. doi:10.1109/TGRS.2006.888937.
Li, F., C. Shen, J. Fan, and C. Shen. 2007. "Image Restoration Combining a Total Variational Filter and a Fourth-Order Filter." *Journal of Visual Communication and Image Representation* 18 (4): 322–330. doi:10.1016/j.jvcir.2007.04.005.
Li, F., and T. Zeng. 2014. "A Universal Variational Framework for Sparsity Based Image Inpainting." *IEEE Transactions on Image Processing* 23 (10): 4242–4254. doi:10.1109/TIP.2014.2346030.
Moser, G., and S. B. Serpico. 2009. "Unsupervised Change Detection from Multichannel SAR Data by Markovian Data Fusion." *IEEE Transactions on Geoscience and Remote Sensing* 47 (7): 2114–2128. doi:10.1109/TGRS.2009.2012407.
Radke, R. J., S. Andra, O. Al-Kofahi, and B. Roysam. 2005. "Image Change Detection Algorithms: A Systematic Survey." *IEEE Transactions on Image Processing* 14 (3): 294–307. doi:10.1109/TIP.2004.838698.
Rignot, E. J. M., and J. J. van Zyl. 1993. "Change Detection Techniques for ERS-1 SAR Data." *IEEE Transactions on Geoscience and Remote Sensing* 31 (4): 896–906. doi:10.1109/36.239913.
Rosin, P. L., and E. Ioannidis. 2003. "Evaluation of Global Image Thresholding for Change Detection." *Pattern Recognition Letters* 24 (14): 2345–2356. doi:10.1016/S0167-8655(03)00060-6.
Schmitt, A., A. Wendleder, and S. Hinz. 2015. "The Kennaugh Element Framework for Multi-Scale, Multi-Polarized, Multi-Temporal and Multi-Frequency SAR Image Preparation." *ISPRS Journal of Photogrammetry and Remote Sensing* 102: 122–139. doi:10.1016/j.isprsjprs.2015.01.007.
Schmitt, A., B. Wessel, and A. Roth. 2009. "Curvelet Approach for SAR Image Denoising, Structure Enhancement, and Change Detection." *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (3/W4): 151–156.
Schmitt, A., B. Wessel, and A. Roth. 2014. "An Innovative Curvelet-Only-Based Approach for Automated Change Detection in Multi-Temporal SAR Imagery." *Remote Sensing* 6 (3): 2435–2462. doi:10.3390/rs6032435.
Viera, A. J., and J. M. Garrett. 2005. "Understanding Interobserver Agreement: The Kappa Statistic." *Family Medicine* 37 (5): 360–363.
Zhang, X., J. Chen, and H. Meng. 2013. "A Novel SAR Image Change Detection Based on Graph-Cut and Generalized Gaussian Model." *IEEE Geoscience and Remote Sensing Letters* 10 (1): 14–18. doi:10.1109/LGRS.2012.2189867.
Zhang, Y. 2010. *An Alternating Direction Algorithm for Nonnegative Matrix Factorization*. Technical report. Advance online publication. http://www.caam.rice.edu/~zhang/reports/tr1003.pdf
Zheng, Y., X. Zhang, B. Hou, and G. Liu. 2014. "Using Combined Difference Image and-Means Clustering for SAR Image Change Detection." *IEEE Geoscience and Remote Sensing Letters* 11 (3): 691–695. doi:10.1109/LGRS.2013.2275738.

Appendix

Proof of Theorem 1

According to (9)–(13), similar to the derivation of solutions for these sub-problems, we have

$$c_1^{k+1} - c_1^k = \frac{\int_\Omega \mathbf{w}_1 \mathbf{l} \mathbf{u}^k \mathrm{d}\boldsymbol{x}}{\int_\Omega \mathbf{w}_1 \mathbf{u}^k \mathrm{d}\boldsymbol{x}} - c_1^k, \tag{46}$$

$$c_2^{k+1} - c_2^k = \frac{\int_\Omega \mathbf{w}_2 \mathbf{I}\left(1 - \mathbf{u}^k\right) \mathrm{d}\boldsymbol{x}}{\int_\Omega \mathbf{w}_2 (1 - \mathbf{u}^k) \mathrm{d}\boldsymbol{x}} - c_2^k, \tag{47}$$

$$\mathbf{u}^{k+1} - \mathbf{u}^k \in \mathcal{C}^{\mathsf{T}}\left(\mathbf{d}^k - \mathbf{b}^k\right) - \theta \mathbf{r}^{k+1} - \partial \chi_U\left(\mathbf{u}^{k+1}\right) - \mathbf{u}^k, \tag{48}$$

$$\mathbf{d}^{k+1} - \mathbf{d}^k = \mathcal{S}\left(\mathcal{C}\mathbf{u}^{k+1} + \mathbf{b}^k, 1/\mu\right) - \mathbf{d}^k, \tag{49}$$

$$\mathbf{b}^{k+1} - \mathbf{b}^k = \mathcal{C}\mathbf{u}^{k+1} - \mathbf{d}^{k+1}, \tag{50}$$

where $\mathbf{r}^{k+1} = \mathbf{w}_1\left(\mathbf{I} - c_1^{k+1}\right)^2 - \lambda_2\left(\mathbf{I} - c_2^{k+1}\right)^2$.

Since $\lim_{k\to\infty}\left(\mathbf{Z}^{k+1} - \mathbf{Z}^k\right) = 0$, the right-hand side of each equality in (25)–(29) goes to zero as $k\to\infty$. Assume the accumulation point of $\{\mathbf{Z}^k\}$ is $\left(c_1^*, c_2^*, \mathbf{u}^*, \mathbf{d}^*, \mathbf{b}^*\right)$. Thus by letting $k\to\infty$, we obtain

$$\frac{\int_\Omega \mathbf{w}_1 \mathbf{I}\mathbf{u}^* \mathrm{d}\boldsymbol{x}}{\int_\Omega \mathbf{w}_1 \mathbf{u}^* \mathrm{d}\boldsymbol{x}} - c_1^* = 0, \tag{51}$$

$$\frac{\int_\Omega \mathbf{w}_2 \mathbf{I}(1 - \mathbf{u}^*) \mathrm{d}\boldsymbol{x}}{\int_\Omega \mathbf{w}_2 (1 - \mathbf{u}^*) \mathrm{d}\boldsymbol{x}} - c_2^* = 0, \tag{52}$$

$$-\mathcal{C}^{\mathsf{T}}(\mathbf{d}^* - \mathbf{b}^*) + \theta \mathbf{r}^* + \partial \chi_U(\mathbf{u}^*) + \mathbf{u}^* \ni 0, \tag{53}$$

$$\mathcal{S}(\mathcal{C}\mathbf{u}^* + \mathbf{b}^*, 1/\mu) - \mathbf{d}^* = 0, \tag{54}$$

$$\mathcal{C}\mathbf{u}^* - \mathbf{d}^* = 0. \tag{55}$$

It is obvious that (30) and (31) are equivalent to (19) and (20), respectively. From (34), we have $\mathbf{u}^* = \mathcal{C}^{\mathsf{T}}\mathbf{d}^*$. Substituting this equality into (32), we get (21). Since (33) is equivalent to

$$\mathbf{d}^* = \arg\min \int_\Omega |\mathbf{d}| \mathrm{d}\boldsymbol{x} + \frac{\mu}{2}\int_\Omega (\mathcal{C}\mathbf{u}^* + \mathbf{b}^* - \mathbf{d})^2 \mathrm{d}\boldsymbol{x}, \tag{56}$$

we have $0 \in \partial|\mathbf{d}^*| - \mu(\mathcal{C}\mathbf{u}^* + \mathbf{b}^* - \mathbf{d}^*)$. By using (34), we can derive (22). Therefore, the accumulation point of $\{\mathbf{Z}^k\}$ satisfies the KKT condition (19)-(23).

We have proved the statement concerning the sequence $\{\mathbf{Z}^k\}_{k=1}^\infty$ and problem (19). The statement concerning the sequence $\{\mathbf{X}^k\}_{k=1}^\infty$ and problem (16) follows directly from the equivalence between the two problems. This completes the proof.