

UNSUPERVISED CHANGE DETECTION IN MULTITEMPORAL MULTISPECTRAL SATELLITE IMAGES: A CONVEX RELAXATION APPROACH

Wei-Cheng Zheng[†], Chia-Hsiang Lin^{*}, Kuo-Hsin Tseng[†], Chih-Yuan Huang[†], Tang-Huang Lin[†],

Chia-Hsiang Wang[†], and Chong-Yung Chi[‡]

[†]Center for Space and Remote Sensing Research, National Central University, Taiwan

^{*}Department of Electrical Engineering, National Cheng Kung University, Taiwan

[‡]Department of Electrical Engineering, National Tsing Hua University, Taiwan

E-mail: 107022004@cc.ncu.edu.tw; chiahsiang.steven.lin@gmail.com;

{khtseng, cyhuang, thlin}@csrsr.ncu.edu.tw; verna.w@gmail.com; cychi@ee.nthu.edu.tw

ABSTRACT

Change detection (CD), enabled by multitemporal multispectral satellite imagery, has many important Earth observation missions such as land cover/use monitoring, for which we observe that change regions are relatively smaller than those caused by disaster (e.g., forest fire) with patterns typically composed of a number of smooth regions. These observations are considered in our new CD criterion, which can effectively mitigate the artifacts and speckle noise suffered by existing statistic-based and *difference image* (DI) analysis based methods. The proposed CD criterion amounts to a large-scale non-convex optimization, which is first reformulated using the convex relaxation trick with associated change map interpreted in the probability sense, followed by adopting an efficient convex solver known as alternating direction method of multipliers (ADMM). The resulted probabilistic change map would be more practical, and can be thresholded at 0.5 to yield the conventional binary-valued one. We also reveal a link between the proposed criterion and the DI-based criterion, and demonstrate the outstanding performance of our fully unsupervised CD algorithm qualitatively and quantitatively.

Index Terms— Change detection, multitemporal imagery, multispectral imagery, convex relaxation, alternating direction method of multipliers.

1. INTRODUCTION

Change detection (CD), playing an important role in natural resource management and monitoring, is enabled by the multitemporal multispectral satellite imagery. Specifically, given two multispectral images, represented as $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_L] \in \mathbb{R}^{M \times L}$ and $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_L] \in \mathbb{R}^{M \times L}$ (the two-dimensional representation of multi-band images [1]), covering the same spatial area but acquired at different time instances, the aim is to unsupervisedly detect the changes between the two images, where L is the number of pixels and M is the number of spectral bands. CD techniques are mainly categorized into two classes, one for disaster mapping and another for land cover/use monitoring (LCUM) [2], where the latter will be the focus of this paper. In comparison with changes caused by disaster (e.g., forest fire), the changes in LCUM application are in general subject to relatively small regions (cf. (A1) in Section 2).

Most CD methods are based on analyzing the so-called *difference image* (DI) $\mathbf{Z} \triangleq \mathbf{X} - \mathbf{Y} \in \mathbb{R}^{M \times L}$ [3], whose ℓ 's pixel is

$\mathbf{z}_\ell \triangleq \mathbf{x}_\ell - \mathbf{y}_\ell \in \mathbb{R}^M$. A straightforward DI-based way is to detect the changed pixels as those with significant differences; precisely, the changes $\mathbf{c} = [c_1, \dots, c_L]^T \in \{0, 1\}^L$ can be detected by

$$c_\ell = \begin{cases} 1, & \text{if } \|\mathbf{z}_\ell\|_2 > \zeta, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where $\zeta > 0$ is a threshold, and $c_\ell = 1$ means that the ℓ th pixel is changed (and $c_\ell = 0$ means otherwise). Some more sophisticated DI-based approaches employ principal component analysis (PCA), including clustering PCA (clPCA) [3] and iterative PCA (itPCA) [4]. However, these methods quite easily suffer from the interference of speckle noise scattered everywhere, as can be seen from Figures 1(e) and 1(f). This motivates us to further assume that the change pattern in LCUM would be composed of a number of smooth regions (cf. (A2) in Section 2).

Another powerful statistic-based method is Bayesian Significant Zero for Image (iBSiZer) [5], which, as far as we know, yields state-of-the-art CD performance and will serve as our key competitor. The speckle noise is significantly reduced by iBSiZer, but there are still some artifacts in the change map; cf. Figure 1(d). In this paper, we design a novel CD criterion to account for both aforementioned assumptions. Such criterion induces a large-scale non-convex optimization, which is reformulated into a convex problem using convex relaxation [6,7] that yields a probabilistic change map. That said, the value of c_ℓ is no longer binary but in the interval $[0, 1]$, interpreted as the change probability of the ℓ th pixel. This would be more practical than conventional binary-valued CD methods because, for pixels not easy to be classified, our method just suggests a value $c_\ell \in [0, 1]$ to indicate the probability, which can be efficiently computed by the alternating direction method of multipliers (ADMM) [7]; when a binary map is preferred, one can threshold the probability map at 0.5 in the maximum-likelihood sense, which is quite effective as will be seen in terms of several quantitative CD performance indices. We also reveal a link between the proposed criterion and the DI-based criterion, and demonstrate the outstanding performance of our fully unsupervised CD algorithm using real satellite images acquired by *Satellite Pour l'Observation de la Terre* (SPOT-7) [8].

Notation: $\mathbf{Diag}(\mathbf{v})$ is a diagonal matrix whose i th diagonal entry is $[\mathbf{v}]_i$ (i.e., the i th entry of vector \mathbf{v}). $\mathbf{diag}(\mathbf{M})$ is a column vector whose i th entry is $[\mathbf{M}]_{(i,i)}$; here, $[\mathbf{M}]_{(i,j)}$ denotes the (i,j) th entry of matrix \mathbf{M} . $\text{conv}(\cdot)$ denotes convex hull. $\text{vec}(\mathbf{M})$ is the vectorization operator, and $\text{vec}_{m \times n}^{-1}(\mathbf{v})$ is the $m \times n$ matrix \mathbf{M} satisfying $\text{vec}(\mathbf{M}) = \mathbf{v}$. \odot is the Hadamard product. $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_F$ are the ℓ_1 -norm, Euclidean norm and Frobenius norm, respectively

This work is supported by the Young Scholar Fellowship Program (Einstein Program) of Ministry of Science and Technology (MOST) in Taiwan, under Grant MOST107-2636-E-006-006.

(resp.). $\mathbf{1}_N$ and $\mathbf{0}_N$ denote the all-one and all-zero N -vectors, resp. $\mathcal{I}_Z \triangleq \{1, \dots, Z\}$ (Z is positive integer).

2. CD CRITERION DESIGN

As discussed above, we adopt the following assumptions:

(A1) the changes are subject to relatively small regions;

(A2) the change pattern is piecewise smooth.

These assumptions will be rigorously formulated later. For now, we further illustrate them using real satellite imagery. In Figure 1(a) (resp., Figure 1(b)), we show false color composition (NIR-R-G) of SPOT-7 image \mathbf{X} (resp., \mathbf{Y}) acquired over Hsinchu County region of Taiwan, on March 24, 2017 (resp., May 24, 2018). This image pair illustrates a challenging scenario for CD study because they were acquired under different illumination conditions. This area (coordinate of upper-left pixel: latitude $24^\circ 50' 4.08''$ N; longitude $121^\circ 6' 3.74''$ E) had been mostly covered by forest, but later some of its subareas incurred deforestation for other purpose of land use; those subareas are marked by white color in Figure 1(c), from which one can see that (A1)-(A2) do well characterize those changes. We acknowledge that these assumptions are particularly made for LCUM (anthropogenic factor), and may not hold for changes caused by disaster (natural factor). Effective CD criterion/algorithm, aided by suitable satellite imagery, is needed for efficient LCUM.

To formulate (A1)-(A2) rigorously, we use $\mathbf{C} \in \{0, 1\}^{L_1 \times L_2}$ to denote the change map for a region with $L_1 \times L_2$ pixels, where $L_1 L_2 = L$, and \mathbf{C} satisfies $\text{vec}(\mathbf{C}) = \mathbf{c}$ (cf. (1)). We are now in place to design our CD criterion, composed of three parts:

1. If the (i, j) th pixel is detected as unchanged (i.e., $[\mathbf{C}]_{(i,j)} = 0$), it means that \mathbf{x}_ℓ and \mathbf{y}_ℓ should be similar, where $\ell = i + (j - 1)L_1$. Thus, the first part of our criterion is to minimize $\|(\mathbf{X} - \mathbf{Y}) \mathbf{Diag}(\text{vec}(\mathbf{1}_{L_1} \mathbf{1}_{L_2}^T - \mathbf{C}))\|_F^2$, which collectively considers all such unchanged pixels sifted by the diagonal matrix.
2. The second part is to account for (A1). Note that the number of changed pixels can be written as $\mathbf{1}_{L_1}^T \mathbf{C} \mathbf{1}_{L_2}$ (i.e., the number of 1's in $\mathbf{C} \in \{0, 1\}^{L_1 \times L_2}$), by minimizing which a solution satisfying (A1) is promoted.
3. The third part is to account for (A2). Promoting piecewise smoothness of \mathbf{C} is equivalent to promoting the sparsity of its gradient map [9], where the latter can be achieved by minimizing the so-called (anisotropic) total variation (TV) regularizer [10] defined as

$$\text{TV}(\mathbf{C}) \triangleq \sum_{(i,j)} \left\| \left[\Delta_{(i,j)}^h \mathbf{C}, \Delta_{(i,j)}^v \mathbf{C} \right]^T \right\|_1,$$

where $\Delta_{(i,j)}^h$ (resp., $\Delta_{(i,j)}^v$) is the horizontal (resp., vertical) first-order difference operator at the (i, j) th pixel; precisely, $\Delta_{(i,j)}^h \mathbf{C} \triangleq [\mathbf{C}]_{(i,j)} - [\mathbf{C}]_{(i,j-1)}$ and $\Delta_{(i,j)}^v \mathbf{C} \triangleq [\mathbf{C}]_{(i,j)} - [\mathbf{C}]_{(i-1,j)}$. Note that the TV function, yielding considerable success in machine learning and imaging sciences over the last decade, is convex, but not everywhere differentiable [9].

All in all, we arrive at the following CD criterion:

$$\min_{[\mathbf{C}]_{(i,j)} \in \{0,1\}, \forall (i,j)} \left\{ \left\| (\mathbf{X} - \mathbf{Y}) \mathbf{Diag}(\text{vec}(\mathbf{1}_{L_1} \mathbf{1}_{L_2}^T - \mathbf{C})) \right\|_F^2 + \lambda \mathbf{1}_{L_1}^T \mathbf{C} \mathbf{1}_{L_2} + \eta \sum_{(i,j)} \left\| \left[\Delta_{(i,j)}^h \mathbf{C}, \Delta_{(i,j)}^v \mathbf{C} \right]^T \right\|_1 \right\}, \quad (2)$$

where $\text{vec}(\mathbf{C}) = \mathbf{c}$, and $\lambda \geq 0$ and $\eta \geq 0$ are the regularization weights for balancing these terms. In Section 3, we solve criterion (2), which will also be linked to statistic-based (via convex relaxation) and DI-based (via Property 1) criteria.

3. CD ALGORITHM DESIGN

In this section, we design an algorithm to solve (2), which is a large-scale, non-convex and non-differentiable optimization problem.

We first handle the non-convexity, caused by the binary-valued constraint $\mathbf{C} \in \{0, 1\}^{L_1 \times L_2}$, by relaxing it as

$$\mathbf{C} \in \text{conv} \left\{ \{0, 1\}^{L_1 \times L_2} \right\} = [0, 1]^{L_1 \times L_2},$$

where the value $[\mathbf{C}]_{(i,j)}$ now belongs to the interval $[0, 1]$ and can be interpreted as the change probability of the (i, j) th pixel. Since the objective function of (2) is already convex, the above convex relaxation technique [6, 7] allows us to reformulate (2) as a convex one, which can then be represented concisely using vector representation:

$$\min_{\mathbf{c} \in [0,1]^L} \|(\mathbf{X} - \mathbf{Y}) \mathbf{Diag}(\mathbf{1}_L - \mathbf{c})\|_F^2 + \lambda \mathbf{1}_L^T \mathbf{c} + \eta \text{TV}(\mathbf{c}), \quad (3)$$

whose constraint is now a convex box constraint.

Although (3) is already convex, it is large-scale and non-differentiable. So, we adopt proximal computing in ADMM to solve it. To this end, by defining $\mathbf{Z} \triangleq \mathbf{X} - \mathbf{Y} \equiv [\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(M)}]^T$, $\Psi \triangleq [\mathbf{Diag}(\mathbf{z}^{(1)}), \dots, \mathbf{Diag}(\mathbf{z}^{(M)})]^T$ and $\psi \triangleq \mathbf{diag}(\mathbf{Z}^T \mathbf{Z})$, and by noticing $\|(\mathbf{X} - \mathbf{Y}) \mathbf{Diag}(\mathbf{1}_L - \mathbf{c})\|_F^2 = \|\mathbf{Z}\|_F^2 + \|\Psi \mathbf{c}\|_2^2 - 2\psi^T \mathbf{c}$, we can equivalently recast (3) into a form required by ADMM:

$$\min_{\mathbf{c}=\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4} \frac{1}{2} \|\Psi \mathbf{c}_1\|_2^2 + \left(\frac{\lambda}{2} \mathbf{1}_L - \psi \right)^T \mathbf{c}_2 + \frac{\eta}{2} \text{TV}(\mathbf{c}_3) + I_B(\mathbf{c}_4), \quad (4)$$

where $I_B(\cdot)$ is the indicator function for the box constraint, i.e., $I_B(\mathbf{c}) = 0$ if $\mathbf{c} \in \mathcal{B} \triangleq [0, 1]^L$ ($I_B(\mathbf{c}) = \infty$, otherwise). Then the ADMM algorithm [7] for solving (4) is detailed in Algorithm 1, referred to as CD-ADMM, in which the augmented Lagrangian of (4) is defined as $\mathcal{L}(\mathbf{c}, \{\mathbf{c}_i\}_{i=1}^4, \{\mathbf{d}_i\}_{i=1}^4) = \frac{1}{2} \|\Psi \mathbf{c}_1\|_2^2 + \left(\frac{\lambda}{2} \mathbf{1}_L - \psi \right)^T \mathbf{c}_2 + \frac{\eta}{2} \text{TV}(\mathbf{c}_3) + I_B(\mathbf{c}_4) + \sum_{i=1}^4 \frac{\mu}{2} \|\mathbf{c} - \mathbf{c}_i - \mathbf{d}_i\|_2^2$, where $\{\mathbf{d}_i\}$ are scaled dual variables, and $\mu > 0$ is the penalty parameter. Before deriving closed-form solutions for the subproblems involved in Algorithm 1, we present some analysis results:

Property 1 *The DI-based criterion (1) is subsumed as a special case by our CD criterion (2).*

Property 2 *The sequence $\{\mathbf{c}^k\}$ generated by Algorithm 1 converges to a global optimum of (3), i.e., the convex relaxation of (2).*

Property 1 (providing a theoretical justification for the empirical observation that (2) dominates DI-based criterion in CD performance) can be proved, by using the pixel index transform $\ell \triangleq i + (j - 1)L_1$ and by investigating the setting $(\eta, \lambda) = (0, \zeta^2)$. Property 2 can be proved, by applying ADMM convergence theory [7] on (4). Detailed proofs for these properties are omitted here due to space limitation. We remark that identifiability analysis is also critical to understand the fundamentals of a criterion [11], and this challenging line is left in the future research.

In case that a binary-valued change map is preferred, one can simply threshold the probabilistic map at 0.5 in the maximum-likelihood sense; this strategy is useful when no further statistic information is available, and quite effective as justified using several

Algorithm 1 The CD-ADMM Algorithm for Solving (4)

- 1: **Given** (\mathbf{X}, \mathbf{Y}) , $\lambda > 0$, $\eta > 0$ and $\mu > 0$.
- 2: Initialize $\mathbf{c}^0 := \mathbf{0}_L$ (or by warm start), and $\mathbf{d}_i^0 := \mathbf{0}_L$, $\forall i \in \mathcal{I}_4$. Set $k := 0$.
- 3: **repeat**
- 4: Update $\{\mathbf{c}_i^{k+1}\}_{i=1}^4 \in \arg \min_{\{\mathbf{c}_i\}} \mathcal{L}(\mathbf{c}^k, \{\mathbf{c}_i\}_{i=1}^4, \{\mathbf{d}_i^k\}_{i=1}^4)$;
- 5: Update $\mathbf{c}^{k+1} \in \arg \min_{\mathbf{c}} \mathcal{L}(\mathbf{c}, \{\mathbf{c}_i^{k+1}\}_{i=1}^4, \{\mathbf{d}_i^k\}_{i=1}^4)$;
- 6: Update $\mathbf{d}_i^{k+1} := \mathbf{d}_i^k + \mathbf{c}_i^{k+1} - \mathbf{c}^{k+1}$, $\forall i \in \mathcal{I}_4$;
- 7: $k := k + 1$;
- 8: **until** the predefined stopping criterion is met.
- 9: **Output** the probabilistic change map $\widehat{\mathbf{C}} := \text{vec}_{L_1 \times L_2}^{-1}(\mathbf{c}^k)$.

quantitative CD performance indices. To complete Algorithm 1, we still need to discuss how to update $\{\mathbf{c}_i^{k+1}\}_{i=1}^4$ and \mathbf{c}^{k+1} next.

3.1. Algorithm Implementation

A nice property of the reformulation (4) is that the five variables (i.e., $(\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4, \mathbf{c})$) can all be decoupled [12]. First, we have $\mathbf{c}_1^{k+1} \in \arg \min_{\mathbf{c}_1} \frac{1}{2} \|\Psi \mathbf{c}_1\|_2^2 + \frac{\mu}{2} \|\mathbf{c}^k - \mathbf{c}_1 - \mathbf{d}_1^k\|_2^2$, whose closed-form solution can be derived as $\mathbf{c}_1^{k+1} = \mu(\Psi^T \Psi + \mu \mathbf{I}_L)^{-1}(\mathbf{c}^k - \mathbf{d}_1^k) = \mu [(\|\mathbf{z}_1\|_2^2 + \mu)^{-1}, \dots, (\|\mathbf{z}_L\|_2^2 + \mu)^{-1}]^T \odot (\mathbf{c}^k - \mathbf{d}_1^k)$, where the latter does not involve the inversion of a huge $L \times L$ matrix. Second, $\mathbf{c}_2^{k+1} \in \arg \min_{\mathbf{c}_2} (\frac{\lambda}{2} \mathbf{1}_L - \psi)^T \mathbf{c}_2 + \frac{\mu}{2} \|\mathbf{c}^k - \mathbf{c}_2 - \mathbf{d}_2^k\|_2^2$, whose closed-form solution can be derived as $\mathbf{c}_2^{k+1} = \mathbf{c}^k - \mathbf{d}_2^k - \frac{1}{\mu} (\frac{\lambda}{2} \mathbf{1}_L - \psi)$. Third, we have $\mathbf{c}_3^{k+1} \in \arg \min_{\mathbf{c}_3} \frac{\eta}{2} \text{TV}(\mathbf{c}_3) + \frac{\mu}{2} \|\mathbf{c}^k - \mathbf{c}_3 - \mathbf{d}_3^k\|_2^2$, which is exactly the total variation denoising operator (with $\mathbf{c}^k - \mathbf{d}_3^k$ considered as the noisy input image). This operator can be efficiently solved using split Bregman method [13, 14], whose implementation is available online¹. Fourth, we have $\mathbf{c}_4^{k+1} \in \arg \min_{\mathbf{c}_4} \mathcal{I}_B(\mathbf{c}_4) + \frac{\mu}{2} \|\mathbf{c}^k - \mathbf{c}_4 - \mathbf{d}_4^k\|_2^2$, whose solution can be verified as $\prod_{\mathcal{B}}(\mathbf{c}^k - \mathbf{d}_4^k)$ (the projection of $\mathbf{c}^k - \mathbf{d}_4^k$ onto the box $\mathcal{B} = [0, 1]^L$), i.e., $[\mathbf{c}_4^{k+1}]_i = \begin{cases} [\mathbf{c}^k - \mathbf{d}_4^k]_i, & \text{if } \mathbf{c}^k - \mathbf{d}_4^k \in [0, 1], \\ 0, & \text{if } \mathbf{c}^k - \mathbf{d}_4^k < 0, \\ 1, & \text{otherwise,} \end{cases} \quad \forall i \in \mathcal{I}_L$. Finally, we have $\mathbf{c}^{k+1} \in \arg \min_{\mathbf{c}} \sum_{i=1}^4 \frac{\mu}{2} \|\mathbf{c} - \mathbf{c}_i^{k+1} - \mathbf{d}_i^k\|_2^2$, whose solution can be easily verified as $\mathbf{c}^{k+1} = \frac{1}{4} \sum_{i=1}^4 (\mathbf{c}_i^{k+1} + \mathbf{d}_i^k)$.

4. EXPERIMENTAL RESULTS WITH SPOT-7 DATA

The real satellite images $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{M \times L}$ ($L = 800 \times 800$ pixels with 1.5 meter ground sampling distance; $M = 4$ spectral bands ranging from 450 to 890 nm) used, as described in Section 2, are acquired by SPOT-7 [8] over Hsinchu County region of Taiwan, on March 2017 (for \mathbf{X}) and May 2018 (for \mathbf{Y}), resp. This pair of images illustrates a challenging scenario for CD study due to the different illumination conditions, as can be seen from their false color compositions; cf. Figures 1(a) and 1(b). The land use of the studied area, originally covered by forest mostly, underwent some changes between the two studied time instances due to anthropogenic deforestation; those changed subareas are marked by white color (i.e., $\mathbf{C}_{(i,j)} = 1$) in Figure 1(c), where pixels with $\mathbf{C}_{(i,j)} = 0$ are marked by black color.

¹<https://www.mathworks.com/matlabcentral/fileexchange/36278-split-bregman-method-for-total-variation-denoising>

Table 1. Quantitative comparison.

Methods	P_{FAR} (\downarrow)	P_{MAR} (\downarrow)	P_{DA} (\uparrow)	κ (\uparrow)
iBSiZer [5]	0.0105	0.0207	0.9888	91.20%
cIPCA [3]	0.0195	0.1788	0.9703	76.36%
itPCA [4]	0.0578	0.2994	0.9267	51.25%
CD-ADMM	0.0032	0.0861	0.9915	92.79%

This image pair was fed into benchmark CD methods, including iBSiZer [5], cIPCA [3] and itPCA [4], as well as the proposed CD-ADMM algorithm (i.e., Algorithm 1 with $(\lambda, \eta, \mu) = (1e-6, 1.08, 0.90)$). The changed pixels detected by these algorithms are marked by white color in Figures 1(d), (e), (f) and (g), resp. The probabilistic version of change map $\widehat{\mathbf{C}}$ obtained by CD-ADMM is also provided in Figure 1(h), which would provide good information for user to double confirm whether a region is changed or not. One can see that the change maps obtained by cIPCA and itPCA suffer from serious speckle noise. The change map $\widehat{\mathbf{C}}$ obtained by iBSiZer is much improved but still has some artifacts. The change map $\widehat{\mathbf{C}}$ obtained by CD-ADMM holds best visual resemblance to the reference map, as also indicated quantitatively by the global detection accuracy measure P_{DA} (to be defined later) in Table 1.

For quantitative comparison, we adopt four commonly used CD performance indices. To define them, let N_1 (resp., N_2) be the number of changed (resp., unchanged) pixels that are detected as changed ones, and let N_3 (resp., N_4) be the number of changed (resp., unchanged) pixels that are detected as unchanged ones ($N_1 + N_2 + N_3 + N_4 = L$). Then, the four indices are defined as follows: 1) false alarm rate $P_{\text{FAR}} \triangleq \frac{N_2}{N_2 + N_4}$ (the smaller, the better); 2) missed alarm rate $P_{\text{MAR}} \triangleq \frac{N_3}{N_1 + N_3}$ (the smaller, the better); 3) detection accuracy $P_{\text{DA}} \triangleq \frac{N_1 + N_4}{L}$ (the larger, the better); 4) Cohen's kappa coefficient (the larger, the better), defined as

$$\kappa \triangleq (P_{\text{DA}} - P_{\text{guess}}) / (1 - P_{\text{guess}}), \quad (5)$$

where $P_{\text{guess}} \triangleq [(N_1 + N_3)(N_1 + N_2) + (N_2 + N_4)(N_3 + N_4)] / L^2$ is the probability of the event that a random algorithm (which guesses a pixel as a changed one with probability $\frac{N_1 + N_2}{L}$) correctly detects the class of a pixel. The results are summarized in Table 1. The proposed CD-ADMM performs best in terms of the false alarm rate, but performs second in terms of missed alarm rate. To investigate deeper, for each algorithm under test, we use yellow color to mark those N_3 changed pixels it missed; since the yellow region associates with some $[\widehat{\mathbf{C}}]_{(i,j)} = 0$ (by definition of N_3), it must be non-overlapping with the white region ($[\widehat{\mathbf{C}}]_{(i,j)} = 1$) and is thus displayed on the same subfigure to save space. From Figure 1, most of the N_3 missed pixels do not really matter in practice, because those yellow regions are mostly located inside the white regions; for this reason, the user (e.g., government) would still be aware of the changes in the yellow regions when the user has been aware of the detected changes (in the white regions) caused by deforestation. Thus, P_{MAR} would not be a critical index in this case study. By contrast, P_{FAR} is more important in the LCUM application because high P_{FAR} will waste unnecessary resources during the field survey stage.

Cohen's kappa coefficient κ is probably the most credible index to measure the CD performance, because it takes into account the possibility of the agreement occurring by chance (i.e., by random guess; cf. (5)), for which CD-ADMM significantly outperforms peer CD methods (about 1.6% higher than the state-of-the-art that we have been aware of); note that CD-ADMM (8.51e+1 sec.) is about 20 times faster than iBSiZer (1.85e+3 sec.) (cIPCA and itPCA took 1.03e+2 and 7.95e+1 sec., resp.). As for the global accuracy indicated by P_{DA} , CD-ADMM again performs best.

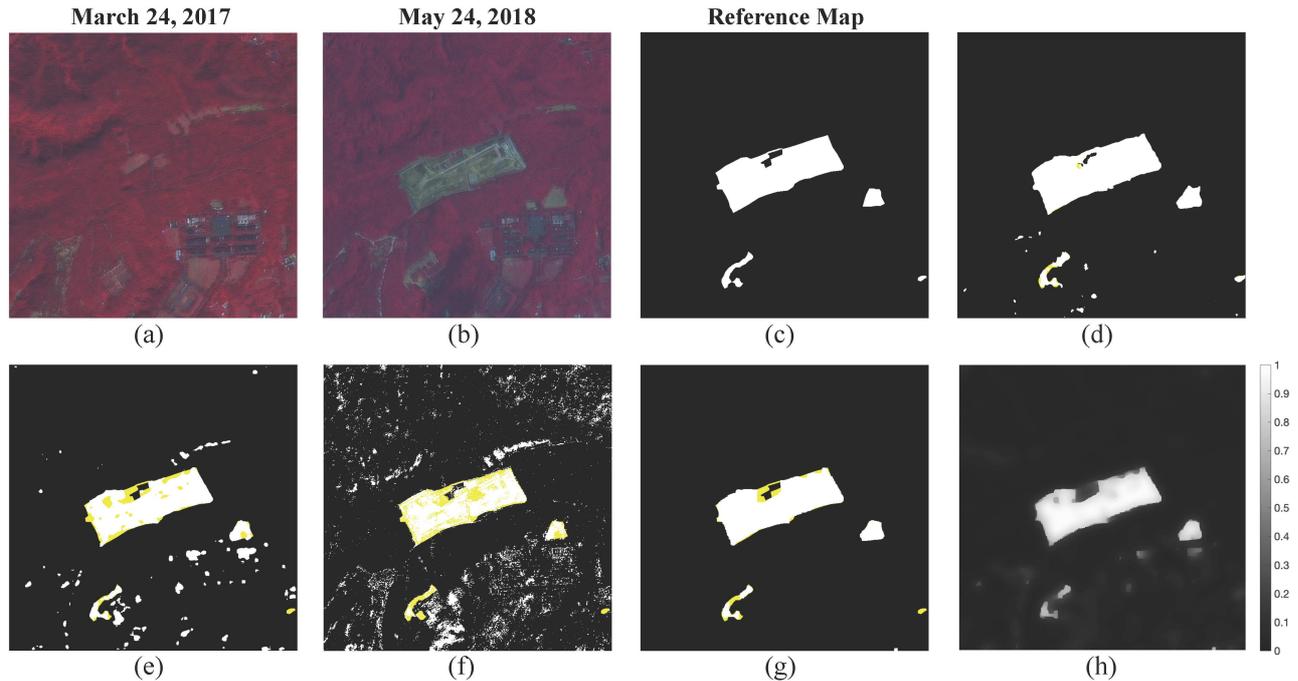


Fig. 1. False color compositions (NIR-R-G) of two SPOT-7 images (X, Y) are displayed in (a) and (b), resp. The reference change map C , labeled artificially, is displayed in (c), where changes are marked by white color. The changes in \hat{C} , detected by iBSiZer, cIPCA, itPCA and the proposed CD-ADMM, are marked by white color in (d), (e), (f) and (g), resp., where the yellow color is used to mark those changed pixels (in C) not detected by these methods. Finally, (h) shows the probabilistic version of \hat{C} obtained by CD-ADMM.

5. CONCLUSION

We designed a new CD criterion under assumptions particularly made for LCUM, and linked it with conventional statistic-based (via convex relaxation) and DI-based (via Property 1) criteria. Then, we adopted proximal computing to develop the CD-ADMM algorithm (i.e., Algorithm 1) for solving our criterion with convergence guarantee (cf. Property 2), where closed-form solutions are derived for efficient implementation. Experiments were conducted on real SPOT-7 satellite images, demonstrating superior efficacy of CD-ADMM in terms of several indices (including Cohen's kappa coefficient). Analyzing the identifiability of the proposed CD criterion is a line deserving future investigation.

6. REFERENCES

- [1] C.-H. Lin, F. Ma, C.-Y. Chi, and C.-H. Hsieh, "A convex optimization-based coupled nonnegative matrix factorization algorithm for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1652–1667, 2018.
- [2] R. H. Topaloğlu, E. Sertel, and N. Musaoğlu, "Assessment of classification accuracies of Sentinel-2 and Landsat-8 data for land cover/use mapping," *Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 41, 2016.
- [3] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k -means clustering," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 4, pp. 772–776, 2009.
- [4] N. Falco, P. R. Marpu, and J. A. Benediktsson, "A toolbox for unsupervised change detection analysis," *International Journal of Remote Sensing*, vol. 37, no. 7, pp. 1505–1526, 2016.
- [5] L. Pasanen and L. Holmström, "Bayesian scale space analysis of temporal changes in satellite images," *Journal of Applied Statistics*, vol. 42, no. 1, pp. 50–70, 2015.
- [6] Y. Liu, F. Condessa, J. Bioucas-Dias, J. Li, and A. Plaza, "Convex formulation for hyperspectral image classification with superpixels," in *IEEE International Geoscience and Remote Sensing Symposium*, 2016, pp. 3294–3297.
- [7] C.-Y. Chi, W.-C. Li, and C.-H. Lin, *Convex Optimization for Signal Processing and Communications: From Fundamentals to Applications*. Boca Raton, FL, USA: CRC Press, 2017.
- [8] V. Parage, B. Vajsova, N. Faget, and P. J. Åstrand, "New sensors benchmark report on SPOT 7," *Citeseer*, 2015.
- [9] M. A. Figueiredo, J. M. Bioucas-Dias, J. P. Oliveira, and R. D. Nowak, "On total variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavelet denoising," in *IEEE International Conference on Image Processing*, 2006, pp. 2633–2636.
- [10] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, 2012.
- [11] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5530–5546, Oct. 2015.
- [12] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2011.
- [13] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.
- [14] C. A. Micchelli, L. Shen, and Y. Xu, "Proximity algorithms for image models: Denoising," *Inverse Problems*, vol. 27, no. 4, p. 045009, 2011.