

# A Circular Invariant Convolution Model-Based Mapping for Multimodal Change Detection

Redha Touati<sup>\*,1,2</sup>, Max Mignotte<sup>1</sup>, Mohamed Dahmane<sup>2</sup>

<sup>1</sup> Vision lab, Département d'Informatique et de Recherche Opérationnelle (DIRO), Université de Montréal, QC, Canada

<sup>2</sup> R&D vision Département, Centre de Recherche Informatique de Montréal (CRIM), Montréal, Qc, Canada

## ARTICLE INFO

Article history:

Received:

Accepted:

Online:

Keywords:

Multisource

Multisensor

Multilooking

Unbalanced satellite images

Circular invariant convolution model

Line minimization process

Change detection

Mapping or projection model

## ABSTRACT

The large and ever-increasing variety of remote sensing sensors used in satellite imagery today explains why detecting changes between identical locations in images that are captured, at two separate times, from heterogeneous capturing systems is a major and challenging recent research problem in the field of satellite imaging for fast and accurate determination of temporal changes. This work presents an original concentric circular invariant convolution model that aims at projecting the first satellite image into the imaging modality of the second image. This allows the two images to have identical statistics so that one can then effectively use a classical monomodal change detection method. The invariant circular convolution kernel parameters are estimated in the least squares sense using a conjugate gradient routine whose optimal direction is determined by a quadratic line search algorithm. After the projection of the before image into the imaging modality domain associated with the after image is achieved, a basic pixel-by-pixel difference permits the estimation of a relevant soft difference map which is segmented into two classes by a multilevel Markovian technique. A series of experiments conducted on several pair of satellite images acquired under different imaging modalities, resolution scales, noise characteristics, change types and events, validates the effectiveness of this strategy. The experimentation shows that the proposed model can process different image pairs with less re-striction about the source images and natural event, coming from different sensors or from the same sensor, for detecting natural changes.

## 1 Introduction

The ever-increasing number of Earth observation satellites, which use today new high-tech sensors, are often technologically quite different from those that have provided the satellite images so far archived. This heterogeneity of satellite image data has thus lately contributed to the advent and development of a novel and rapidly growing research interest in remote sensing and geoscience imaging commonly known as multi-modal (or heterogeneous) change detection (MCD). The purpose of the multimodal CD (MCD) [1][2] is to detect and precisely locate any change in land cover between at least two satellite images taken in the same place, at different times and under different acquisition conditions. It usually concerns CD models that process heterogeneous satellite images, *i.e.* provided by different satellite sensor types which may combine active and passive sensors such as SAR/optical or by one sensor type, with two heterogeneous SAR, optical or others or finally by different satellites using the same satellite sensor but under different specifications,

looks, wavelengths, or calibrations [3].

Recently, MCD has aroused a lot of attention and interest in satellite imagery thanks to its consistency and coherence with our environment favouring the emergence of increasingly heterogeneous images. Nevertheless, it is a difficult procedure because MCD has to be flexible enough to treat a multi-modal image pair, for solving the problems usually considered by the traditional single-modality CD methods [4, 5] including damage, land, environmental monitoring or city planning.

In the literature, the proposed MCD approaches can be classified under four classes. The first class includes the simplest methods which use local descriptors [6] or similarity measures [7, 8, 6, 9], using invariance properties relative to the imaging modality processed. The second category gathers the methods that do not have assumptions about the data distributions and are thus non-parametric. This includes algorithms using machine learning techniques that learn from the training samples [10, 11, 12, 13, 3, 14, 15] [16, 17, 18, 19]

\*Corresponding Author: touatire@iro.umontreal.ca

or from unsupervised parameter, without requiring any training phase, for example the least squares (LSQ) Multidimensional scaling (MDS) and energy-based model, integrating data-driven pixel pairwise constraints, introduced in [20]. The third group relies on parametric methods trying to model common statistical features or relationship between the different imaging modalities or between the different multisource data *via* a set of multivariate distributions [21, 22, 23, 24, 25, 26, 27] or through a pixel pairwise modeling integrated in a Markov Random Field (MRF) model [28]. Finally, the last category regroups algorithmic procedures essentially relying on a projective transformation of the heterogeneous images to a common intermediate feature space, where the projected satellite images would ultimately share the same statistical characteristics and on which any conventional homogeneous CD models could be subsequently used [29, 30, 31, 32, 33, 34, 35, 36, 37, 38]. This category can encompass also the procedures projecting the first image in the imaging modality of the second image or inversely [39]. The method presented in this work fits fully into this sub-category.

More precisely, this work presents a concentric circular invariant convolution mapping which aims at projecting the *before* image onto the imaging modality of the *after* image. In this way, we ensure that the statistics of the pre and after-event satellite images are nearly identical and that a classical monomodal CD procedure can then be efficiently used, as easily as if the two images came from the same imaging modality or sensor with the same settings (specifications/calibrations/wavelength).

To this end, we will show that, once the mapping is done by the proposed convolutive model, the pixel-wise difference image estimated from the image pair shows good likelihood distributions properties corresponding to the “change” and “no change” label classes; that is to say, presenting a mixture of distributions not too far from normal distributions and also not too flat (*i.e.* large or non-informative) so that a binary segmentation with an unsupervised Markovian model to be efficient, despite the big difference between the imaging modalities that could be confronted in multi-modal satellite imagery and which will be evaluated in this paper. As an amelioration of our convolution model-based mapping [1], in the present research we propose a circular invariant model. Furthermore, a three dimensional (3D) convolution based mapping strategy was adopted in the specific case of MCD using two color images (or multispectral). A battery of tests was conducted to quantify the benefits of such improvements.

It is worth mentioning that convolution model estimation has also been widely experienced and analyzed in different image processing problems as image restoration [40] or 2D or 3D deconvolution issues [41, 42, 43, 44].

The reminder of this paper is structured as follows: Section 2 describes the proposed projection model based on a concentric circular invariant convolution representation and the unsupervised Markovian approach. Section 3 presents the results and an experimental comparison with state of the art multimodal CD methods. Section 4 concludes the paper.

## 2 Proposed MCD Model

Let  $I_b$  and  $I_a$ , be the bi-temporal remote sensing image pair of  $N$  pixels, captured in the same area, at two consecutive times (*i.e.*, before and after a given event), and coming from different imaging systems or sensors. Let  $I_b^\#$  and  $I_a^\#$  also be the informational part of the image or the semantic information of the scene concretely representing the set of real objects and materials imaged in  $I_b$  and  $I_a$ . The acquisition system, related to respectively the pre-event and post-event image can be modeled by the following linear models:  $I_b = I_b^\# * u_b$  and  $I_a = I_a^\# * u_a$  where “\*” is the convolution operator and  $u_b$  and  $u_a$  mathematically model the underlying structure of the Point Spread Function (PSF) of the pre-event and post-event data acquisition system [45].

This modeling framework defines a reliable and efficient way for projecting the before-change image to the domain or imaging system of the after-change image. It consists to apply the following operation to the pre-event image:  $(I_b * u_b^{*-1}) * u_a = I_b^\# * u_a$  where “ $u_b^{*-1}$ ” represents the inverse convolution operator of  $u_b$  (giving  $u_b^{*-1} * u_b = \delta$ , such as  $\delta$  is the Dirac impulse function) or by commutativity and associativity of the convolution product, applying the convolution operator  $u_{b \rightarrow a} = (u_b^{*-1} * u_a) = (u_a * u_b^{*-1})$  to  $I_b$ . This operation allows us to convert the original MCD issue into a conventional monomodal CD method which is used in the case of a pair of images having the same acquisition mode and therefore having the same statistics.

In order to reliably find  $u_{b \rightarrow a}$ , the simplest strategy is to search for the convolution filter parameters that will minimize in the LSQ context, the energy function given by equation 1:

$$\hat{u}_{b \rightarrow a} = \arg \min_{u_{(x,y) \in S_{nc}}} \sum_{(x,y) \in S_{nc}} \overbrace{[I_b(x,y) * u_{b \rightarrow a}(x,y)] - I_a(x,y)}^{E(u_{b \rightarrow a})}^2 \quad (1)$$

where  $S_{nc}$  is the set of pixels that do not belong to the class label “change” and which must be evaluated in the final change detection map. As first approximation for  $\hat{u}_{b \rightarrow a}$ , we can take for  $S_{nc}$ , the set of all pixels in the image. This can be justified because the area of change is usually quite small in proportion to the image which, also, can be considered (in a very first approximation) as noisy observations or outliers in the image data with which it is shown that the LSQ estimator performs well (especially thanks to the fact that the function  $E(u_{b \rightarrow a})$  to be minimized is convex with regard to the parameter vector of the convolution filter  $u_{b \rightarrow a}$ ). As a result, different efficient minimization procedures can be adopted [47].

In this work, we use a combination of conjugate gradient and a line minimization search routine guaranteeing a considerably increased convergence speed compared to other minimization methods. Another advantage of such a minimization technique consists on its ability to integrate some hard constraints on  $u_{b \rightarrow a}$ . In our case, since the two imaging modalities are circular invariant, we constraint  $u_{b \rightarrow a}$  to be also circular invariant. To this end, at each iteration  $k$  of the conjugate gradient descent (CGD), we simply average each of the  $(sz \times sz)$  parameter values of  $u_{b \rightarrow a}^{[k]}$  that are equidistant (in the  $L_1$  distance sense) from the center of the convolution filter

Table 1: Heterogeneous datasets description

No	Dates	Area	Image size	Event	Satellite
1	Sept. 1995 - Jul. 1996	Sardinia, It	412×300	Lake overflow (30 m.)	Landsat-5 (NIR band) / Optical
2	Nov. 1999 - Oct. 2000	Gloucester, Uk	330×590	Flooding ( $\approx 40$ m.)	SAR-ERS / Spot
3	July 2006 - July 2007	Gloucester, Uk	2325×4135	Flooding (0.65 m.)	TerraSAR-X / QuickBird 2
4	Feb. 2009 - July 2013	Toulouse, Fr	4404×2604	Construction (2 m.)	TerraSAR-X / Pleiades
5	May 2012 - July 2013	Toulouse, Fr	2000×2000	Construction (0.52 m.)	Pleiades / WorldView 2
6	2001 - Jan. 2002	Congo, Af.	400×800	Nyiragongo volcano eruption (10 m.)	Radarsat (3 / 5-looks)
7	June 2008 - June 2013	Island town, Chn.	415×403	Urbanization (8 m.)	Radarsat-2 / Google Earth
8	June 2008 - Sept. 2012	Shuguang village, Chn	419×342	Building construction (8 m.)	Radarsat-2 / Google Earth
9	June 2008 - June 2009	Yellow river, Chn.	409×367	River's drying up (8 m.)	Radarsat-2 (1 / 4-looks)
10	Aug. 2013 - Aug. 2015	Wehe river, Xi'an city, Chn.	392×303	flooding (2 m.)	GF-1 ( $\neq$ bands) [46]
11	1999 - 2000	Gloucester, Uk	990×554	flooding ( $\approx 25$ m.)	SPOT/NDVI [17]
12	May 2002 - July 2009	Hubei Campus, Chn	240×240	Building construction (2.44 m.)	QuickBird / IKONOS
13	April 2005 - July 2009	Wuhan University, Chn.	400×400	Building construction (3.28 m.)	QuickBird / IKONOS

(see Algorithm 1). This allows us to integrate this constraint with a linear complexity with respect to the  $(sz \times sz)$  parameters of the convolution filter. This overall minimization routine is presented in our basic model [1].

#### Algorithm 1 Concentric Circular Constraint

$u(x, y)$  Convolution filter of dim.  $(sz \times sz)$  [input]  
 $u^\circ(x, y)$  Circular invariant filter [output]  
 $TAB[x][y]$  2D table (length,width)=( $2sz, 2$ ) of floats

• Fill the table  $TAB[i][j]$  with zeros

**for** each cell of  $u(i, j)$  with coordinates  $(i, j)$  **do**

- $d \leftarrow L_1$  distance from  $(i, j)$  to the center of  $u$
- $TAB[d][0] \leftarrow TAB[d][0] + u(i, j)$
- $TAB[d][1] \leftarrow TAB[d][1] + 1$

**for** each cell of  $u^\circ(i, j)$  with coordinates  $(i, j)$  **do**

- $d \leftarrow L_1$  distance from  $(i, j)$  to the center of  $u^\circ$
- $u^\circ(i, j) \leftarrow TAB[d][0]/TAB[d][1]$

In order to improve this approximation of  $\hat{u}_{b \rightarrow a}$ , we can formulate the LSQ estimation problem of the convolution filter parameters as a fixed point problem involving a contraction mapping. Technically speaking, the first estimation of  $\hat{u}_{b \rightarrow a}$  allows us to obtain  $I_{b \rightarrow a}(x, y) = I_b(x, y) * u_{b \rightarrow a}(x, y)$ , corresponding to the image pre-event image projected on the modality of the post-event image. A simple pixel-by-pixel difference between  $I_{b \rightarrow a}(x, y)$  and  $I_a(x, y)$  enables us then to produce an output image of absolute difference that is subsequently binarized to the “change” and the “non-change” class label. To this end, a Gaussian kernel is used to model the distribution of each class and the Expectation Maximization (EM) [48] algorithm is used both to estimate the parameters of this weighted distribution mixture but also to give an approximate estimate of  $S_{nc}$

(i.e., the pixels that does not belong to the class label “change”) with a binarization scheme in the Maximum Likelihood (ML) sense. This process allows us to finally better estimate  $\hat{u}_{b \rightarrow a}$  (see Eq. 1). This process is repeated until the stability of the algorithmic steady state fixed point thus defined is reached. This fixed point-based strategy yields to the following iterative procedure:

$$\hat{u}_{b \rightarrow a}^{[k+1]} = \arg \min_u \sum_{(x,y) \in S_{nc}^{[k]}} [I_b(x, y) * u_{b \rightarrow a}(x, y) - I_a(x, y)]^2 \quad (2)$$

$$\Phi_D^{[k+1]} = \text{EM} \left\{ \underbrace{[I_b(x, y) * u_{b \rightarrow a}^{[k+1]}(x, y) - I_a(x, y)]}_{D^{[k+1]}(x,y)} \right\} \quad (3)$$

$$S_{nc}^{[k+1]} = \text{ML}_{\text{binarization}} \{ \Phi_D^{[k+1]}, D^{[k+1]}(x, y) \} \quad (4)$$

where  $S_{nc}^{[0]}$  stands for the whole image pixels and  $\Phi_D^{[k+1]}$  is the parameter vector of the two weighted normal kernels evaluated by the EM procedure (at iteration  $k + 1$ ) on the difference image  $D(x, y)$ .  $\text{ML}_{\text{binarization}} \{ \Phi_D^{[k+1]}, \cdot \}$  represents the binarization algorithm, based on  $\Phi_D^{[k+1]}$ , in the Maximum Likelihood sense.

A second enhancement can be incorporated in this iterative procedure by simply inverting the direction of temporality of the two (before and after) satellite images in the algorithm introduced in our basic model [1]. Concretely, the estimation of  $\hat{u}_{a \rightarrow b}$  can be done by  $I_{a \rightarrow b}(x, y) = I_a(x, y) * u_{a \rightarrow b}(x, y)$ . This allows us to obtain another difference map between  $I_{a \rightarrow b}(x, y)$  and  $I_b(x, y)$ . The reverse estimate can be used along with the original one to obtain an average estimate ensuring an improved (less noisy) difference image  $D(x, y)$  (see Algorithm 2).

In the case where the image before or after is a single-band image and the second image (possibly a multi-band) is then converted into gray levels and this pair of greyscale images are used as explained in Algorithm 3. A third improvement is proposed in the particular case where we have at our disposal two images before and after in colors (or with  $b > 1$  color bands). In this case, the mapping and a difference map  $D(b, x, y)$  is estimated for each of the  $b$  images individually and sequentially. A difference map is defined finally as

the  $L_\infty$  distance, for each pixel  $(x, y)$ , between the  $b$ -components of each pixel ( $D(x, y) \leftarrow \max_b D(b, x, y)$ ) as illustrated in Algorithm 2.

After the fixed point becomes stable (concretely when  $S_{nc}^{[k]} \approx S_{nc}^{[k+1]}$ , or after  $L_{\max}^P$ , a maximal number of iterations is reached), the parameter vector of the Gaussian kernel mixture  $\Phi_D$  (given by the EM procedure) is used to classify the difference map into two classes, once again, in the Sequential Maximum *A Posteriori* (SMAP) sense by using the multiscale, fine-to-coarse and coarse-to-fine segmentation procedure [49].

This overall minimization routine is summarized in Algorithm 2 and in Algorithm 3 for a N-band image pair.

---

#### Algorithm 2 Convolution-Based Multimodal Change Detection

---

$S(x, y)$	Binary change detection map [output]
$I_b(x, y)$	The before image [input]
$I_a(x, y)$	The after image [input]
$D(x, y)$	Difference map
$\Phi_D$	Vector parameters
$L_{\max}^{EM}$	Maximal number of iterations for EM
$L_{\max}^P$	Maximal number of fixed-point iterations

##### 1. Initialization

- Set  $S(x, y)$  with labels “no change” for each pixel

##### 2. Fixed Point Estimation Step

Repeat

- Estimate  $u_{b \rightarrow a}$  and  $u_{a \rightarrow b}$  (see [1])
- Circular constraint on  $u_{b \rightarrow a}$  and  $u_{a \rightarrow b}$
- $D(x, y) = |[I_b(x, y) * u_{b \rightarrow a}(x, y)] - I_a(x, y)|$   
 $+ |[I_a(x, y) * u_{a \rightarrow b}(x, y)] - I_b(x, y)|$
- $\Phi_D \leftarrow$  EM estimation ( $L_{\max}^{EM}$  iter.) on  $D(x, y)$
- $S^{[l]} \leftarrow$  ML binarization of  $D(x, y)$  based on  $\Phi_D$
- $k++$

Until  $k < L_{\max}^P$  (or  $S(x, y)^{[l]} \neq S(x, y)^{[l-1]}$ );

##### 3. Segmentation Step

- $S(x, y) \leftarrow$  SMAP on  $D(x, y)$  based on  $\Phi_D$
- 

---

#### Algorithm 3 CBMCD For N-Band Image pair

---

$S(x, y)$	Binary change detection map [output]
$I_b(b, x, y)$	The before image with $b < N$ bands [input]
$I_a(b, x, y)$	The after image with $b < N$ bands [input]
$D(x, y)$	Difference map
$\Phi_D$	Vector parameters
$D(b, x, y)$	Temporary map with $b < N$ bands
$L_{\max}^{EM}$	Maximal number of iterations for EM
$L_{\max}^P$	Maximal number of fixed-point iterations

##### 1. Initialization

- Set  $S(x, y)$  with labels “no change” for each pixel and fill  $D(k, x, y)$  with zeros

##### 2. Fixed Point Estimation Step

Repeat

for each band  $b$  of  $[I_b(b, x, y), I_a(b, x, y)]$  do

- Estimate  $u_{b \rightarrow a}$  and  $u_{a \rightarrow b}$
- Circular constraint on  $u_{b \rightarrow a}$  and  $u_{a \rightarrow b}$
- $D(b, x, y) = |[I_b(b, x, y) * u_{b \rightarrow a}] - I_a(b, x, y)|$   
 $+ |[I_a(b, x, y) * u_{a \rightarrow b}] - I_b(b, x, y)|$

- $\Phi_D \leftarrow$  EM ( $L_{\max}^{EM}$  iterations) on  $D(b, x, y)$
- $D(x, y) \leftarrow \max_b D(b, x, y)$
- $S \leftarrow$  ML binarization of  $D(x, y)$  based on  $\Phi_D$
- $k \leftarrow k + 1$

Until  $k < L_{\max}^P$  (or  $S(x, y)^{[l]} \neq S(x, y)^{[l-1]}$ );

- $\Phi_D \leftarrow$  EM ( $L_{\max}^{EM}$  iterations) on  $D(x, y)$

##### 3. Segmentation Step

- $S(x, y) \leftarrow$  SMAP on  $D(x, y)$  based on  $\Phi_D$
- 

## 3 Experimental Results

### 3.1 Heterogeneous Dataset Description

Our strategy is validated by conducting a set of evaluation scenarios on thirteen dissimilar heterogeneous datasets exhibiting different multi modality imaging data such as cross-sensor or multisensor (#1, #5, #12, #13), multisource (#2, #3, #4, #7, #8), multi-looking (#6, #9, #11) or multispectral images (#10) with different resolution levels and image sizes and imaging a wide variety of changed events which are degraded by a wide variety of both noise type and levels as depicted in Table 1.

In all the experimental results, we recall that if one of the two images is in grayscale, we must convert the other image (colors or with  $b > 1$  bands) into grayscale to apply algorithm 2. Otherwise, if the two images contain the same number of bands ( $b$ ), we rely on algorithm 3.

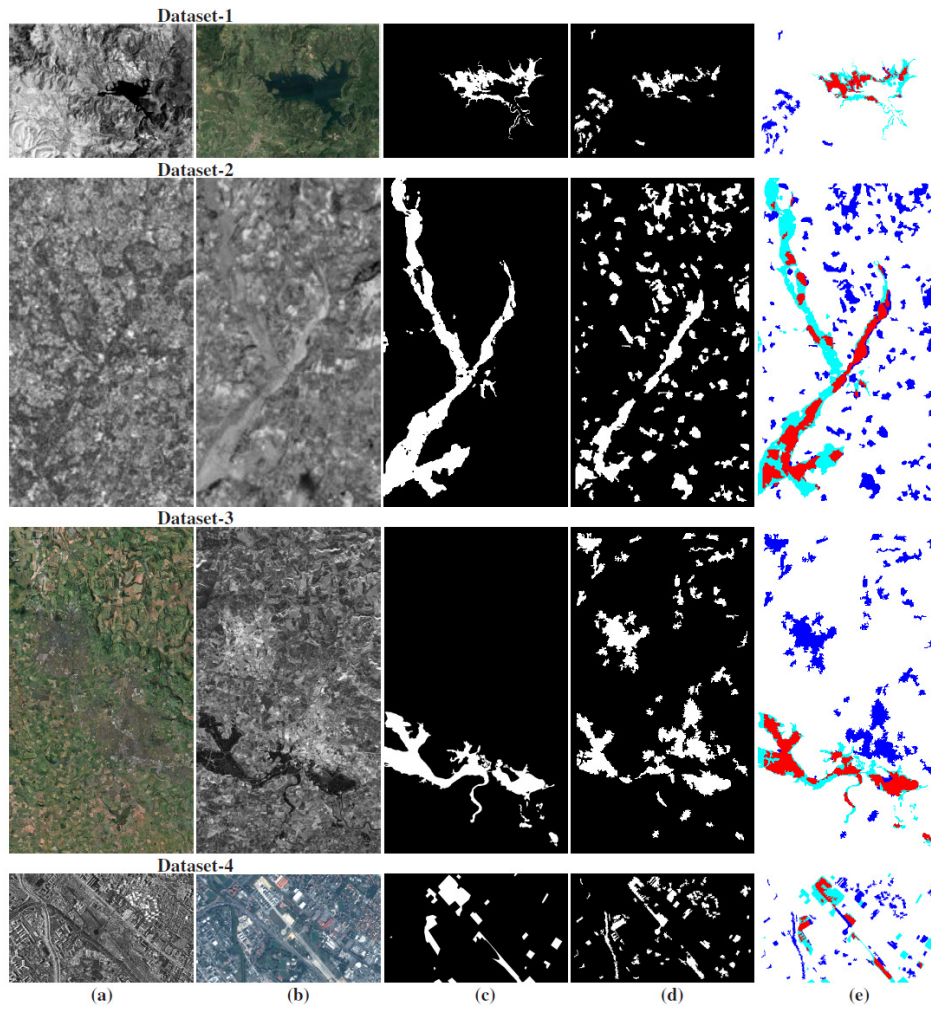


Figure 1: Multimodal datasets. (a) image  $t_1$ , (b) image  $t_2$ , (c) the ground truth; (d) binary map segmentation, (e) confusion map estimated by our model with white (TN), red (TP), blue (FP), cyan (FN) colors

In addition, for computational reasons, we have reduced the image size so that the widest edge of it is at most equal to 500 pixels. For the estimation stage, the size  $sz$  of the convolutional (square) filter was set to 9, with a maximal iteration number  $L_{\max}^G$  set to 200 and a fixed-point iteration number  $L_{\max}^P$  equal to 2. The iteration number  $L_{\max}^{EM}$  used in the EM algorithm is 12 iterations (see our basic model [1]). For the segmentation stage, the depth of the SMAP was set to  $d=9$  [49]. The regularization parameter  $\theta$  was fixed to 0.9 (a value commonly used for this SMAP fine-to-coarse, coarse-to-fine segmentation algorithm [49, 54]).

### 3.2 Convolution Filter Estimation Step Result

For all the test scenarios, the maximum iteration number used in the conjugate gradient descent combined with the minimization line process, is sufficient to ensure the convergence of the minimization procedure. As example, we show, in Figure 4, how evolves the error rate along the minimization process. In addition, for all the scenarios, we estimate for  $u_{b \rightarrow a}$  or  $u_{a \rightarrow b}$  a convolution filter (that can be likened to a point spread function (PSF)) which is quite different from one dataset to another since the frequency response

(or modulation transfer function) of the imaging modality is also quite different for each pair of imaging multi-modalities. Besides, the convolution filter size  $sz$  is also justified, enough because small values of parameters are estimated at the edges of the rectangular filter in all the tested cases (see Fig. 5) for an example of convolution filter estimate).

### 3.3 Results & Evaluation

We evaluate and compare the obtained results using the classification rate accuracy that measures the correct changed and unchanged pixels percentages: ACC and the F-measure Fm which can be defined in the equations 5 and 6.

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} \quad (5)$$

$$Fm = \frac{2TP}{2TP + TN + FN + FP} \quad (6)$$

where TP and TN are the true positives and negatives. FN and FP designate the false negatives and positives.



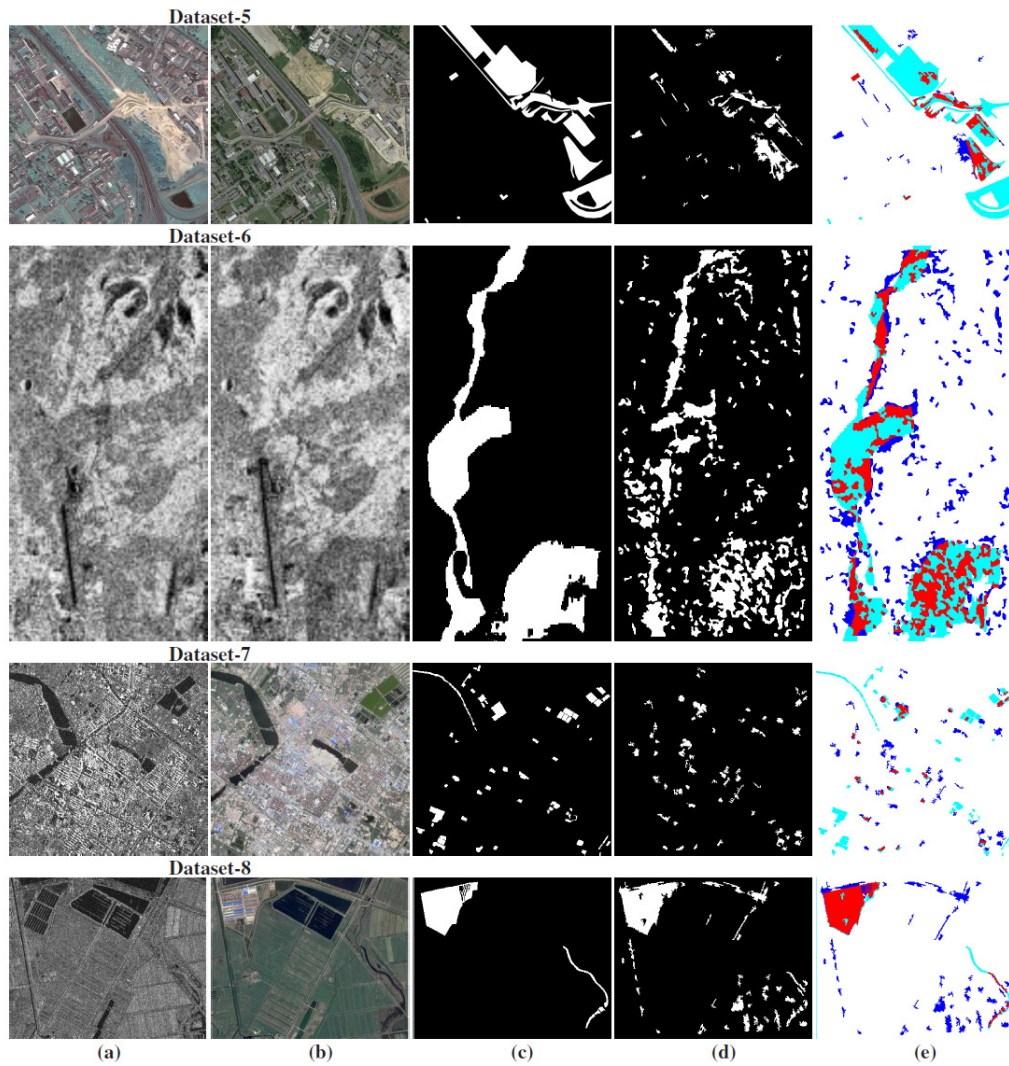


Figure 2: Multimodal datasets. (a) image  $t_1$ , (b) image  $t_2$ , (c) the ground truth; (d) binary map segmentation, (e) confusion map estimated by our model with white (TN), red (TP), blue (FP), cyan (FN) colors

Table 2 shows a comparison between our segmentation results and the different supervised and unsupervised state-of-the-art methods [24, 23, 51, 32] [25, 27, 26] [12, 9, 53, 20] [38, 16, 17, 52].

Figures 1, 2 and 3 compare qualitatively the obtained results. The average accuracy rate of our proposed model is higher than the supervised and unsupervised state-of-the-art methods, obtained by our proposed MCD model on the thirteen heterogeneous dataset is 87.78% with a F-measure equals to 0.444. Compared to these results, our preliminary model (non-circular invariant) [1] gave essentially the same result in terms of rate accuracy but with a much lower F-measure which did not exceed 0.389.

Globally, we notice that our approach works efficiently when a) the considered MCD problem tends towards the simple monomodal CD problem; in other words, in the case of a simple imaging modality which is not too different between the pre-event and post-event image (*e.g.*, when there are similar sensors involving slightly different noise laws between the before and after images), or b) in the case of a couple of images with high resolution, or c) in the case of

simple changed events images with a simple uniform texture; as example, a city is flooding by a river (inducing an uniform texture for a new zone or also in building construction or structure with a uniform texture very different from the rest of the image). Conversely, our approach is a little less suitable when it comes to processing a bi-temporal images in which the variability between the before and after images is more important. Otherwise said, in the case of an imaging modality resulting from the combination of active and passive sensors, thus inducing a combination of multiplicative and additive noises in the before and after images with a dissimilar noise law. Also for images pairs with a low spatial resolution and showing complex changed events; as example, urban site construction (replacing a complex texture zone by the new complex texture area).

It requires about 54 seconds to the MCD to process an image. The processing time depends on the estimation problem complexity, the size of image, and the mono or multichannel based estimation (or 660 seconds for the set of 13 image pairs considered in this validation study) using a code with non-optimized C++ version

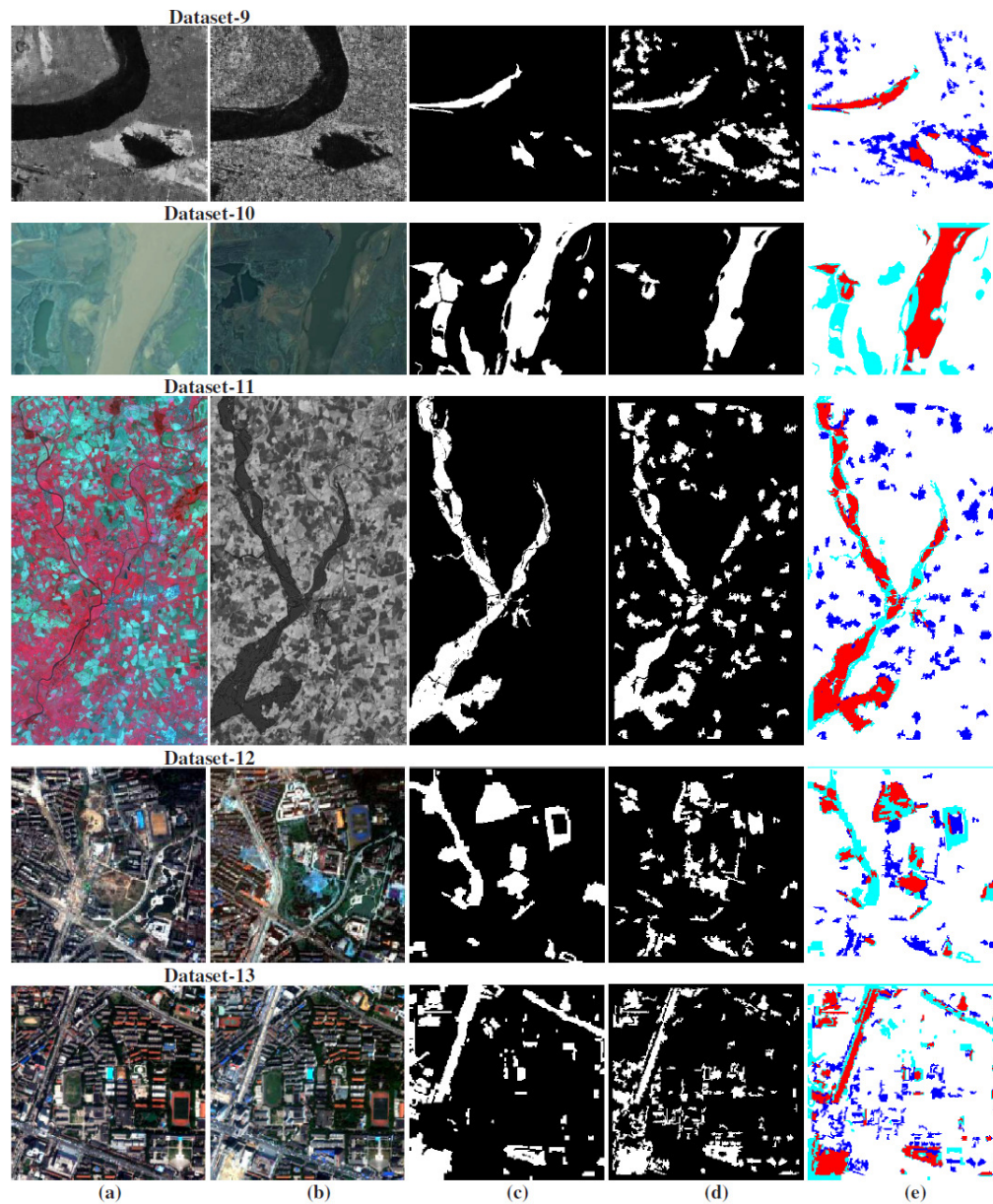


Figure 3: Multimodal datasets. (a) image  $t_1$ , (b) image  $t_2$ , (c) the ground truth; (d) binary map segmentation, (e) confusion map estimated by our model with white (TN), red (TP), blue (FP), cyan (FN) colors

running on *i7 – 930* Intel CPU with 2.8 GHz Linux machine.

## 4 Conclusion

We have presented in this work an efficient mapping model for the change detection problem in multimodal imagery relying on the parametric estimation of an invariant circular convolution CD model. Optimization of the model was performed on a convex quadratic error function using a conjugate gradient routine, whose optimal direction is determined by a quadratic line minimization algorithm, and formulated as a fixed point problem involving a contraction mapping. This allows us to project the before image in the imaging modality space of the second image and inversely so that a differ-

ence change map is then computed pixel by pixel and efficiently binarized with a multiscale Markov model. Our proposed strategy seems well suited for the multimodal change detection issue and robust enough against images provided under different resolution levels, type or level of noise and exhibiting a variety of natural event.

## Acknowledgement

The authors thank the researchers whose shared the *multi-modal* dataset.

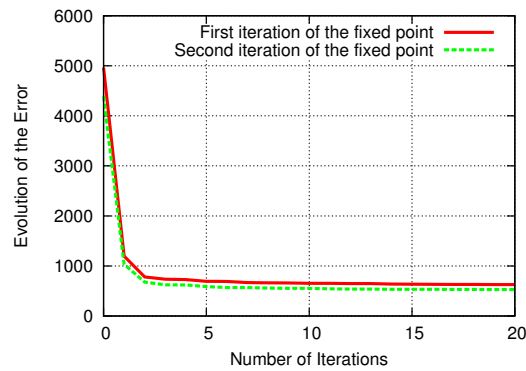


Figure 4: Cost function evolution (see Eq. (1) for the estimation of  $u_{b \rightarrow a}$  associated to dataset #1

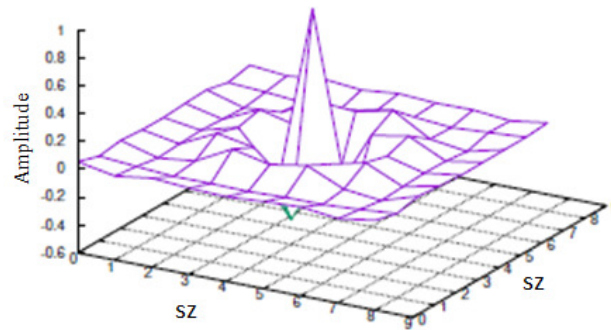


Figure 5:  $9 \times 9$  Convolution filter estimate  $u_{b \rightarrow a}$  associated to dataset #1

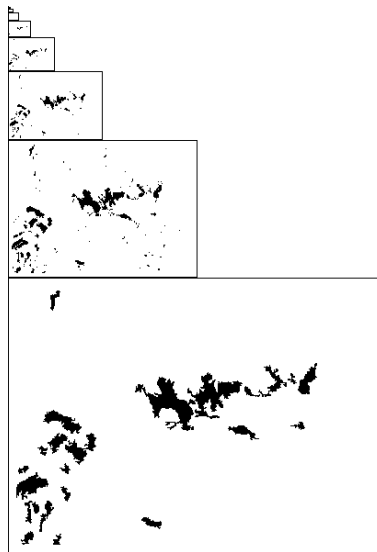


Figure 6: SMAP-based segmentation obtained for the dataset #1

## References

- [1] R. Touati, M. Mignotte, M. Dahmane, "Multimodal change detection using a convolution model-based mapping," in eighth International Conference on Image Processing Theory, Tools and Applications, IPTA 2019, Istanbul, Turkey, November 2019, 2019.
- [2] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot, Q. Du, "Multi-Modal Change Detection, Application to the Detection of Flooded Areas: Outcome of the 2009-2010 Data Fusion Contest," IEEE J. Sel. Topics Appl. Earth Observ., **5**(1), 331–342, 2012.
- [3] L. Su, M. Gong, P. Zhang, M. Zhang, J. Liu, H. Yang, "Deep Learning and Mapping Based Ternary Change Detection for Information Unbalanced Images," Pattern Recognition, **66**(C), 213–228, 2017.
- [4] R. Hedjam, M. Kalacska, M. Mignotte, H. Z. Nafchi, M. Cheriet, "Itera-



- tive Classifiers Combination Model for Change Detection in Remote Sensing Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, **54**(12), 6997–7008, 2016.
- [5] R. Touati, M. Mignotte, "A multidimensional scaling optimization and fusion approach for the unsupervised change detection problem in remote sensing images," in *sixth IEEE International Conference on Image Processing Theory, Tools and Applications, IPTA'16, Oulu, Finland, December 2016*, 1–6, 2016.
  - [6] G. Liu, J. Delon, Y. Gousseau, F. Tupin, "Unsupervised change detection between multi-sensor high resolution satellite images," in *24th European Signal Processing Conference, EUSIPCO 2016, Budapest, Hungary, August 29 - September 2, 2016*, 2435–2439, 2016.
  - [7] V. Alberga, "Similarity Measures of Remotely Sensed Multi-Sensor Images for Change Detection Applications," *Remote Sensing*, **1**(3), 122–143, 2009.
  - [8] D. Brunner, G. Lemoine, L. Bruzzone, "Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, **48**(5), 2403–2420, 2010.
  - [9] R. Touati, M. Mignotte, M. Dahmane, "A new change detector in heterogeneous remote sensing imagery," in *7th IEEE International Conference on Image Processing Theory, Tools and Applications (IPTA 2017), Montreal, Canada, Qc, 2017*.
  - [10] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. L. Rojo-Álvarez, M. Martínez-Ramón, "Kernel-Based Framework for Multitemporal and Multisource Remote Sensing Data Classification and Change Detection," *IEEE Trans. Geoscience and Remote Sensing*, **46**(6), 1822–1835, 2008.
  - [11] P. Du, S. Liu, J. Xia, Y. Zhao, "Information Fusion Techniques for Change Detection from Multi-temporal Remote Sensing Images," *Inf. Fusion*, **14**(1), 19–27, 2013.
  - [12] P. Zhang, M. Gong, L. Su, J. Liu, Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, **116**, 24–41, 2016.
  - [13] M. Gong, P. Zhang, L. Su, J. Liu, "Coupled Dictionary Learning for Change Detection From Multisource Data," *IEEE Trans. Geoscience and Remote Sensing*, **54**(12), 7077–7091, 2016.
  - [14] W. Zhao, Z. Wang, M. Gong, J. Liu, "Discriminative Feature Learning for Unsupervised Change Detection in Heterogeneous Images Based on a Coupled Neural Network," *IEEE Trans. Geoscience and Remote Sensing*, **55**(12), 7066–7080, 2017.
  - [15] N. Merkle, P. F. S. Auer, R. Muller, "On the Possibility of Conditional Adversarial Networks for Multi-Sensor Image Matching," in *Proceedings of IGARSS 2017, IGARSS 2017, 1–4, Fort Worth, Texas, USA, 2017*.
  - [16] J. Liu, M. Gong, K. Qin, P. Zhang, "A Deep Convolutional Coupling Network for Change Detection Based on Heterogeneous Optical and Radar Images," *IEEE Transactions Neural Netw. Learn. Syst.*, 2018.
  - [17] Z. Liu, G. Li, G. Mercier, Y. He, Q. Pan, "Change Detection in Heterogeneous Remote Sensing Images via Homogeneous Pixel Transformation," *IEEE Transactions on Image Processing*, **27**(4), 1822–1834, 2018.
  - [18] R. Touati, M. Mignotte, M. Dahmane, "Anomaly feature learning for unsupervised change detection in heterogeneous images: A deep sparse residual model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **13**(1), 588–600, 2020.
  - [19] R. Touati, M. Mignotte, M. Dahmane, "Partly uncoupled siamese model for change detection from heterogeneous remote sensing imagery," *Journal of Remote sensing and GIS*, **9**(1), 2020.
  - [20] R. Touati, M. Mignotte, "An energy-based model encoding non-local pairwise pixel interactions for multi-sensor change detection," *IEEE Transactions on Geoscience and Remote Sensing*, **56**(1), 2018.
  - [21] B. Stovrik, G. Stovrik, R. Fjørtoft, "On the Combination of Multisensor Data Using Meta-Gaussian Distributions," *IEEE Trans. Geoscience and Remote Sensing*, **47**(7-2), 2372–2379, 2009.
  - [22] G. Mercier, G. Moser, S. Serpico, "Conditional copula for change detection on heterogeneous SAR data," in *2007 IEEE International Geoscience and Remote Sensing Symposium*, 2394–2397, 2007.
  - [23] G. Mercier, G. Moser, S. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, **46**(5), 1428–1441, 2008.
  - [24] F. Chatelain, J. Y. Tournet, J. Inglada, "Change Detection in Multisensor SAR Images Using Bivariate Gamma Distributions," *IEEE Transactions on Image Processing*, **17**(3), 249–258, 2008.
  - [25] J. Prendes, M. Chabert, F. Pascal, A. Giros, J. Tournet, "A New Multivariate Statistical Model for Change Detection in Images Acquired by Homogeneous and Heterogeneous Sensors," *IEEE Transactions on Image Processing*, **24**(3), 799–812, 2015.
  - [26] J. Prendes, M. Chabert, F. Pascal, A. Giros, J. Tournet, "Performance assessment of a recent change detection method for homogeneous and heterogeneous images," *Revue Française de Photogrammétrie et de Télédétection*, **209**, 23–29, 2015.
  - [27] J. Prendes, New statistical modeling of multi-sensor images with application to change detection, Ph.D. thesis, Toulouse, 2015.
  - [28] R. Touati, M. Mignotte, M. Dahmane, "Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based Markov Random Field model," *IEEE Trans. on Image Processing*, **29**(1), 757–767, 2020.
  - [29] M. Xu, C. Cao, H. Zhang, Y. Xue, Y. Li, J. Guo, C. Chang, Q. He, M. Gao, X. Li, "Change detection of the Tangjiashan barrier lake based on multi-source remote sensing data," in *2009 IEEE International Geoscience and Remote Sensing Symposium*, volume 4, IV–303–IV–306, 2009.
  - [30] M. Volpi, F. de Morsier, G. Camps-Valls, M. Kanevski, D. Tuia, "Multi-sensor change detection based on nonlinear canonical correlations," in *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS, 1944–1947, 2013*.
  - [31] C. Wu, B. Du, L. Zhang, "Slow Feature Analysis for Change Detection in Multispectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, **52**(5), 2858–2874, 2014.
  - [32] Z. G. Liu, G. Mercier, J. Dezert, Q. Pan, "Change Detection in Heterogeneous Remote Sensing Images Based on Multidimensional Evidential Reasoning," *IEEE Geoscience and Remote Sensing Letters*, **11**(1), 168–172, 2014.
  - [33] M. Volpi, G. Camps-Valls, D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS Journal of Photogrammetry and Remote Sensing*, **107**, 50–63, 2015.
  - [34] X. Chen, J. Li, Y. Zhang, L. Tao, "Change Detection with Multi-Source Defective Remote Sensing Images Based on Evidential Fusion," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 125–132, 2016.
  - [35] D. Tuia, D. Marcos, G. Camps-Valls, "Multi-temporal and Multi-source Remote Sensing Image Classification by Nonlinear Relative Normalization," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016.
  - [36] Z. Liu, L. Zhang, G. Li, Y. He, "Change detection in heterogeneous remote sensing images based on the fusion of pixel transformation," in *20th International Conference on Information Fusion, FUSION 2017, Xi'an, China, July 10-13, 2017*, 1–6, 2017.
  - [37] L. T. Luppino, S. N. Anfinsen, G. Moser, R. Jenssen, F. M. Bianchi, S. B. Serpico, G. Mercier, "A Clustering Approach to Heterogeneous Change Detection," in *Image Analysis - 20th Scandinavian Conference, SCIA 2017, Tromsø, Norway, June 12-14, 2017, Proceedings, Part II*, 181–192, 2017.
  - [38] R. Touati, M. Mignotte, M. Dahmane, "Change detection in heterogeneous remote sensing images based on an imaging modality-invariant MDS representation," in *25th IEEE International Conference on Image Processing (ICIP'18), Athens, Greece, 2018*.

- [39] L. T. Luppino, F. M. Bianchi, G. Moser, S. N. Anfinsen, "Remote sensing image regression for heterogeneous change detection," *CoRR*, **abs/1807.11766**, 2018.
- [40] S. Benameur, M. Mignotte, J.-P. Soucy, J. Meunier, "Image restoration using functional and anatomical information fusion with application to SPECT-MRI images," *International Journal of Biomedical Imaging*, **2009**, 12 pages, 2009.
- [41] D. Kundur, D. Hatzinakos, "Blind image restoration via recursive filtering using deterministic constraints," in *Proc. International Conference on Acoustics, Speech, and Signal Processing*, volume 4, 547–549, 1996.
- [42] M. Mignotte, J. Meunier, "Unsupervised restoration of brain SPECT volumes," in *Vision Interfaces*, VI'00, 55–60, Montréal, Québec, Canada, 2000.
- [43] M. Mignotte, J. Meunier, "Blind deconvolution of SPECT images using a noise model estimation," in *SPIE Conference on Medical Imaging*, volume 3979, 1370–1377, San Diego, CA, USA, 2000.
- [44] M. Mignotte, J. Meunier, "Three-dimensional blind deconvolution of SPECT images," *IEEE trans. on Biomedical Engineering*, **47**(2), 274–281, 2000, blind deconvolution.
- [45] N. B. Smith, A. Webb, *Introduction to Medical Imaging: Physics, Engineering and Clinical Applications*, Cambridge Texts in Biomedical Engineering, Cambridge University Press, 2010, doi:10.1017/CBO9780511760976.
- [46] T. Zhan, M. Gong, J. Liu, P. Zhang, "Iterative feature mapping network for detecting multiple changes in multi-source remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, **146**, 38 – 51, 2018.
- [47] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press, Cambridge, USA, 2nd edition, 1992.
- [48] A. Dempster, N. Laird, D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Royal Statistical Society*, 1–38, 1976.
- [49] C. Bouman, M. Shapiro, "A Multiscale Random Field Model for Bayesian Image Segmentation," *IEEE Trans. on Image Processing*, **3**(2), 162–177, 1994.
- [50] J. Prendes, M. Chabert, F. Pascal, A. Giros, J.-Y. Tournet, "Change detection for optical and radar images using a Bayesian nonparametric model coupled with a Markov random field," in *Proc. IEEE International Conference on Acoustic, Speech, and Signal Processing, ICASSP'15*, Brisbane, Australia, 2015.
- [51] O. D. Team, "The ORFEO Toolbox Software Guide," 2014, available at <http://orfeo-toolbox.org/>.
- [52] T. Zhan, M. Gong, J. Liu, P. Zhang, "Iterative feature mapping network for detecting multiple changes in multi-source remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, **146**, 38–51, 2018.
- [53] Y. Tang, L. Zhang, "Urban Change Analysis with Multi-Sensor Multispectral Imagery," *Remote Sensing*, **9**(3), 2017.
- [54] M. Mignotte, C. Collet, P. Pérez, P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. on Image Processing*, **9**(7), 1216–1231, 2000.

Table 2: Results comparison.

		<b>Dataset [#3]</b>		<b>Accuracy rate</b>
		<b>Proposed model</b>		<b>0.896</b>
		Touati et al. [20]		0.949
		Touati et al. [9]		0.932
		Touati et al. [38]		0.943
		Prendes et al. [26]		0.844
		Correlation [26]		0.670
		Mutual Information [26]		0.580
		<b>Dataset [#4]</b>		<b>Accuracy rate</b>
		<b>Proposed model</b>		<b>0.882</b>
		Touati et al. [20]		0.867
		Touati et al. [38]		0.878
		Touati et al. [38]		0.881
		Prendes et al. [50, 27]		0.918
		Prendes et al. [25]		0.854
		Copulas [23, 25]		0.760
		Correlation [23, 25]		0.688
		Mutual Information [23, 25]		0.768
		Difference [51, 25]		0.782
		Ratio [51, 25]		0.813
<b>Dataset [#5]</b>				<b>Accuracy rate</b>
<b>Proposed model</b>				<b>0.866</b>
Touati et al. [20]				0.853
Touati et al. [9]				0.870
Touati et al. [38]				0.877
Prendes et al. [26, 27]				0.844
Correlation [26, 27]				0.679
Mutual Information [26, 27]				0.759
Difference. [51, 27]				0.708
Ratio [51, 27]				0.661
		<b>Dataset [#6]</b>		<b>Accuracy rate</b>
		<b>Proposed model</b>		<b>0.818</b>
		Chatelain et al. [24]		0.749
		Correlation [24]		0.713
		Ratio edge [24]		0.737
		<b>Dataset [#7]</b>		<b>Accuracy rate</b>
		<b>Proposed method</b>		<b>0.946</b>
		zhan et al. [52]		0.979
		<b>Dataset [#8]</b>		<b>Accuracy rate</b>
		<b>Proposed model</b>		<b>0.949</b>
		Touati et al. [38]		0.967
		Liu et al. [16]		0.976
		PCC [16]		0.821
<b>Dataset [#9]</b>		<b>Dataset [#10]</b>		<b>Accuracy rate</b>
<b>Proposed model</b>		<b>Proposed model</b>		<b>0.862</b>
zhan et al. [52]		zhan et al. [52]		0.937
		<b>Dataset [#12]</b>		<b>Accuracy rate</b>
		<b>Proposed model</b>		<b>0.835</b>
		Tang et al. [53]		0.986
		Multiscale [53]		0.991
		<b>Optical/Optical [#13]</b>		<b>Accuracy rate</b>
		<b>Proposed model</b>		<b>0.815</b>
		Tang et al. [53]		0.959
		Multiscale [53]		0.966