

Article

# Change Detection Based on Multi-Feature Clustering Using Differential Evolution for Landsat Imagery

Mi Song , Yanfei Zhong \*  and Ailong Ma \*

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; songmi@whu.edu.cn

\* Correspondence: zhongyanfei@whu.edu.cn (Y.Z.); maailong007@whu.edu.cn (A.M.)

Received: 31 August 2018; Accepted: 16 October 2018; Published: 21 October 2018



**Abstract:** Change detection (CD) of natural land cover is important for environmental protection and to maintain an ecological balance. The Landsat series of satellites provide continuous observation of the Earth's surface and is sensitive to reflection of water, soil and vegetation. It offers fine spatial resolutions (15–80 m) and short revisit times (16–18 days). Therefore, Landsat imagery is suitable for monitoring natural land cover changes. Clustering-based CD methods using evolutionary algorithms (EAs) can be applied to Landsat images to obtain optimal changed and unchanged clustering centers (clusters) with minimum clustering index. However, they directly analyze difference image (DI), which finds itself subject to interference by Gaussian noise and local brightness distortion in Landsat data, resulting in false alarms in detection results. In order to reduce image interferences and improve CD accuracy, we proposed an unsupervised CD method based on multi-feature clustering using the differential evolution algorithm (M-DECD) for Landsat Imagery. First, according to characteristics of Landsat data, a multi-feature space is constructed with three elements: Wiener de-noising, detail enhancement, and structural similarity. Then, a CD method based on differential evolution (DE) algorithm and fuzzy clustering is proposed to obtain global optimal clusters in the multi-feature space, and generate a binary change map (CM). In addition, the control parameters of the DE algorithm are adjusted to improve the robustness of M-DECD. The experimental results obtained with four Landsat datasets confirm the effectiveness of M-DECD. Compared with the results of conventional methods and the current state-of-the-art methods based on evolutionary clustering, the detection accuracies of the M-DECD on the Mexico dataset and the Sardinia dataset are very close to the best results. The accuracies of the M-DECD in the Alaska dataset and the large Canada dataset increased by about 3.3% and 11.9%, respectively. This indicates that multiple features are suitable for Landsat images and the DE algorithm is effective in searching for an optimal CD result.

**Keywords:** change detection; detail enhancement; differential evolution; multi-feature clustering; noise robust; remote sensing imagery; structural similarity

## 1. Introduction

Human activities and frequent natural disasters accelerate natural land-cover changes on the Earth. Over the past few decades, remote sensing image CD has become the primary solution to effectively monitor changes in large areas. CD is a process that aims at identifying differences in land cover by analyzing the multi-temporal images acquired in the same geographical area [1]. Timely and accurate CD of natural land cover can provide foundational data for disaster assessment, environmental protection, sustainable development, and maintenance of ecological balance [2,3].

The Landsat series of satellites has been providing continuous observation of the Earth's surface since 1972; its spatio-temporal historical extension is ripe with information that detect land cover changes. Furthermore, the Landsat program has launched seven satellites (Landsat 1, 2, 3, 4, 5, 7, 8) and

is equipped with five types of sensors—Multispectral Scanner (MSS), Thematic Mapper (TM), Enhanced Thematic Mapper (ETM+), Operational Land Imager (OLI), and Thermal Infrared Sensor (TIRS)—which have moderate spatial resolution (15–80 m) and short revisit times (16–18 days). These sensors are suitable for observing natural land cover, such as water, soil, coastlines, and vegetation. In addition, since 2008, the US Geological Survey (USGS) has provided more than 11 million current and historical Landsat images free of charge to users over the Internet (<https://earthexplorer.usgs.gov/>), with ease of access. Thus, Landsat imagery has been widely used in monitoring the natural land cover changes caused by floods [4], forest fires [5,6], and glacial melting [7].

Interferences in Landsat images include Gaussian noise and local brightness distortion; it is not sensitive to the changes of small objects. On one hand, the process of acquiring a remote sensing image—from acquisition to transmission—is inevitably affected by different degrees of noise, e.g., speckle noise in SAR images and Gaussian noise in optical satellite images. On the other hand, Landsat satellites carry passive sensors that record electromagnetic energy; however, the passive sensors are sensitive to the difference in illumination and atmospheric conditions. Therefore, the multi-temporal Landsat images generally have a difference in local brightness [8]. In addition, since intra-class variation and inter-class variation in the Landsat image are relatively small, the expression of changed details is weak.

As the spatial resolution of the Landsat image is not high ( $\geq 15$  m) and the intra-class variability in the image is small, pixel-based methods are suitable for Landsat image CD, which can be further divided into threshold-based or classification-based [3]. In the threshold-based method [9–13], it is crucial to analyze the statistical distribution of DI and find the optimal threshold to distinguish changed pixels from the unchanged ones. For example, Kittler and Illingworth proposed a minimum-error thresholding algorithm (KI), which determines the threshold by optimizing the average pixel classification error rate [11]. The expectation maximization (EM) algorithm automatically selects the decision threshold by minimizing the overall CD error probability under the assumption that pixels in the DI are independent [9]. Zanetti constructed a compound model to describe the statistical distribution of the DI, and the threshold value was determined by the EM algorithm and Bayes decision [12]. Threshold-based methods are efficient and useful. However, they require an accurate estimation of the DI statistical distribution and decision threshold, and are sensitive to image interferences. The Markov random field (MRF) model was introduced to integrate the spatial-contextual information included in the neighborhood of each pixel, and it can remove small false detection regions [9]. However, MRF also requires the selection or estimation of a model for the statistical distribution of the changed and unchanged classes.

As CD is essentially a classification problem, there exist in literature both supervised and unsupervised approaches (clustering-based) to address it [14–21]. A set of training data is required for the supervised methods, while there is no need for training data in the clustering-based methods. Thus, the clustering-based methods are commonly applied in real applications. Clustering is a classification process under the assumption that samples of the same cluster are more similar to one another than samples belonging to different clusters. In clustering-based CD methods, because the image pixels belonging to different clusters cannot be separated with sharp boundaries, one of the most popular methods is the fuzzy c-means (FCM) algorithm [17–21]. The conventional FCM algorithm is not robust to image noise and local brightness distortion because the spatial-contextual information is not taken into account [17]. Krindis and Chatzis proposed a fuzzy local information c-means clustering algorithm (FLICM), which uses a fuzzy local similarity measure to resist image noise. In particular, a fuzzy factor is introduced into its objective function to enhance the clustering performance [18]. Gong et al. proposed a reformulated fuzzy local-information c-means clustering algorithm to classify the changed and unchanged regions in the fused DI [19]. However, fuzzy clustering-based CD methods are sensitive to the initialization of the clusters and easily get stuck to the local minima because the distribution of DI is complex due to image interferences.

Since EAs have promising optimization ability, which can jump out of the local minima, EAs-based optimal clustering has been applied to band selection [22], image classification [23–25], and change detection [8,20,21,26–32]. Celik used the genetic algorithm (GA) to generate optimal CM with minimum mean square error (MSE) [20]. This method is suitable for CD on different types of satellite images. However, the algorithm is sensitive to image interferences and converges slowly due to its encoding strategy, which is based on the whole binary mask. Ghosh et al. proposed a context-sensitive CD method, which adopts a mean filter to remove the image noise in DI and uses GA to evolved the changed and unchanged clusters [21]. Li et al. proposed an evolutionary multi-objective fuzzy clustering method, which can obtain a set of clusters with different trade-offs between removing noise and preserving detail [28]. However, filtering will blur the weak changed details, especially on boundaries or edges, and it cannot resist local brightness distortion in the multi-temporal images. Thus, although the above EAs-based methods can obtain optimal clusters, they are overly dependent on the DI and are affected by image interferences, resulting in inaccurate detection results.

In order to reduce image interferences and improve CD accuracy, we propose an unsupervised CD method based on multi-feature clustering using DE algorithm (M-DECD) for Landsat Imagery. In M-DECD, according to characteristics of Landsat images, the Wiener filter, edge detection and structural similarity (SSIM) index are used to extract spectral-spatial features from the multi-temporal images and DI. To obtain optimal clusters in the multi-feature space, a CD method based on DE algorithm and FCM is proposed and an adaptive strategy adopted to adjust DE's control parameters. Compared to previous works, the major contributions of this paper are as follows.

- (1) **The optimal multi-feature clustering-based CD framework.** This framework is composed of two steps. In the first step, a suitable multi-feature space is constructed according to characteristics of the data. In the second step, an EA is adopted to search for optimal CD clusters using the minimum clustering index. The binary CM can be generated according to the distance between the pixels and the clusters.
- (2) **A multi-feature space for Landsat images CD.** Three complementary spectral-spatial features are extracted from the multi-temporal images and DI, including the Wiener de-noising, structural similarity, and detail enhancement, which are designed to resist the interference of the Gaussian noise and local brightness distortion, while preserving weak changes.
- (3) **Adaptive DE algorithm for the optimization of CD results.** Since the DE algorithm has promising global optimization capabilities, it is employed to search for the optimal changed and unchanged clusters in the complex multi-feature space. The two clusters are encoded into an individual, and the objective function derived from the FCM is utilized to evaluate the quality of clusters. The genetic operators, i.e., crossover, mutation and selection are used for evolutionary processes. In addition, the control parameters of DE, i.e., the scaling factor  $F$  and crossover rate  $CR$ , which have significant influence on optimization performance, have been adaptively adjusted to improve the robustness of the proposed method.

The experimental results show that the M-DECD is robust to interferences of Landsat images, and it can preserve some weak changed edges of small objects. Compared with the conventional methods and the state-of-the-art methods based on evolutionary clustering, M-DECD can generate a better CM and improve detection accuracy, especially in the Canada dataset.

This paper is organized as follows. Section 2 briefly introduces the FCM algorithm and the basic DE algorithm. Section 3 presents the methodology of the proposed M-DECD, including the construction of multi-feature space and the adaptive DE algorithm. Dataset description, experimental results, and parameters analyses are presented in Section 4. Finally, the conclusions are drawn in Section 5.

## 2. Background

### 2.1. Fuzzy Clustering-Based Change Detection

Each pixel of the DI can be mapped to a data point in the feature space according to its gray value. Clustering-based CD approaches attempt to divide the DI into the changed and unchanged classes by searching two clusters in the feature space. Since the pixels of DI belonging to the two clusters are highly overlapped in the feature space, the fuzzy clustering is more suitable to partition the DI than the crisp clustering in which the data points are divided into distinct clusters by hard boundaries, such as  $k$ -means. In fuzzy clustering, data points can potentially belong to multiple clusters, which have robust characteristics for ambiguity. Thus, fuzzy clustering is effective in separating the overlapping clusters. One of the most widely used fuzzy clustering algorithm is the FCM algorithm [3,17–21,28,31]. It is put in a tabular form in Table 1.

**Table 1.** FCM algorithm.

Input unlabeled data, output clusters, fuzzy membership matrix.	
Set number of classes, fuzzy component, termination condition.	
Step 1.	Initialize clusters randomly and compute the membership matrix $U^{(0)}$ using (2)
Step 2.	Calculate the new clusters $G^{(k)}$ with fuzzy membership matrix $U^{(k)}$ using (3)
Step 3.	Update the fuzzy membership $U^{(k)} \rightarrow U^{(k+1)}$ using (2)
Step 4.	Repeat Step 2 and Step 3, until the termination condition is met.

FCM is an iterative method that searches the clusters and produces the partition by minimizing the objective function  $Jm$  distance [3].

$$Jm = \sum_{n=1}^N \sum_k^c u_{n,k}^m \|X_n - G^k\| \quad (1)$$

For the fuzzy clustering-based CD, partition  $c = \{0, 1\}$ .  $N$  is the size of the image,  $X_n$  denotes the  $n$ -th pixel ( $n$ -th data point in the feature space).  $G^1$  and  $G^0$  correspond to the changed and unchanged clusters, respectively.  $u_{n,k}$  is the fuzzy membership of belonging of pixel  $X_n$  to the cluster  $G^k$ .  $m$  is the fuzzy component.  $\|\cdot\|$  represents the Euclidean distance.

The membership  $u_{n,k}$  and the clusters  $G^k$  are updated by using Equations (2) and (3), respectively.

$$u_{n,k} = \frac{\|X_n - G^k\|^{-1/(m-1)}}{\sum_k^c \|X_n - G^k\|^{-1/(m-1)}} \quad (2)$$

$$G(k) = \frac{\sum_{n=1}^N u_{n,k}^m X_n}{\sum_k^c u_{n,k}^m} \quad (3)$$

However, the FCM is sensitive to the initialization of the clusters and usually converges to a local minimum, which can affect detection accuracy. In order to find the optimal clusters with the minimum  $Jm$  distance, we can combine FCM with a global optimization algorithm, such as the DE algorithm.

### 2.2. Differential Evolution Algorithm

The DE algorithm was first proposed by Storn and Price [33], and it is a real value-encoded EA with promising global optimization capability. DE often performs well in all types of optimization problems without any assumption, and it has been proven effective in image clustering [23]. Zhong et al.

proposed a clustering method based on the adaptive multi-objective DE algorithm, which could achieve high accuracy [24]. Ma et al. proposed a multi-objective memetic algorithm for remote sensing image clustering, which adopt DE in the global search step [25].

The classical DE algorithm is put in a tabular form in Table 2. DE is a population-based stochastic heuristic global search method and one of the most powerful real-parameter optimization algorithms in current use [34,35]. The purpose of DE is to search for a global optimal point, which can minimize the specific optimization problem, in a  $D$  dimensional real parameter space, defined as:

$$\min f(G_1, \dots, G_j, \dots, G_D) \text{ s.t. } Low_j \leq G_j \leq Up_j, j = 1, 2, \dots, D \quad (4)$$

where  $G_j$  is the  $j$ -th dimension of point  $G$ .  $Low_j$  and  $Up_j$  denote the minimum and maximum boundaries of the  $j$ -th dimension of the real parameter space.

DE begins with a randomly initiated population of  $NP$  real-valued vectors, and evolves them through evaluation, mutation, crossover, and selection. The crucial idea of DE is to perturb the current generation population members with the scaled differences of randomly selected and distinct population members. After each generation, DE preserves the elitist vectors and abandons the inferior ones. Each vector is then considered the target vector and the evolutionary process is repeated until the termination condition is met, e.g., reach the maximum iterations.

In this paper, DE is employed to optimize the objective function of CD and search for the optimal clusters in the multi-feature space due to its simplicity, ease of implementation, fast convergence, and robustness. To the best of the authors' knowledge, the use of DE for optimization of CD result has not been reported in the literature. Details are described in Section 3.2.

**Table 2.** Differential evolution algorithm.

Model the objective function according to the specific optimization problem. Set population size $NP$ , termination condition, crossover rate $CR$ , and scale factor $F$ .	
Step 1.	<b>Initialization:</b> Initialize the population of $NP$ real-valued parameter vectors randomly. Each vector, also known as individual, forms a candidate solution to the optimization problem.
Step 2.	Evolutionary process. <ol style="list-style-type: none"> <li>2.1. <b>Fitness evaluation:</b> Fitness is the objective function value of each individual, which is used to evaluate the quality of the solution.</li> <li>2.2. <b>Mutation:</b> For each target individual, randomly choose other two individuals from the population and generate a difference vector. A mutant individual is created by adding the weighted difference vector to the target vector under the scale factor <math>F</math>.</li> <li>2.3. <b>Crossover:</b> The trial vector is generated by mixing the components of the mutant vector with the target vector under the crossover rate <math>CR</math>.</li> <li>2.4. <b>Greedy selection:</b> For a minimization problem, the solution is better with smaller fitness; thus the target individual will survive to the next generation if it has smaller fitness than the trial individual, and vice versa.</li> </ol>
Step 3.	Repeat the evolutionary process until the termination condition is met, then output the optimal solution with minimum fitness value.

### 3. Proposed Methodology

The framework of the proposed M-DECD (shown in Figure 1) is composed of two main steps: (1) Constructing a multi-feature space from the multi-temporal images and DI; and (2) adopting the adaptive DE algorithm to search for the optimal changed and unchanged clusters and generate the binary CM. For a better understanding, the meaning of commonly used symbols is listed in Table 3.

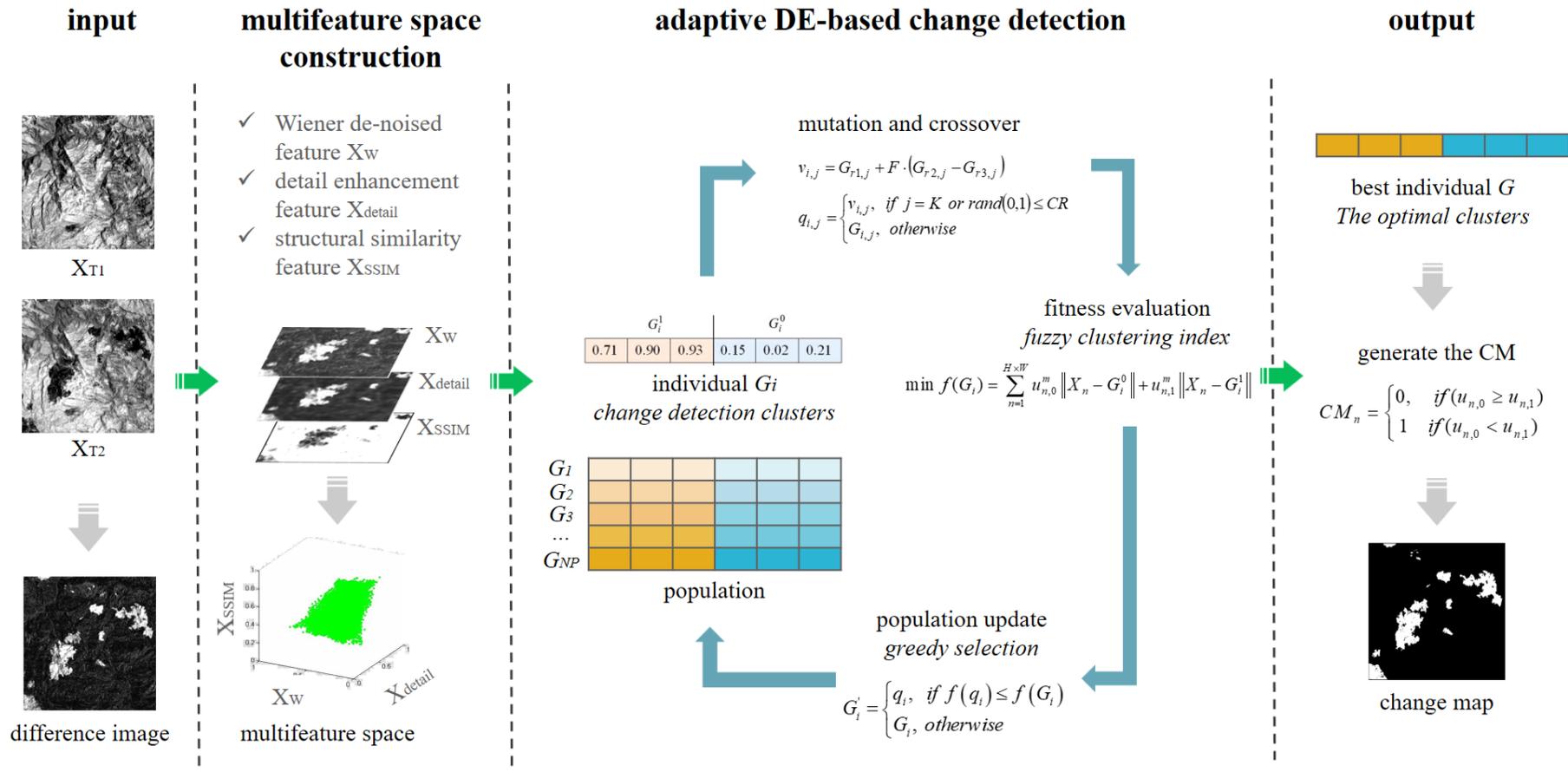
**Table 3.** Symbol annotation.

Symbol	Meaning
$X_{T1}, X_{T2}$	multi-temporal images
$X_d$	difference image
$H, W$	image height and width
$CM$	binary change map
$X_{edge}$	edge feature
$X_w$	wiener de-noised feature
$X_{detail}$	detail enhancement feature
$X_{SSIM}$	structural similarity feature
$X$	multi-feature
$F$	scale factor
$CR$	crossover rate

We assume that two co-registered images  $X_{T1}$  and  $X_{T2}$  of size  $H \times W$  are captured over the same geographical area at two different times,  $T_1$  and  $T_2$ , respectively. The purpose of the M-DECD is to generate an accurate CM,  $CM = \{cm(i, j) | 1 \leq i \leq H, 1 \leq j \leq W\}$ , where  $cm(i, j) \in \{0, 1\}$ .

The first step of the M-DECD is to calculate a DI using multi-temporal images  $X_{T1}$  and  $X_{T2}$ . For the Landsat images, we use the images subtraction to compute the DI ( $X_d$ ), which calculates the absolute value of the spectral difference.

$$X_d = |X_{T1} - X_{T2}| \quad (5)$$



**Figure 1.** Framework of the proposed M-DECD. (a) Given two registered images, the DI is generated by images subtraction. (b) Three image features are extracted to construct a multi-feature space, including the Wiener de-noised feature, detail enhancement feature and structural similarity feature. (c) The optimal clusters in the multi-feature space are obtained by the adaptive DE algorithm. (d) Generate the binary CM according to the optimal clusters and fuzzy membership matrix.

### 3.1. Multifeature Space Construction

According to the characteristics of Landsat images, a multi-feature space is constructed to resist the interferences of Gaussian noise and local brightness distortion, while preserving important weak changes. The multi-feature space consists of three complementary elements: Wiener de-noising, the detail enhancement and structural similarity, which are extracted from multi-temporal images and DI.

#### 3.1.1. Wiener De-Noised Feature

Image denoising is an important step in digital image processing, and it can effectively reduce the small false detection regions in CD. The evolutionary clustering-based CD methods apply mean filtering to the difference image and utilize the average value of the neighborhood pixels as the spatial feature [21,28]. Although the mean filter can remove some image noise, it will blur important changed details, such as boundaries or edges.

For Landsat images, Gaussian noise is generated by the resistive components of the receiver. Since the Wiener filter can effectively remove Gaussian noise and retain the high-frequency information of the image (e.g., boundaries and edges), in M-DECD, we apply Wiener filtering to the DI and generate the Wiener de-noised feature. The Wiener filter removes the noise by minimizing the MSE between the output image and the expected result, and it can adaptively conduct smoothing according to the local image variance. When the local variance is large, it performs little smoothing, and vice versa. Therefore, the quality of filtering result will not decrease drastically with an increase of the filter window size [36].

The Wiener filter [37] estimates the local mean  $\mu$  and variance  $\sigma^2$  around each pixel using Equations (6) and (7), respectively. The Wiener de-noised feature  $X_w$  is then calculated using Equation (8).  $\eta$  is the  $n \times m$  local neighborhood of each pixel in  $X_d$ , and  $v^2$  is the average of the local estimated variances.

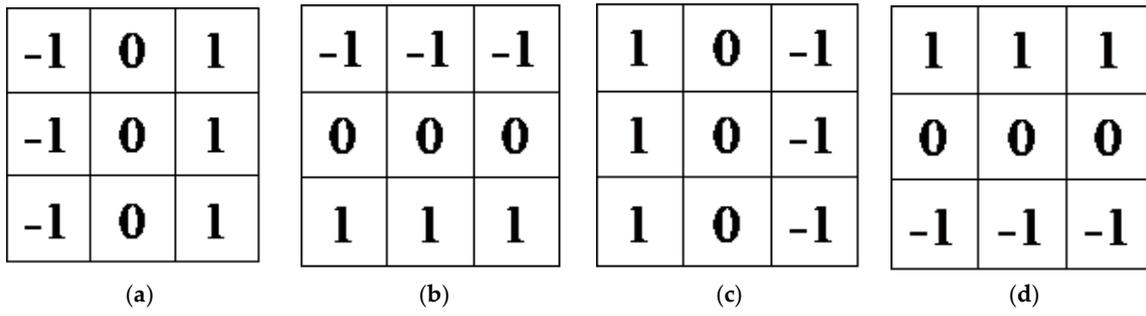
$$\mu = \frac{1}{nm} \sum_{i,j \in \eta} X_d(i,j) \quad (6)$$

$$\sigma^2 = \frac{1}{nm} \sum_{i,j \in \eta} (X_d^2(i,j) - \mu^2) \quad (7)$$

$$X_w(i,j) = \mu + \frac{\sigma^2 - v^2}{\sigma^2} (X_d(i,j) - \mu) \quad (8)$$

#### 3.1.2. Detail Enhancement Feature

For Landsat images, the intra-class variation and inter-class variation are relatively small, and the expression of changed details is weak, especially at the edge of the changed region. Thus, except for the noise removal, another challenge of Landsat images CD is detail preservation. In [38], a modified Perona-Malik filter was employed to remove image noise and extract stable edges for CD. In [39], a novel edge-preserving MRF method for CD was proposed, which integrated the edge and detail-preserving information in the MRF modeling. In the proposed M-DECD, the image details are enhanced by fusing the edge features with the DI. In order to obtain a full range of edge features, four edge detection operators of different directions are applied to the DI, including  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ . The convolution templates are shown in Figure 2.



**Figure 2.** The convolution templates for edge detection. (a) 0°. (b) 90°. (c) 180°. (d) 270°.

After extraction of multiple edge features  $X_{edge}$ , in order to prevent abnormal sample data possibly arising out of the multiple features from interfering with the subsequent change judgment, all the features need be normalized to  $[0, 1]$ . The DI is then combined with the edge features using the weighted sum strategy and the detail enhancement feature  $X_{detail}$  is defined as:

$$X_{detail} = X_d + \sum_{k=1}^4 |X_{edge}(k)| \quad (9)$$

### 3.1.3. Structural Similarity Feature

In order to resist the local brightness difference between the multi-temporal Landsat images, the SSIM [40] was utilized to measure the similarity between the multi-temporal images, and is calculated based on local luminance, contrast, and structure comparisons. The SSIM has been used in various applications, such as image compression [40], video surveillance, and object detection [41]. In [8], the difference image was obtained by the SSIM.

The SSIM uses three components, including luminance  $l$ , contrast  $c$ , and structure  $s$ , to compute the structural similarity feature  $X_{SSIM}$  using Equation (10), where  $\alpha$ ,  $\beta$ , and  $\gamma$  are the exponents used to adjust the influence of each measurement. In M-DECD,  $\alpha = \beta = \gamma = 1$ . The variables  $l$ ,  $c$ , and  $s$  are calculated using Equation (11).

$$X_{SSIM} = [l(X_{T1}, X_{T2})]^\alpha [c(X_{T1}, X_{T2})]^\beta [s(X_{T1}, X_{T2})]^\gamma \quad (10)$$

$$\begin{cases} l(X_1, X_2) = \frac{2\mu_{X_{T1}}\mu_{X_{T2}}}{\mu_{X_{T1}} + \mu_{X_{T2}} + \varepsilon_1} \\ c(X_{T1}, X_{T2}) = \frac{2\sigma_{X_{T1}}\sigma_{X_{T2}}}{\sigma_{X_{T1}} + \sigma_{X_{T2}} + \varepsilon_2} \\ s(X_{T1}, X_{T2}) = \frac{\sigma_{X_{T1}X_{T2}}}{\sigma_{X_{T1}}\sigma_{X_{T2}} + \varepsilon_3} \end{cases} \quad (11)$$

where  $\mu$ ,  $\sigma$ , and  $\sigma_{X_{T1}X_{T2}}$  are the mean, standard deviation, and cross correlation between  $X_{T1}$  and  $X_{T2}$ , respectively.  $\varepsilon_1$ ,  $\varepsilon_2$ , and  $\varepsilon_3$  are small positive constants to avoid instability. The image statistics ( $\mu$ ,  $\sigma$ , and  $\sigma_{X_{T1}X_{T2}}$ ) are computed as proposed in [40].

If  $X_{SSIM}$  is close to 1, it indicates that the local structures of the two-phase images have a high similarity. On the contrary, if  $X_{SSIM}$  is close to 0, it means that the local structures of the two-phase images are quite different.

After extracting the Wiener de-noised feature, the detail enhancement feature and the structural similarity feature, a multi-feature space  $X = [X_w, X_{detail}, X_{SSIM}]$  is constructed by stacking these feature maps. Then, each pixel in the DI can be mapped to a data point in the multi-feature space and the CD is transformed to a classification problem, and it is crucial to find the optimal changed and unchanged clusters in the multi-feature space.

### 3.2. Adaptive Differential Evolution-Based Change Detection

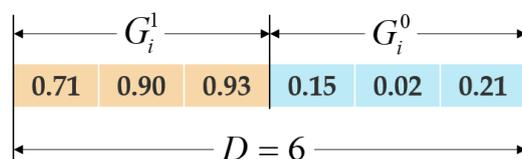
In M-DECD, a CD method based on adaptive DE algorithm and FCM is proposed. The DE is employed to search the optimal clusters in the multi-feature space by minimizing the objective function derived from FCM. Input the multi-feature data  $X = [X_w, X_{detail}, X_{SSIM}]$ , then randomly generate a set of potential clusters in the multi-feature space, and produce new clusters through mutation and crossover of DE algorithm. The  $Jm$  distance of FCM is used as the objective function of DE to evaluate the quality of these clusters, and those elite clusters with smaller objective function value will survive to the next generation until the process reaches the maximum iterations. Finally, output the optimal clusters and the multi-feature data can be divided into changed and unchanged classes according to their fuzzy membership to the optimal clusters. The adaptive DE-based CD method is composed of four steps, which are described as follows:

#### 3.2.1. Population Initialization

The population consists of  $NP$  individuals,  $G = \{G_1, G_2, \dots, G_{NP}\}$ . Each individual represents a candidate solution of the CD problem, i.e., a vector of changed and unchanged clusters, as shown in Figure 3. Each individual in the population is randomly generated in the multi-feature space.

$$G_{i,j} = \min X_j + \text{rand}(0, 1) \cdot (\max X_j - \min X_j) \tag{12}$$

where  $i = 1, 2, \dots, NP, j = 1, 2, \dots, D, G_{i,j}$  represents the  $j$ -th dimension of the  $i$ -th individual in the population, and  $D$  is the dimension of the two clusters.  $\min X_j$  and  $\max X_j$  correspond to the minimum and maximum values in the  $j$ -th dimension of the input multi-feature data, respectively, and  $\text{rand}(a, b)$  is a random value between  $a$  and  $b$ .



**Figure 3.** Individual represents the clusters of changed and unchanged classes.  $G_i^1$  is the cluster of the changed class, and  $G_i^0$  is the cluster of the unchanged class.

#### 3.2.2. Fitness Evaluation

The objective function is modeled according to the fuzzy clustering-based CD problem. In this paper, we adopted the  $Jm$  distance of FCM, which is designed to minimize the intra-class distance, as the objective function of DE algorithm. Fitness is the objective function value of each individual, which can evaluate the quality of the solution. The purpose of DE is to obtain CD clusters with the minimum fitness value. The fitness of individual  $G_i$  can be calculated using Equation (13).

$$\min f(G_i) = \sum_{n=1}^{H \times W} u_{n,0}^m \|X_n - G_i^0\| + u_{n,1}^m \|X_n - G_i^1\|, \quad i = 1, 2, \dots, NP \tag{13}$$

where  $G_i$  is a  $D$ -dimension individual consisting of changed cluster  $G_i^1$  and unchanged cluster  $G_i^0$ .  $X_n$  is a  $\frac{D}{2}$ -dimension vector which denotes the  $n$ -th point of the multi-feature data  $X = [X_w, X_{detail}, X_{SSIM}]$ .  $u_{n,k}, k = \{0, 1\}$  is the fuzzy membership which represents the probability of belonging of point  $X_n$  to the cluster  $G_i^k$ , and it is calculated according to Equation (2).  $u_{n,0}^m \|X_n - G_i^0\|$  represents the intraclass distance of unchanged pixels, and  $u_{n,1}^m \|X_n - G_i^1\|$  represents the intra-class distance of changed pixels.

#### 3.2.3. Mutation and Crossover

DE upsets the current population with the scaled difference of randomly selected and distinct individuals. For each target individual  $G_i$ , we randomly pick three exclusive individuals  $G_{r1}, G_{r2}, G_{r3}$

from the current population and generate a mutant individual  $v_i$  using Equation (14). The mutant individual is a vector of potential clusters around the target individual.  $F \in [0, 2]$  is the scaling factor.  $v_i$  has to be constrained according to the domain of multi-feature space.

$$v_{i,j} = G_{r1,j} + F \cdot (G_{r2,j} - G_{r3,j}) \quad (14)$$

In order to produce a desirable vector of clusters, the crossover operator is then applied to the mutant individual  $v_i$  to generate a trial individual  $q_{i,j}$  using Equation (15), where  $CR \in [0, 1]$  is the crossover rate.  $K$  is a random integer between 0 and  $D$ .

$$q_{i,j} = \begin{cases} v_{i,j}, & \text{if } j = K \text{ or } \text{rand}(0, 1) \leq CR \\ G_{i,j}, & \text{otherwise} \end{cases} \quad (15)$$

Two control parameters (scaling factor  $F$ , crossover rate  $CR$ ) affect the optimization result. If properly designed, an adaptive strategy can enhance the robustness of the algorithm and improve the convergence rate. In the M-DECD, a self-adaptive strategy is adopted, and it can dynamically adjust the parameters according to the fitness of the individual (details can be found in [42]). The mutation rate  $p_m$  is adaptively determined by Equation (16), which means that a vector of clusters with a smaller fitness is more likely to retain its control parameters. Before performing mutation and crossover, the new control parameters  $F'$  and  $CR'$  are updated using Equations (17) and (18), respectively.

$$p_m = \frac{f(G_i) - \min f(G_i)}{\max f(G_i) - \min f(G_i)} \quad (16)$$

$$F' = \begin{cases} 1 - \text{rand}_1(0, 1)^{\left(1 - \frac{\text{gen}}{\text{maxgen}}\right)^2}, & \text{if } \text{rand}_2(0, 1) < p_m \\ F, & \text{otherwise} \end{cases} \quad (17)$$

$$CR' = \begin{cases} \text{rand}_3(0, 1), & \text{if } \text{rand}_4(0, 1) < p_m \\ CR, & \text{otherwise} \end{cases} \quad (18)$$

where  $\text{rand}_k(0, 1)$ ,  $k \in \{1, 2, 3\}$  denotes the uniform random values within the range  $[0, 1]$ ,  $\text{gen}$  is the iteration number,  $\text{maxgen}$  is the maximum iterations. In the initial generations, adaptive DE tends to search the solutions in the multi-feature space uniformly and, in the later generations, it tends to search the solutions near the optimal clusters.

#### 3.2.4. Selection

To keep the population size a constant, DE applies the greedy selection strategy to decide whether the new solutions can survive to the next generation or not, which is executed between the target individual  $G_i$  and trial individual  $q_i$  using Equation (19), where  $G_i'$  is the  $i$ -th individual in the next generation, and  $f(\cdot)$  is the objective function.

$$G_i' = \begin{cases} q_i, & \text{if } f(q_i) \leq f(G_i) \\ G_i, & \text{otherwise} \end{cases} \quad (19)$$

After selection, potential clusters with smaller  $Jm$  distance are preserved. The evolutionary process should be repeated until the number of iterations reach the maximum, Finally, output the optimal clusters and its fuzzy membership  $u_{n,k}$ ,  $n = \{1, 2, \dots, N\}$ ,  $k = \{0, 1\}$ .

### 3.2.5. Generation of the Change Map

The optimal CM can be obtained by assigning a label (0: unchanged, 1: changed) to each pixel according to its fuzzy membership.

$$CM_n = \begin{cases} 0, & \text{if } (u_{n,0} \geq u_{n,1}) \\ 1, & \text{if } (u_{n,0} < u_{n,1}) \end{cases} \quad (20)$$

### 3.3. Computational Complexity of the M-DECD Algorithm

The computational complexity of M-DECD can be split into two parts: (1) the construction of the multi-feature space; (2) the adaptive DE clustering-based CD. We assume that the size of the input image is  $M \times N$ , and the dimension of the multi-feature space is  $D$ , then the computational complexity for part (1) is  $O(M \times N \times D)$ . The adaptive DE clustering-based CD can be further split into six steps as follows:

- (1) *Population initialization.* The computational complexity should be  $O(L \times NP)$ , where  $L$  denotes the length of the individual, i.e., dimension of the changed and unchanged clusters.
- (2) *Fitness evaluation.* The fitness of each individual is calculated based on the input data, so the computational complexity of the fitness evaluation of  $NP$  individuals should be  $O(M \times N \times NP)$ .
- (3) *Population optimization.* This step is composed of fitness evaluation, mutation, crossover and selection. The whole computational complexity should be  $O(L \times NP \times G)$ , where  $G$  is the maximum generations.
- (4) *Generation of the change map.* The computational complexity should be  $O(M \times N)$ .

## 4. Experiments and Analysis

In order to demonstrate the effectiveness of the M-DECD, four real Landsat datasets were tested in the experiments. Seven pixel-based methods for remote sensing images with moderate spatial resolution are presented, including a threshold-based method (KI [11]), a spatial-contextual method (Markov random field expectation maximization, MRF+EM [9]), fuzzy clustering-based methods (FCM [17]; fuzzy local information c-means algorithm, FLICM [18]), and EA-based methods (genetic algorithm encoding change mask, GA-mask [20]; genetic algorithm combined with FCM algorithm, GA-FCM [21]; multi-objective EA with spatial information, MOEA/D [28]).

The CD result is a binary CM, where the white pixels in the map represent the changed objects, and the black pixels represent the unchanged objects. In order to quantitatively analyze the experiment results, we compare the CMs with the reference image. Generally speaking, the false alarms (FAs), missed alarm (MAs), overall error (OE), and kappa coefficient are commonly used as evaluation indexes to assess the result of CD. MA represents the number of pixels that are classified into the unchanged class but are changed in the reference image. FA denotes the number of pixels, which are classified into the changed class but are unchanged in the reference image. OE is defined as:

$$OE = FA + MA \quad (21)$$

The kappa coefficient is usually applied to evaluate the effect of the classification. A higher value of kappa indicates a better detection result. Kappa can be calculated as:

$$Kappa = \frac{PCC - PRE}{1 - PRE} \quad (22)$$

where  $PCC$  represents the percentage of correct classification,  $PRE$  represents the proportion of expected agreement.  $PCC$  and  $PRE$  are defined by Equations (23) and (24).

$$PCC = (TP + TN) / (TP + FA + TN + MA) \quad (23)$$

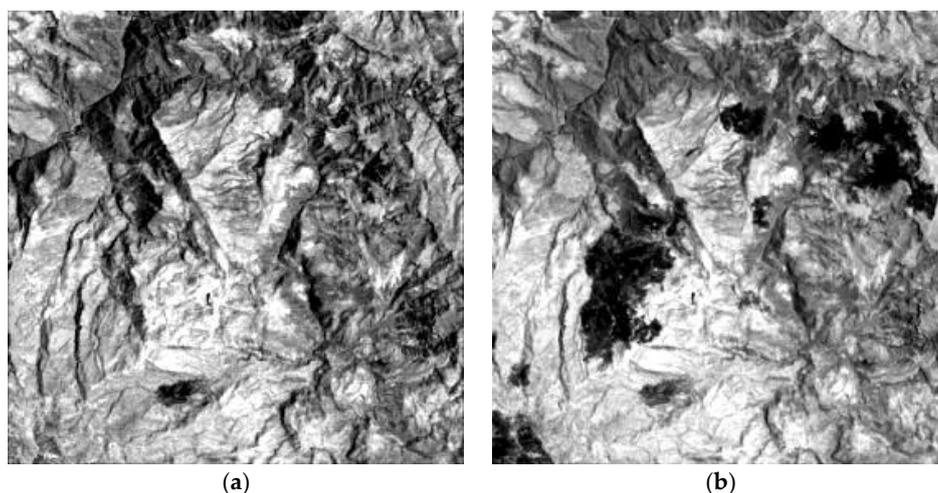
$$PRE = ((TP + FA) \times N_c + (MA + TN) \times N_u) / N^2 \quad (24)$$

where  $N$  is the number of image pixels.  $TP$  is the number of changed pixels that are correctly classified, and  $TN$  is the number of unchanged pixels that are correctly classified.  $N_c$  and  $N_u$  are the number of changed pixels and unchanged pixels in the reference image, respectively.

#### 4.1. Dataset Description

##### 4.1.1. Mexico

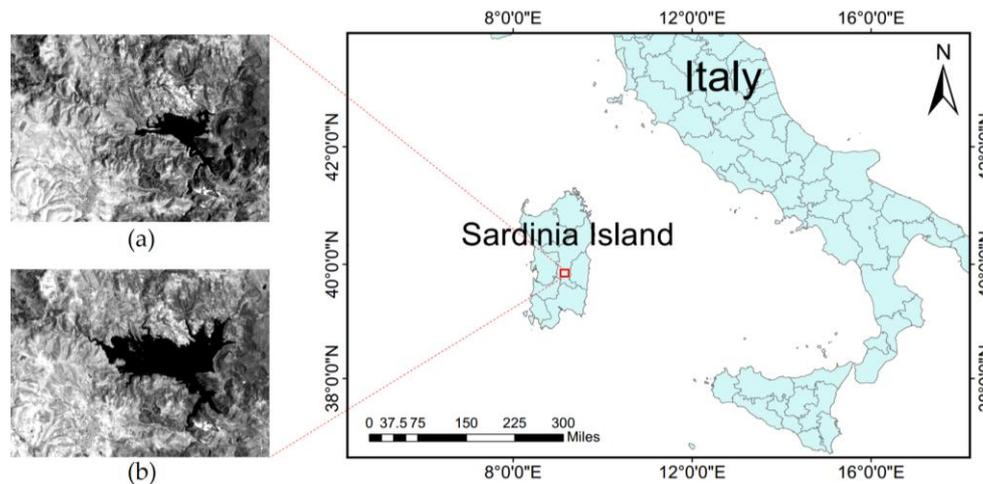
The Mexico dataset is an open-access dataset for change detection evaluation. It is made up of two images acquired by the ETM+ sensor of the Landsat-7 satellite in an area of Mexico on 18 April 2000 and 20 May 2002. The spatial resolution was 30 m. A section of  $512 \times 512$  pixels was selected as the test site. Between the two acquisition dates, forest fires destroyed a large proportion of the vegetation. Band 4 (spectral range from  $0.775 \mu\text{m}$  to  $0.900 \mu\text{m}$ ) was selected in our experiment, because band 4 of the ETM+ sensor is the near-infrared band, which is sensitive to the difference of vegetation type. Figure 4a,b shows the band 4 of the images 2000 and 2002, respectively. The reference map was manually defined according to a detailed visual analysis of the available multi-temporal images and the difference image. The reference map contained 25,599 changed pixels and 236,545 unchanged pixels. No radiometric correction was applied.



**Figure 4.** Mexico dataset. (a) Image acquired in April 2000. (b) Image acquired in May 2002.

##### 4.1.2. Sardinia Island

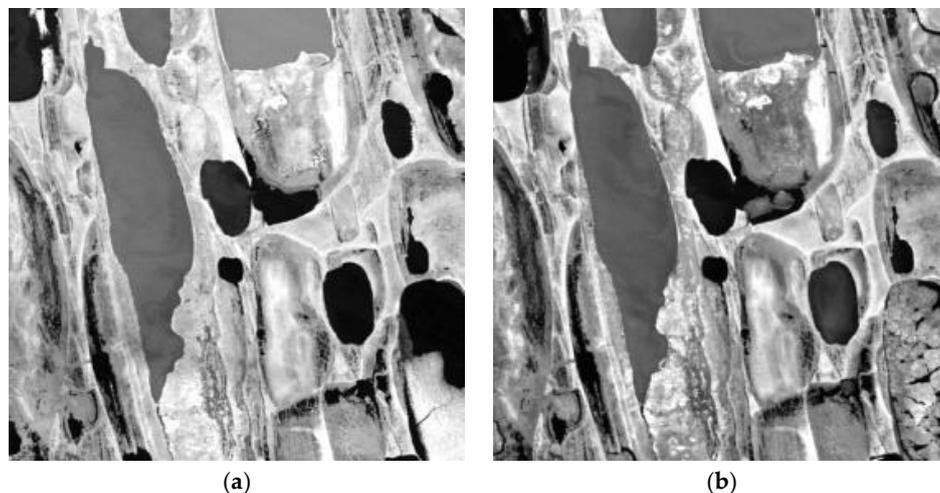
The second dataset tested in our experiment is the Sardinia island dataset, which is composed of two multispectral images acquired by the TM sensor of the Landsat-5 satellite in September 1995 and July 1996. The spatial resolution was 30 m. The test site was a section of  $300 \times 412$  pixels of a scene covered Lake Mulargia on the island of Sardinia. Between the two acquisition dates, the water level of the lake increased due to the flood. Band 4 (spectral range from  $0.760 \mu\text{m}$  to  $0.960 \mu\text{m}$ ) of the TM sensor is the near-infrared band, because water has a strong absorption effect in the near-infrared band, which makes the water profile clear in band 4. Figure 5a,b show the tested image 1995 and image 1996, respectively. The reference map was manually defined, in which 7480 changed pixels and 116,120 unchanged pixels were identified. No radiometric correction was applied to the multi-temporal images. The images were co-registered with 12 ground control points.



**Figure 5.** Sardinia island dataset. (a) Image acquired in September 1995. (b) Image acquired in July 1996.

#### 4.1.3. Alaska

The third dataset is the Alaska dataset, which is made up of two multispectral images acquired by the TM sensor of the Landsat-5 satellite in July 1985 and July 2005. The spatial resolution was 30 m. Band 4 was selected for the experiment for the same reasons mentioned above. A section of  $400 \times 400$  pixels was selected for the experiment. Figure 6a,b display the tested images. The reference map contains 9741 changed pixels and 150,259 unchanged pixels. No radiometric correction was applied to the multi-temporal images. The former image was registered to the latter one.

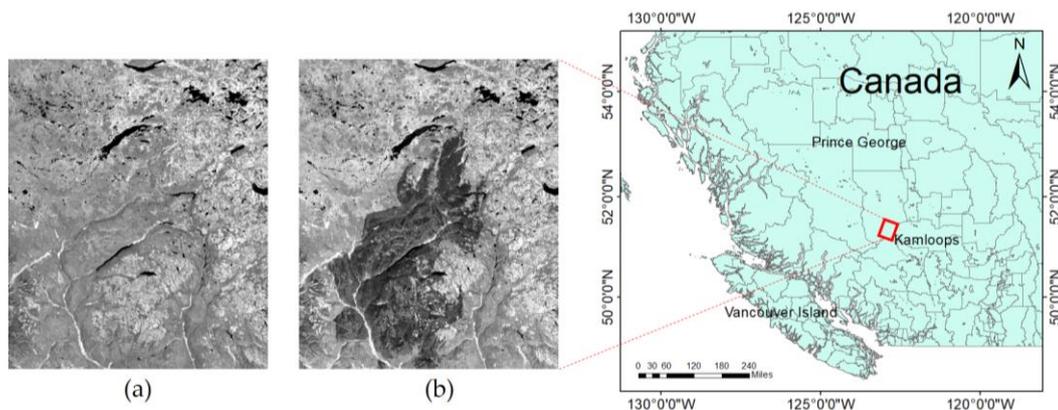


**Figure 6.** Alaska dataset. (a) Image acquired in July 1985. (b) Image acquired in July 2005.

#### 4.1.4. Canada

The Canada dataset is composed of two multispectral images acquired by the Landsat-8 OLI sensor on 2 July 2017 and 22 August 2017. The spatial resolution was 30 m. The test site is a scene near Prince George, British Columbia, Canada. The size is  $3000 \times 2500$  pixels, which is much bigger than the above three datasets. The forest fires lasted more than two weeks, which led to the suspension of three major timber producers, and more than 14,000 people had to be evacuated. In order to accurately detect the disaster area in the forest, we selected band 5 for our experiments. Band 5 (spectral range from  $0.845 \mu\text{m}$  to  $0.885 \mu\text{m}$ ) belongs to the near-infrared band, which is sensitive to changes in vegetation and suitable for the detection of fire scars. Figure 7a and (b) show band 5 of the multi-temporal images. According to a detailed visual analysis, 1,795,840 changed pixels and 5,704,160 unchanged pixels were

identified in the reference map. No radiometric correction or images registration was applied to the bi-temporal images.



**Figure 7.** Canada dataset. (a) Image acquired on 2 July 2017. (b) Image acquired on 22 August 2017.

#### 4.2. Parameter Setting

The proposed method is composed of two main steps: multi-feature space construction and DE clustering-based CD. In the first step, three image features are extracted and the parameters are adjusted by trial and error. The window size of the Wiener filtering was set to be 13, 13, 3, and 13 for the Mexico, Sardinia, Alaska, and Canada datasets, respectively. SSIM illustrates a locally isotropic property, and we used a  $53 \times 53$  circular symmetric Gaussian weighting function with a standard deviation of one sample for all of datasets. In the second step, the pixels in the multi-feature space are classified into either changed or unchanged classes using the DE algorithm. The population size for all tests is set to be 30, and the initial scale factor  $F_0$  and the initial crossover rate  $CR_0$  are set to be 0.8 and 0.2, respectively. The maximum number of generations is set to be 100. The fuzzy component  $m$  in the objective function is set to be 4.1, 2.0, 2.0, and 3.0 for the Mexico, Sardinia, Alaska, and Canada datasets, respectively.

For the fuzzy clustering-based CD methods, i.e., FCM, FLICM, GA-FCM, and MOEA/D, their fuzzy components are the same as the proposed M-DECD. For the EAs-based CD methods, i.e., GA-mask, GA-FCM and MOEA/D, the population size is set to be 30. The maximum number of iterations of GA-FCM and MOEA/D is set to be 100. Especially, the maximum number of iterations of GA-mask is set to be 100,000, because the GA-mask method directly encodes the whole binary CM, resulting in slow convergence. In contrast, the GA-FCM, MOEA/D and M-DECD encode the CD clusters, whose length is much lower than the binary CM. Thus, these methods can converge in 100 generations.

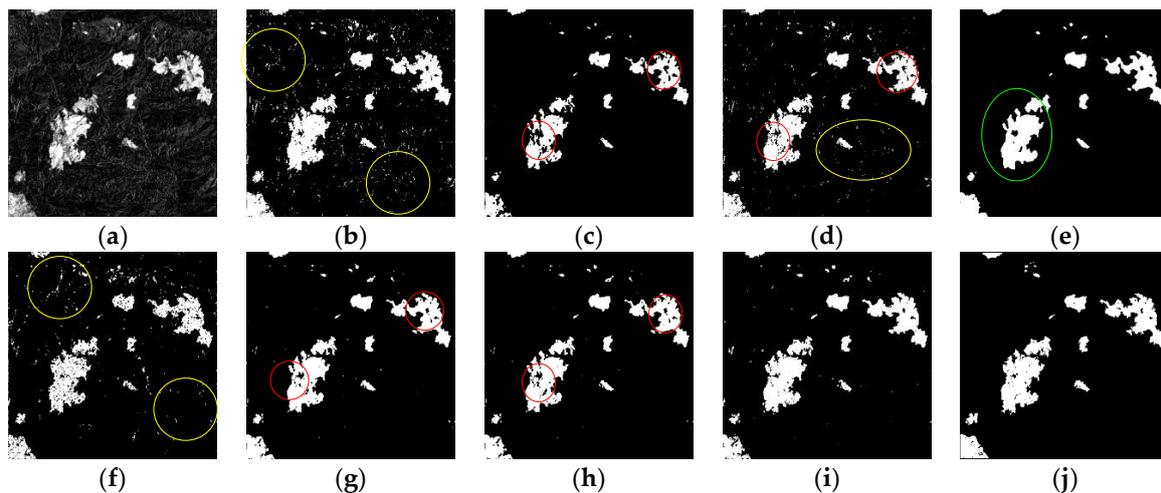
#### 4.3. Experimental Results

##### 4.3.1. Results of the Mexico Dataset

The CMs of the experiments tested on the Mexico dataset are shown in Figure 8. From Figure 8a, we can see that the DI of the Mexico dataset is relatively simple, because the change area is obvious, and the background interference is small. Therefore, the CD of the Mexico dataset is not very challenging. Therefore, the tested methods obtained similar results.

As shown in Figure 8, compared with the reference map (see Figure 8j), we see that many white noise spots exist in the CMs of KI (see Figure 8b), FCM (see Figure 8d), and GA-mask (see Figure 8f)), because these methods assume that the pixels are independent of each other and only consider their spectral information. It should be mentioned that GA-mask is very time-consuming due to its mask-based encoding strategy, and it took about 8 h to process the Mexico dataset. In contrast, MRF+EM (see Figure 8c), FLICM (see Figure 8e), GA-FCM (see Figure 8g), and MOEA/D (see

Figure 8h) consider the influence of the neighborhood pixels, so these methods are more robust to background interference, but they lose some changed details. For example, some weak changed regions in the middle of the DI are not detected. The result of the proposed method is shown in Figure 8i, where M-DECD shows better detail-preserving capability for the weak changes, due to the detail enhancement feature. In addition, the proposed method has fewer false detections, due to the Wiener denoising feature. However, we did find that the detail enhancement feature enhances background interference, so some pixels are wrongly identified as changed in the left of the CM. As to the algorithm efficiency, the proposed method takes about 3 min for the Mexico dataset. The DE algorithm can converge in 100 iterations and obtain a satisfactory result.

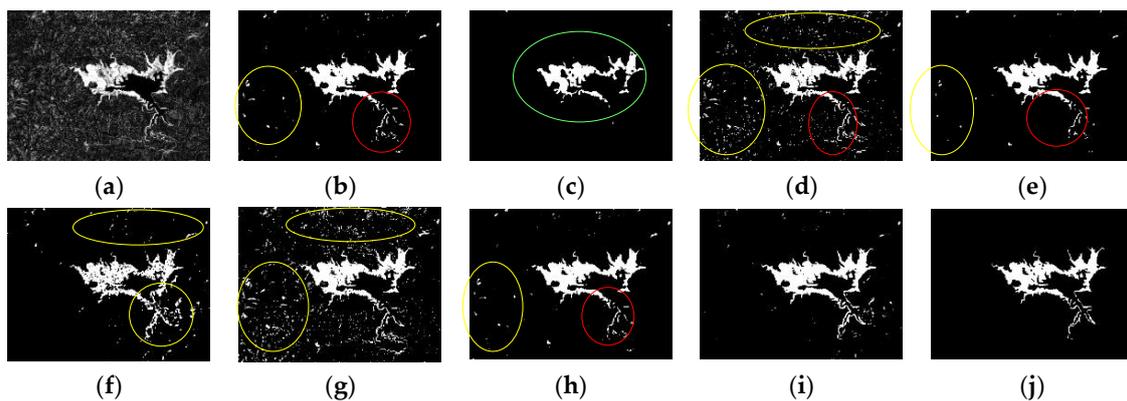


**Figure 8.** Results of the Mexico dataset. (a) difference image. (b) KI. (c) MRF+EM. (d) FCM. (e) FLICM. (f) GA-mask. (g) GA-FCM. (h) MOEA/D. (i) Proposed M-DECD. (j) Reference map. Yellow ovals refer to the false detection areas of the comparison algorithms compared with the M-DECD, and the red ovals refer to the miss-detection areas of the comparison algorithms. Green ovals refer to over-smoothing region.

#### 4.3.2. Results of the Sardinia Dataset

The CMs of the Sardinia dataset are shown in Figure 9. We can see that the DI (see Figure 9a) of the Sardinia dataset has some background interference caused by the Gaussian noise. In addition, the lake boundary is weak and slim. Therefore, it is a challenging work to remove the noise while preserving the lake boundary.

We can see that some methods cannot resist the effect of the noise due to ignoring of spatial-contextual information, such as KI (see Figure 9b), FCM (see Figure 9d), and GA-mask (see Figure 9f), in which many pixels are wrongly detected as changed. MRF+EM (see Figure 9c) shows a good performance in noise removal, but its CM appears over-smoothing and the weak changed boundary of the lake is not detected. Although the average of the neighborhood pixels has been considered, there are some small false detected regions in the CMs of FLICM (see Figure 9e), GA-FCM (see Figure 9g), and MOEA/D (see Figure 9h). The result of the proposed method is shown in Figure 9i; it was found that the M-DECD can not only effectively resist background interference and generate a clean CM, but also retain the weak changed boundary of the lake, especially in the red oval regions.

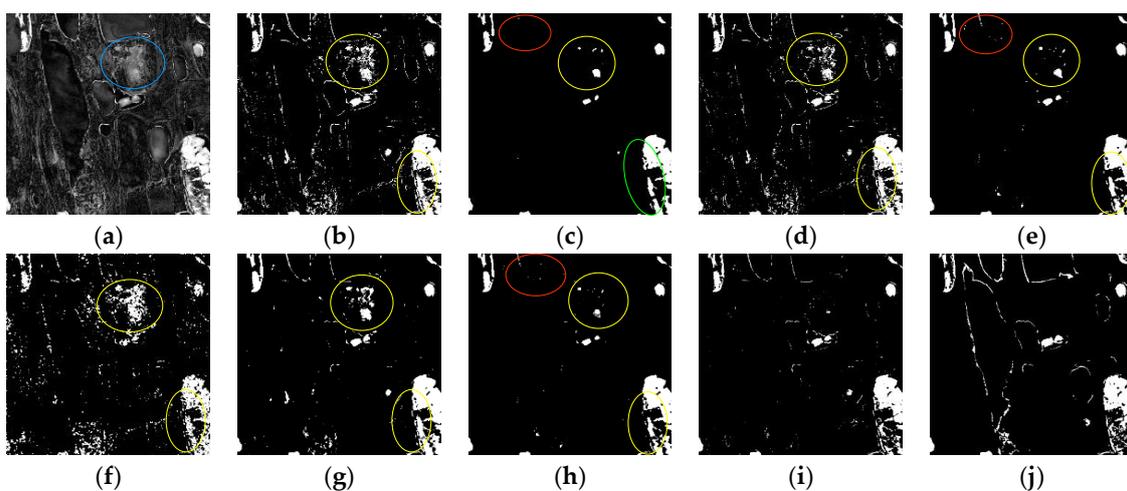


**Figure 9.** Results of the Sardinia dataset. (a) Difference image. (b) KI. (c) MRF+EM. (d) FCM. (e) FLICM. (f) GA-mask. (g) GA-FCM. (h) MOEA/D. (i) Proposed M-DECD. (j) Reference map.

#### 4.3.3. Results of the Alaska Dataset

The results of the Alaska dataset are shown in Figure 10. In Figure 10a, we see that the Alaska dataset has complex background interference, caused by both Gaussian noise and local brightness distortion (marked by the blue ovals). In addition, there are many weak and slim edges. Therefore, CD of the Alaska dataset is quite challenging.

As shown in Figure 10, compared with the reference map (see Figure 10j), we see that some methods are edge-sensitive, including KI (see Figure 10b), FCM (see Figure 10d), and GA-mask (see Figure 10f). However, these methods are also noise-sensitive, and many pixels are wrongly detected as changed. MRF+EM (see Figure 10c), FLICM (see Figure 10e), GA-FCM (see Figure 10g), and MOEA/D (see Figure 10h) can reduce some false alarms by considering the spatial information of the neighborhood pixels; however, the spatial smoothness may blur the edges and it cannot resist local brightness distortion. Therefore, these methods misjudge some pixels in brightness distortion regions. Some methods lose many weak changed edges due to over-smoothing (MRF+EM, FLICM, and MOEA/D). The results of the M-DECD are shown in Figure 10i; it was found that the M-DECD can resist the effect of Gaussian noise and generate a clean CM due to the Wiener de-noised feature. Moreover, the M-DECD performs well in brightness distortion regions due to the structural similarity feature.

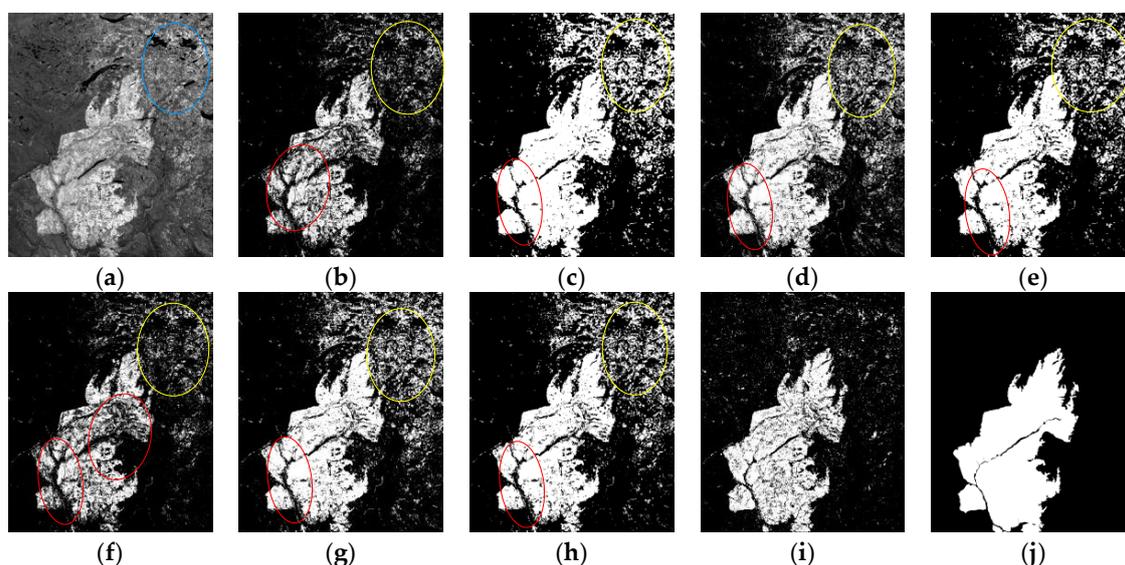


**Figure 10.** Results of the Alaska dataset. (a) Difference image, the blue ovals refer to the local brightness deformation region. (b) KI. (c) MRF+EM. (d) FCM. (e) FLICM. (f) GA-mask. (g) GA-FCM. (h) MOEA/D. (i) Proposed M-DECD. (j) Reference map.

#### 4.3.4. Results of the Canada Dataset

The CMs of the experiments tested on the Canada dataset are shown in Figure 11. First, the size of the Canada dataset is larger than the other three datasets. In addition, from Figure 11a, we see that the Canada dataset has strong background interference, caused by both Gaussian noise and local brightness distortion (marked by the blue ovals), especially in the top right corner, where has severe local brightness distortion. Moreover, the spectral contrast between the changed and unchanged regions is low. Therefore, CD of the Canada dataset is the most difficult in our experiments.

As shown in Figure 11, compared with the reference map (see Figure 11j), we can see that the advantage of the proposed M-DECD is obvious. Some methods are sensitive to the image noise and local brightness distortion, including KI (see Figure 11b), FCM (see Figure 11d), and GA-mask (see Figure 11f). However, some other methods can resist the image noise, including MRF+EM (see Figure 11c), FLICM (see Figure 11e), GA-FCM (see Figure 11g), and MOEA/D (see Figure 11h). However, they cannot resist local brightness distortion, and many regions are wrongly detected as changed, especially in the top right corner of the CMs. The result of the M-DECD is shown in Figure 11i; it was found that the M-DECD performs well in brightness distortion regions, while preserving the changed details.



**Figure 11.** Results of the Canada dataset. (a) Difference image, the blue ovals refer to the local brightness deformation region. (b) KI. (c) MRF+EM. (d) FCM. (e) FLICM. (f) GA-mask. (g) GA-FCM. (h) MOEA/D. (i) Proposed M-DECD. (j) Reference map.

#### 4.3.5. Quantitative Analysis

The quantitative evaluation of the experimental results is listed in Table 4. The optimal results of each evaluation index are marked in bold, and the suboptimal results are underlined.

Among the four datasets, CD of the Mexico dataset is the easiest, and CD of the Sardinia dataset is more difficult due to background noise and weak changed edges. The OE and kappa of M-DECD are very close to the best results, because the FA of M-DECD is a little higher than the best result due to edge expansion, which is caused by the detail enhancement feature. The performance of the M-DECD could be further improved by adopting a more appropriate edge detection method.

CD of the Alaska dataset is quite challenge, due to the complex background interference caused by Gaussian noise and local brightness distortion, and there are many weak changed edges. CD of the Canada dataset is the most difficult, due to its large size and strong brightness distortion caused by illumination difference. The accuracies of the M-DECD on the Alaska dataset and the large Canada

dataset increased by about 3.3% and 11.9%, respectively. This indicates that the multi-feature space is suitable for Landsat images and the DE clustering-based CD method is effective.

**Table 4.** Quantitative evaluation of the experimental results obtained with four datasets.

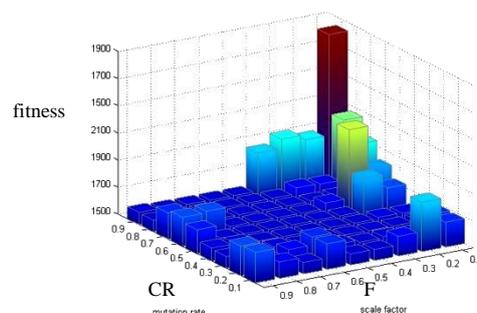
Dataset	Index	KI	MRF+EM	FCM	FLICM	GA-Mask	GA-FCM	MOEA/D	M-DECD
Mexico	MA	<b>1918</b>	6103	3236	3320	4681	3270	2795	<u>2114</u>
	FA	5389	<b>429</b>	2357	1014	3453	1049	<u>1247</u>	2068
	OE	7308	6532	5593	4334	8134	4319	<b>4042</b>	<u>4182</u>
	kappa	0.8508	0.8430	0.8770	0.9022	0.8200	0.9027	<b>0.9101</b>	<u>0.9094</u>
Sardinia	MA	<u>739</u>	2266	910	1038	1478	1004	877	<b>485</b>
	FA	5448	<b>250</b>	3834	1078	2617	1408	<u>1392</u>	1944
	OE	6187	2516	4744	<b>2116</b>	4095	2412	<u>2269</u>	2429
	kappa	0.6645	0.7994	0.7189	<b>0.8525</b>	0.7325	0.8355	<u>0.8463</u>	<u>0.8442</u>
Alaska	MA	<u>2160</u>	4117	2378	3505	3927	<b>2116</b>	3644	3095
	FA	6114	<u>1064</u>	4550	1447	7076	2853	1500	<b>984</b>
	OE	8274	5181	6928	<u>4952</u>	9741	4969	5144	<b>4079</b>
	kappa	0.6199	0.6682	0.6571	0.6997	0.4776	<u>0.7184</u>	0.6866	<b>0.7519</b>
Canada	MA	823,031	<b>189,863</b>	406,428	293,202	831,436	778,996	338,566	386,034
	FA	<u>331,124</u>	1,002,245	855,901	<u>722,737</u>	325,081	364,579	723,305	<b>205,690</b>
	OE	1,154,155	1,192,108	1,262,329	1,015,939	1,156,517	1,143,575	1,061,871	<b>591,724</b>
	kappa	0.5337	0.6222	0.5744	<u>0.6563</u>	0.5314	0.6120	0.6379	<b>0.7757</b>

In addition, we can find that the CD results of the EAs clustering-based CD methods (e.g., GA-FCM, MOEA/D, and M-DECD) perform better than those conventional clustering-based CD methods (FCM, FLICM), which demonstrates the promising optimization capability of the EAs for clustering-based CD. The M-DECD obtains stable results with high accuracy in all of the tested datasets, which proves that adaptive DE algorithm is robust and effective in searching the optimal CD results.

#### 4.4. Experiment Analysis

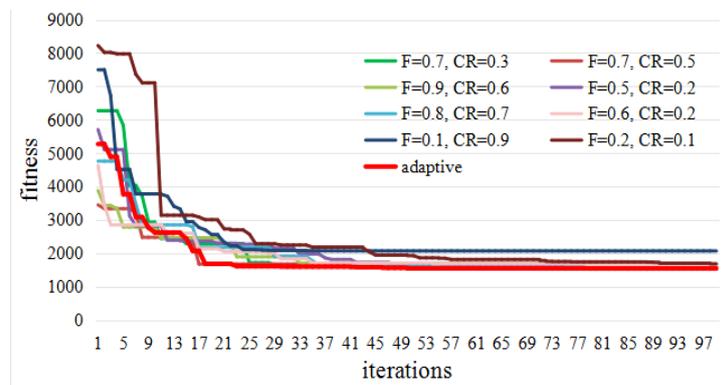
##### 4.4.1. Analysis of the Self-Adaptive Strategy

The proposed M-DECD adopted a self-adaptive parameter adjustment strategy [42] for DE. In order to prove the significance of the adaptive strategy, the relationship between the DE performance and different combinations of control parameters was analyzed. Without loss of generality, the experimental result obtained with the Alaska dataset is presented in Figure 12. The abscissas are parameters  $CR$  and  $F$ , respectively, and the ordinate is the fitness value of the optimal solution. It can be seen that the different combinations of the control parameters have a significant influence on the optimization result. DE with an inappropriate combination of  $CR$  and  $F$  converges to the local minima, which will affect the detection result. Therefore, it is crucial to use an adaptive strategy to ensure that DE converges to the global optimal solution.



**Figure 12.** Relationship between the search performance (fitness) and different combinations of control parameters ( $F$  and  $CR$ ).

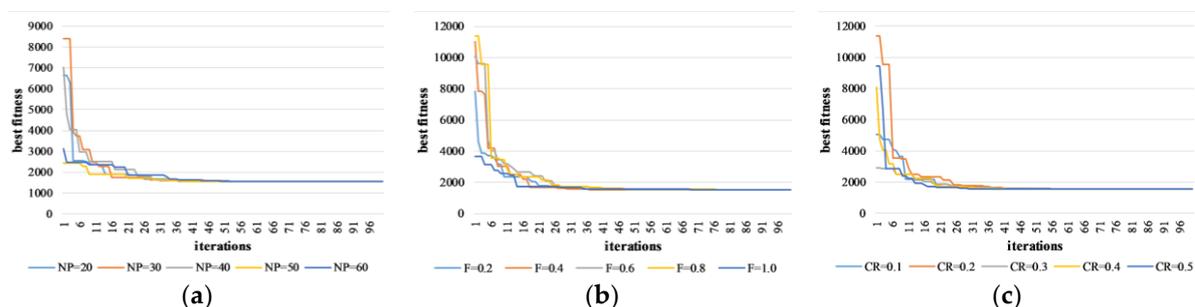
In order to prove the effectiveness of the adopted self-adaptive strategy, we analyzed the relationship between the algorithm convergence process and different combinations of control parameters, as shown in Figure 13. The abscissa is the number of DE evolutionary iterations, and the ordinate is the fitness value of the optimal solution in each generation. It can be concluded that the adopted strategy (red line) can not only improve convergence speed, but also enhance the robustness of the algorithm.



**Figure 13.** Relationship between the search performance (fitness) and different combinations of control parameters ( $F$  and  $CR$ ).

#### 4.4.2. A Sensitivity Analysis on Adaptive DE Algorithm

There are three parameters that have to be set before the adaptive DE-based CD step, i.e., population size  $NP$ , initial scale factor  $F_0$ , and initial crossover rate  $CR_0$ . In order to prove the robustness of the adaptive DE, we tested the influence of these parameters. Without loss of generality, the tested results of Alaska dataset are shown in Figure 14. For demonstration purposes, after each generation, we recorded the fitness of the best individual as the Y-axis, i.e., the  $Jm$  distance of the best clusters, and the X-axis is the parameter of DE. Population size is an important parameter, which affects the distribution of population and convergence behavior. We fixed  $F_0 = 0.8$  and  $CR_0 = 0.2$ , then recorded the performance of DE at different population sizes, as shown in Figure 14a. Further, we fixed  $NP = 30$  and  $CR_0 = 0.2$ , and recorded the performance of DE with various initial scale factors, as shown in Figure 14b. Furthermore, Figure 14c depicts the performance of DE with different initial crossover rates, where  $NP = 30$  and  $F_0 = 0.8$ .



**Figure 14.** Relationship between the performance and different parameters of DE. (a) Performance with different population size  $NP$ . (b) Performance with different initial scale factor  $F_0$ . (c) Performance with different initial crossover rate  $CR_0$ .

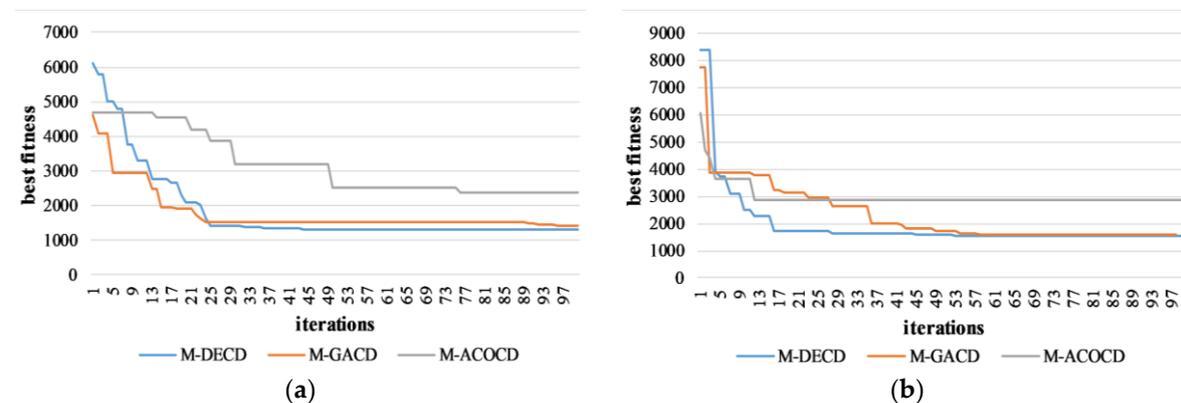
From the results, it can be concluded that the performance of the adaptive DE algorithm is very stable. The population size  $NP$ , initial scale factor  $F_0$ , and initial crossover rate  $CR_0$  have little effect on convergence behavior. On one hand, the optimization of CD clusters is not difficult owing to the low-dimensional encoding strategy. On the other hand, the parameter self-adaptive strategy can

dynamically adjust the scale factor and the crossover rate, and improve the search efficiency of DE. Therefore, the adaptive DE algorithm can converge to satisfactory optimization results in 100 iterations under various parameters.

#### 4.4.3. Performance Comparison of GA, ACO, and DE

The DE algorithm is employed to search for the optimal changed and unchanged clusters with minimum  $J_m$  distance. However, DE is just a case study of the proposed optimal multi-feature clustering-based CD framework, and can be replaced by other EAs with powerful optimization ability. In this section, we further test the performance of other EAs, i.e., genetic algorithm (GA [43]), ant colony optimization algorithm (ACO [44]) in the proposed framework, namely M-GACD and M-ACOD. The purpose is to compare the convergence process of GA, ACO, and DE algorithms using the proposed objective function. In the GA algorithm, considering that the individual is encoded with real-value, we use intermediate recombination and real-value mutation operators [43] to produce new individuals. As for the ACO algorithm, we adopt the real-value mutation operator to construct new solutions, and the transition probability and pheromone are updated according to the basic ACO [44,45]. These EAs randomly generate  $NP$  clusters, and evolve the clusters through different evolutionary operators. For fairness, their population size is set to 30, and the maximum number of generations is set to 100. The Sardinia dataset and the Alaska dataset are selected as examples, and Figure 15 depicts the convergence process of the compared EAs. The abscissa is the number of generations, and the ordinate is the best fitness of the optimal clusters in each generation.

From the results of the Sardinia dataset and the Alaska dataset, we find that the initial populations of GA and ACO are better than DE, because the populations are randomly generated. However, eventually, the DE algorithm converges to the best solution with minimal fitness, which demonstrates that DE has a powerful global search capability. The GA algorithm can obtain a better solution than the ACO algorithm. Since the ACO algorithm is more suitable for combinatorial optimization, such as the traveling salesman problem [46], the proposed CD problem is a continuous optimization problem, so the ACO algorithm tends to fall into the local minima, resulting in failure to find optimal clusters.



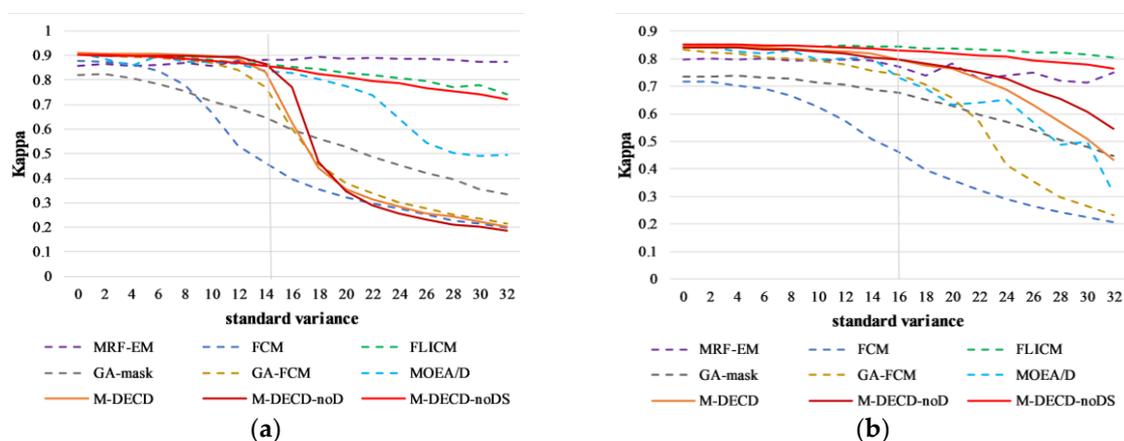
**Figure 15.** Convergence analysis of GA, ACO, and DE in the multi-feature clustering-based CD framework. (a) Convergence process of different EAs for the Sardinia dataset. (b) Convergence process of different EAs for the Alaska dataset.

#### 4.4.4. Robustness Test

Taking into consideration interferences in the Landsat image, three features are extracted to construct a multi-feature space—detail enhancement, Wiener de-noising, and structural similarity. In this section, we test the robustness of the proposed M-DECD in resisting the interferences of the Landsat images, and analyze the effect of the three features.

The Mexico dataset and the Sardinia dataset are selected for the Gaussian noise robustness test, because the background interference of the two datasets is mainly composed of Gaussian

noise with little brightness distortion. The Gaussian noise, which obeys the normal distribution  $N(\mu, \sigma^2)$ ,  $\mu = 0, \sigma = 0, 2, \dots, 32$ , is added into the tested datasets. The accuracy (Kappa coefficient) of the comparison methods is recorded as the Y-axis, and the X-axis is the standard variance of the Gaussian noise, as shown in Figure 16. It can be found that the MRF-EM and FLICM methods can remove strong Gaussian noise, because their CD is conducted with spatial information derived from the neighborhood pixels in DI. Although the proposed M-DECD method has employed the Wiener filter to remove some Gaussian noise, the detail enhancement feature strengthens the interference when the noise is too strong ( $\sigma > 14$  for the Mexico dataset and  $\sigma > 16$  for the Sardinia dataset), resulting in reduced accuracy. Therefore, we further tested the M-DECD method without the detail enhancement feature and the M-DECD method only with the Wiener de-noise feature, namely M-DECD-noD and M-DECD-noDS, respectively. We found that the M-DECD-noDS can obtain similar results as the FLICM when the noise is strong. The M-DECD-noD performs better than M-DECD, but it still cannot resist strong image noise, which means the structural similarity feature also enhances the image noise to some extent. Thus, the proposed method could be further improved by combining the multiple features in a more appropriate way.



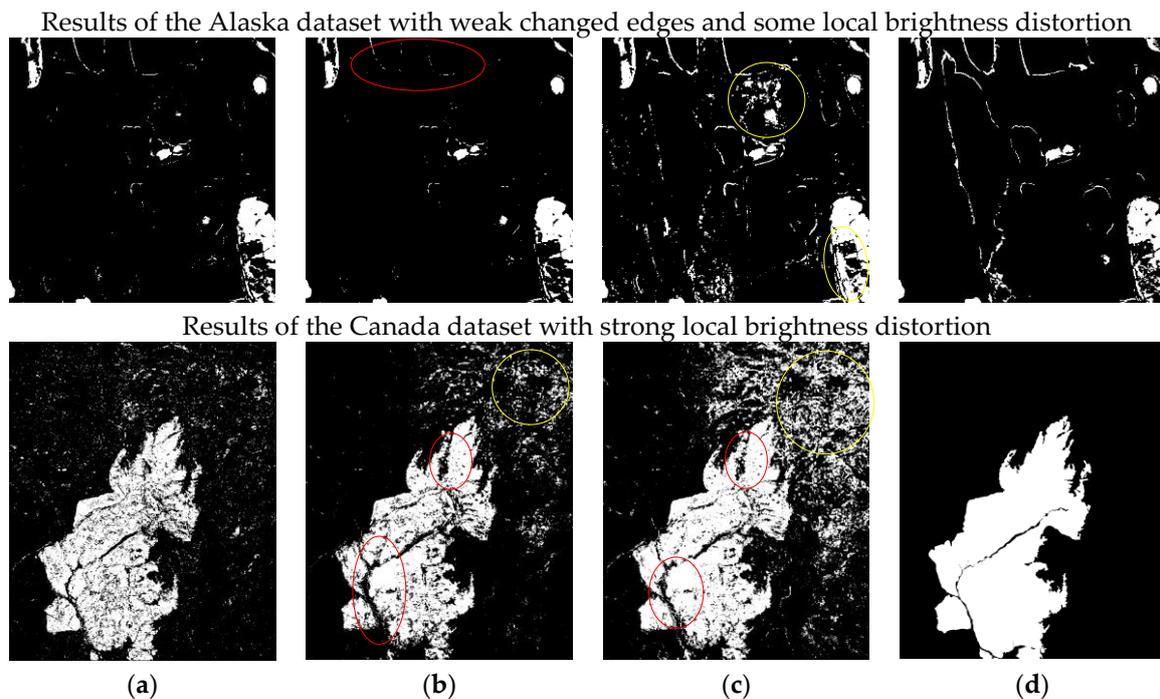
**Figure 16.** Gaussian noise robustness analysis of the comparison methods. (a) Accuracy of the Mexico dataset with respect to different noise level. (b) Accuracy of the Sardinia dataset with respect to different noise levels.

Furthermore, the Alaska dataset and the Canada dataset are tested to prove the effectiveness of the detail enhancement feature and the structural similarity feature, because the Alaska dataset has many weak changed edges and some local brightness distortion, while the Canada dataset is affected by stronger local brightness distortion (marked by the blue ovals in Figure 11a). For demonstration, we tested the M-DECD method without the detail enhancement feature (M-DECD-noD) to analyze the effect of detail preservation, and we also tested the M-DECD without structural similarity feature (M-DECD-noS) to analyze the robustness of the algorithm to local brightness distortion. The CMs are depicted in Figure 17 and their accuracy is listed in Table 5.

The CMs of the Alaska dataset show that the M-DECD-noD lost some weak changed edges (marked by red oval), compared to the M-DECD, which indicates that the M-DECD has better detail preservation ability due to the detail enhancement feature. The multi-feature space of M-DECD-noS is composed of the detail enhancement feature and the Wiener de-noise feature; thus, its CM contains several changed details. However, without the structural similarity feature, the CM of M-DECD-noS has many false detected pixels in the brightness distortion region (marked by yellow oval). From the CMs of the Canada dataset, we find that the M-DECD-noD and the M-DECD-noS miss a lot of changed pixels, which prove that the detail enhancement feature and the structural similarity feature can preserve local details of the changed object. Since the spectral contrast between the changed and unchanged regions is low in the Canada dataset, it needs as much information as possible for the

accurate CD. The M-DECD-noD does not consider the spectral value of DI, so its CM has some false detected pixels in the right top corner. On comparing the CMs of the M-DECD and the M-DECD-noS, we can conclude that the structural similarity feature is effective in resisting strong local brightness.

The accuracy of the M-DECD and M-DECD-noD demonstrates that the detail enhancement feature can decrease the miss alarm, and the accuracy of the M-DECD and M-DECD-noS proves that the structural similarity feature can significantly decrease the false alarm, especially when the interference of images is mainly caused by local brightness distortion.



**Figure 17.** Results of the M-DECD series of methods with various combinations of image features. (a) CMs of the M-DECD. (b) CMs of the M-DECD-noD. (c) CMs of the M-DECD-noS. (d) Reference map. Yellow ovals refer to the false detection areas compared with the M-DECD, and the red ovals refer to the miss-detection areas compared with the M-DECD.

**Table 5.** Accuracy of the results obtained with various combinations of image features.

Dataset	Index	M-DECD	M-DECD-noD	M-DECD-noS
Alaska	MA	3095	3524	1925
	FA	984	624	5902
	OE	4079	4148	7827
	Kappa	0.7519	0.736619	0.640776
Canada	MA	386,034	458,155	310,156
	FA	205,690	230,708	795,020
	OE	591,724	688,863	1,105,176
	Kappa	0.7757	0.736385	0.629674

## 5. Discussion and Conclusions

In this paper, we proposed an unsupervised CD method based on multi-feature clustering using adaptive differential evolution (M-DECD) for Landsat imagery. The proposed method constructs an effective multi-feature space for Landsat image CD, which can resist Gaussian noise and local brightness distortion, while preserving weak changes. Three features are extracted from the multi-temporal images and DI—Wiener denoising, detail enhancement, and structural similarity. In order to obtain optimal CD results, the M-DECD combines the DE algorithm with FCM to search the

changed and unchanged clusters in the multi-feature space by minimizing the  $J_m$  distance. In addition, the control parameters of the DE algorithm have been adaptively adjusted to improve the robustness of M-DECD. Four Landsat datasets with typical natural land cover changes have been tested and the experiment results demonstrate the effectiveness and superiority of the M-DECD when compared with seven conventional and state-of-the-art methods, especially for the Canada dataset, whose detection accuracy is increased by 11.9%.

However, the proposed method treats all the features equally. It could be further improved by analyzing the noise level and brightness distortion level of the data, then incorporating the multiple image features in a more appropriate way. In addition, the detail enhancement feature strengthens image noise, so the M-DECD is not suitable for processing SAR image with strong speckle noise.

**Author Contributions:** All the authors made significant contributions to the work. Y.Z., M.S. and A.M. proposed the conception and designed the methodology.

**Funding:** This work was supported by National Key Research and Development Program of China under Grant No. 2017YFB0504202, National Natural Science Foundation of China under Grant Nos. 41622107 and 41771385, and Natural Science Foundation of Hubei Province in China under Grant No. 2016CFA029.

**Acknowledgments:** The authors would like to thank Liangpei Zhang for providing advice for experiments and the preparation of the paper. The authors would also like to thank the National Aeronautics and Space Administration (NASA) and the US Geological Survey (USGS) for acquiring and providing the Landsat images used in this study, and Zhang Hua (China University of Mining and Technology, Jiangsu, China) for sharing the Mexico, Sardinia, and Alaska datasets.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Singh, A. Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* **1989**, *10*, 989–1003. [[CrossRef](#)]
2. Bruzzone, L.; Prieto, D.F. An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. *IEEE Trans. Image Process.* **2002**, *11*, 452–466. [[CrossRef](#)] [[PubMed](#)]
3. Mishra, N.S.; Ghosh, S.; Ghosh, A. Fuzzy clustering algorithms incorporating local information for change detection in remotely sensed images. *Appl. Soft Comput.* **2012**, *12*, 2683–2692. [[CrossRef](#)]
4. Yoon, G.W.; Yun, Y.B.; Park, J.H. Change vector analysis: Detecting of areas associated with flood using landsat TM. *Int. Geosci. Remote Sens.* **2003**, *5*, 3386–3388.
5. Chen, W.; Moriya, K.; Sakai, T.; Koyama, L.; Cao, C.X. Mapping a burned forest area from landsat tm data by multiple methods. *Geomat. Nat. Hazards Risk.* **2016**, *7*, 384–402. [[CrossRef](#)]
6. Schroeder, T.A.; Wulder, M.A.; Healey, S.P.; Moisen, G.G. Detecting post-fire salvage logging from landsat change maps and national fire survey data. *Remote Sens. Environ.* **2012**, *122*, 166–174. [[CrossRef](#)]
7. Yavasli, D.D.; Tucker, C.J.; Melocik, K.A. Change in the glacier extent in turkey during the landsat ERA. *Remote Sens. Environ.* **2015**, *163*, 32–41. [[CrossRef](#)]
8. Yavariabdi, A.; Kusetogullari, H. Change detection in multispectral landsat images using multiobjective evolutionary algorithm. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 414–418. [[CrossRef](#)]
9. Bruzzone, L.; Prieto, D.F. Automatic analysis of the difference image for unsupervised change detection. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 1171–1182. [[CrossRef](#)]
10. Otsu, N. Threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
11. Kittler, J.; Illingworth, J. Minimum error thresholding. *Pattern Recognit.* **1986**, *19*, 41–47. [[CrossRef](#)]
12. Zanetti, M.; Bruzzone, L. A theoretical framework for change detection based on a compound multiclass statistical model of the difference image. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 1129–1143. [[CrossRef](#)]
13. Zanetti, M.; Bovolo, F.; Bruzzone, L. Rayleigh-rice mixture parameter estimation via em algorithm for change detection in multispectral images. *IEEE Trans. Image Process.* **2015**, *24*, 5004–5016. [[CrossRef](#)] [[PubMed](#)]
14. Peiman, R. Pre-classification and post-classification change-detection techniques to monitor land-cover and land-use change using multi-temporal landsat imagery: A case study on pisa province in italy. *Int. J. Remote Sens.* **2011**, *32*, 4365–4381. [[CrossRef](#)]

15. Ghosh, S.; Bruzzone, L.; Patra, S.; Bovolo, F.; Ghosh, A. A context-sensitive technique for unsupervised change detection based on hopfield-type neural networks. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 778–789. [[CrossRef](#)]
16. Bovolo, F.; Bruzzone, L.; Marconcini, M. A novel approach to unsupervised change detection based on a semisupervised svm and a similarity measure. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2070–2082. [[CrossRef](#)]
17. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM—The fuzzy c-means clustering-algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [[CrossRef](#)]
18. Krinidis, S.; Chatzis, V. A robust fuzzy local information c-means clustering algorithm. *IEEE Trans. Image Process.* **2010**, *19*, 1328–1337. [[CrossRef](#)] [[PubMed](#)]
19. Gong, M.G.; Zhou, Z.Q.; Ma, J.J. Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering. *IEEE Trans. Image Process.* **2012**, *21*, 2141–2151. [[CrossRef](#)] [[PubMed](#)]
20. Celik, T. Change detection in satellite images using a genetic algorithm approach. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 386–390. [[CrossRef](#)]
21. Ghosh, A.; Mishra, N.S.; Ghosh, S. Fuzzy clustering algorithms for unsupervised change detection in remote sensing images. *Inf. Sci.* **2011**, *181*, 699–715. [[CrossRef](#)]
22. Wang, Q.; Zhang, F.; Li, X. Optimal clustering framework for hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5910–5922. [[CrossRef](#)]
23. Zhong, Y.F.; Ma, A.L.; Ong, Y.S.; Zhu, Z.X.; Zhang, L.P. Computational intelligence in optical remote sensing image processing. *Appl. Soft Comput.* **2018**, *64*, 75–93. [[CrossRef](#)]
24. Zhong, Y.F.; Zhang, S.; Zhang, L.P. Automatic fuzzy clustering based on adaptive multi-objective differential evolution for remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2290–2301. [[CrossRef](#)]
25. Zhong, Y.F.; Ma, A.L.; Zhang, L.P. An adaptive memetic fuzzy clustering algorithm with spatial information for remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1235–1248. [[CrossRef](#)]
26. Celik, T. Image change detection using gaussian mixture model and genetic algorithm. *J. Vis. Commun. Image Represent.* **2010**, *21*, 965–974. [[CrossRef](#)]
27. Celik, T.; Yetgin, Z. Change detection without difference image computation based on multiobjective cost function optimization. *Turk. J. Electr. Eng. Comput. Sci.* **2011**, *19*, 941–956.
28. Li, H.; Gong, M.G.; Wang, Q.; Liu, J.; Su, L.Z. A multiobjective fuzzy clustering method for change detection in sar images. *Appl. Soft Comput.* **2016**, *46*, 767–777. [[CrossRef](#)]
29. Aghababae, H.; Tzeng, Y.C.; Amini, J. Swarm intelligence and fractals in dual-pol synthetic aperture radar image change detection. *J. Appl. Remote Sens.* **2012**, *6*, 063596. [[CrossRef](#)]
30. Kusetogullari, H.; Yavariabdi, A.; Celik, T. Unsupervised change detection in multitemporal multispectral satellite images using parallel particle swarm optimization. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2151–2164. [[CrossRef](#)]
31. Singh, K.K.; Mehrotra, A.; Nigam, M.J.; Pal, K. Unsupervised change detection from remote sensing images using hybrid genetic FCM. In Proceedings of the 2013 Students Conference on Engineering and Systems (Sces): Inspiring Engineering and Systems for Sustainable Development, Allahabad, India, 12–14 April 2013.
32. Shang, R.H.; Qi, L.P.; Jiao, L.C.; Stolkin, R.; Li, Y.Y. Change detection in sar images by artificial immune multi-objective clustering. *Eng. Appl. Artif. Intell.* **2014**, *31*, 53–67. [[CrossRef](#)]
33. Storn, R.; Price, K. Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* **1997**, *11*, 341–359. [[CrossRef](#)]
34. Das, S.; Mullick, S.S.; Suganthan, P.N. Recent advances in differential evolution—An updated survey. *Swarm Evol. Comput.* **2016**, *27*, 1–30. [[CrossRef](#)]
35. Das, S.; Suganthan, P.N. Differential evolution: A survey of the state-of-the-art. *IEEE Trans. Evol. Comput.* **2011**, *15*, 4–31. [[CrossRef](#)]
36. Wang, D.; Zhang, X.F.; Liu, Y.; Zhao, Z.W.; Song, Z.J. Remote sensing image denoising with iterative adaptive wiener filter. *Springer Proc. Phys.* **2017**, *192*, 361–370.
37. Lim, J.S. *Two-Dimensional Signal and Image Processing*; Prentice Hall: Englewood Cliffs, NJ, USA, 1990; 710p.
38. Stephane, M.; Charlotte, P. Primal sketch of image series with edge preserving filtering application to change detection. In Proceedings of the 2015 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp), Annecy, France, 22–24 July 2015.

39. Moser, G.; Serpico, S.B. Unsupervised change detection with high-resolution sar images by edge-preserving markov random fields and graph-cuts. *Int. Geosci. Remote Sens.* **2012**, 1984–1987. [[CrossRef](#)]
40. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
41. Loza, A.; Mihaylova, L.; Bull, D.; Canagarajah, N. Structural similarity-based object tracking in multimodality surveillance videos. *Mach. Vis. Appl.* **2009**, *20*, 71–83. [[CrossRef](#)]
42. Zhong, Y.F.; Zhang, L.P. Remote sensing image subpixel mapping based on adaptive differential evolution. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2012**, *42*, 1306–1329. [[CrossRef](#)] [[PubMed](#)]
43. Mühlenbein, H.; Schlierkamp-Voosen, D. Predictive models for the breeder genetic algorithm I. Continuous parameter optimization. *Evol. Comput.* **1993**, *1*, 25–49. [[CrossRef](#)]
44. Dorigo, M. Optimization, learning and natural algorithms. PhD Thesis, Politecnico di Milano, 1992.
45. Dorigo, M.; Blum, C. Ant colony optimization theory: A survey. *Theor. Comput. Sci.* **2005**, *344*, 243–278. [[CrossRef](#)]
46. Dorigo, M.; Caro, G.D. Ant colony optimization: A new meta-heuristic. In Proceedings of the 1999 Congress on Evolutionary Computation (CEC 99), Washington, DC, USA, 6–9 July 1999; Volume 2, pp. 1470–1477.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).