

## Firefly optimization algorithm for change detection in urban areas using remote sensing images

Mahdi Moradi<sup>1\*</sup>, Mahmoud Reza Sahebi<sup>1</sup>, Shaheen Ghayourmanesh<sup>2</sup>

<sup>1</sup> Department of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Tehran, Iran

<sup>2</sup> Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, NB, Canada

### Article history:

Received: 30 May 2018, Received in revised form: 15 October 2018, Accepted: 20 October 2018

### ABSTRACT

Urban areas experience rapid changes due to natural and manmade factors. Monitoring these changes is essential for urban planning, resource management, and updating geospatial information systems. Therefore, change detection is an interesting topic for researchers in the remote sensing field, especially with the availability of high spatial resolution images in recent years. However, the use of high-resolution imagery increases the variability within homogenous land-cover classes and leads to low-accuracy change detection results. To overcome this problem and to generate a more accurate change mask, several features have been used to extract spatial information from images. The firefly algorithm (FA), as one of the recently developed optimization algorithms is evaluated for finding the optimum subspace. The urban areas under study are Azadshahr (in Tehran province, Iran) and Shiraz (Iran). Two high-resolution images at two different time points were captured from each study area. To detect intra-class changes, a two-class classification of differential features was used, which also helps with the poor radiometric condition of the images (especially in Azadshahr images). The performance of FA was then compared with a particle swarm optimization (PSO) algorithm and a genetic algorithm (GA). The results show that FA outperformed both PSO and GA algorithms with an overall accuracy and kappa coefficient of [95.17%, 0.90] versus [93.45%, 0.87] and [91.03%, 0.82] in the Azadshahr study area, and [94.87%, 0.90] versus [94.44%, 0.89] and [93.16%, 0.86] in the Shiraz study area, respectively. The proposed methodology was also compared with the results of two other studies conducted on the Azadshahr area and outperformed them as well. To analyze the contribution and importance of each feature type in change detection results, three indices, i.e. Effectiveness, Partial Effectiveness and Overall Effectiveness, were introduced in this paper. The result shows that the features extracted from a grey level co-occurrence matrix and the features of other color spaces are the most effective features selected by FA to be used in change detection of high-resolution images. Moreover, these indices revealed the weakness of using only spectral information for change detection of high-resolution images.

### KEYWORDS

Firefly Algorithm  
Change Detection  
Particle Swarm  
Optimization Genetic Algorithm  
Optimization  
Effectiveness Index

### 1. Introduction

Land use and land cover (LULC) are not static phenomena, especially in urban areas, and are constantly affected by human interference and natural factors. In particular, rapid industrial and technological developments have greatly increased the speed of these changes, while geospatial information systems lag behind in updating the changes in their systems, resulting in loss of their credibility. So,

monitoring the changes and reflecting them in LULC information systems are essential for urban planning, resource management, and damage assessment that are dependent on up-to-date geospatial information systems (Du et al., 2012; Phalke, 2006).

In recent years, several techniques have been developed for change detection in urban areas using high-resolution imagery. Despite all these efforts, each technique and

\* Corresponding author

E-mail addresses: [m\\_moradi@kntu.ac.ir](mailto:m_moradi@kntu.ac.ir) (M. Moradi); [sahebi@kntu.ac.ir](mailto:sahebi@kntu.ac.ir) (M.R. Sahebi); [shaheen.gh@unb.ca](mailto:shaheen.gh@unb.ca) (Sh. Ghayourmanesh)

DOI: 10.22059/eoge.2018.263964.1029

method has its own advantages and disadvantages and choosing the most appropriate method and algorithm is not an easy task. Researchers have tried to categorize these different approaches based on different criteria. Hall and Hay (2003), categorized these methods into three main levels of pixels, features and objects. While Sui et al. (2008) categorized these methods into seven groups of direct comparison, classification, object-oriented method, model method, time-series analysis (TSA), visual analysis and hybrid method. Hussain et al. (2013), introduced categorization of pixel-based, object-based and spatial data mining. Mehrotra et al. (2014) introduced Algebra, transformation, classification, and other approaches as four major categories of change detection. Devi and Jij (2015), in similar categories provided two major groups of pixel-based and object-based. Considering all the above methods and categorizations, a generalized super categorization for the study of change detection of urban areas can be formed with four major categories of direct comparison-post classification pair, object based-pixel based pair, supervised-unsupervised pair and feature, texture and spatial information analysis based methods.

The first pair in the super categories of change detection is the direct comparison-post classification approach. Post-classification comparison methods separately apply classification techniques on each image and then, the change detection is carried out by comparing the corresponding pixels in the classified images (Yuan et al., 2005). The major advantage of a post-classification comparison method is that the images are separately categorized, thus the results are not influenced by the radiometric differences between the images. It also produces "from-to" change information. The remarkable point is that each image source and sensor type and the classification method used for each image can be different (Van Oort, 2007). This is one of the most common methods used in past studies (El-Hattab, 2016; Mosammam et al., 2016; Galdavi et al., 2013; Li et al., 2015). Recently, new types of post-classification comparisons have been introduced for change detection in urban areas (Bhatt et al., 2016; Wen et al., 2016; Huang et al., 2014), based on the main elements of urban areas such as water, soil, vegetation, and buildings. However, there are several limitations that should be considered when using a post-classification technique. In all post-classification techniques, the accuracy of the change detection depends on the combined accuracy of each classified images. For example, if each classification has a producer accuracy of 90%, the accuracy of the post-classification change detection will be about 81% (Stow et al., 1980). Also, in some cases, choosing the correct index parameters and specifying the required thresholds are challenging. In addition to these limitations, another limitation of post-classification techniques that has not been considered in previous researches is that they can only detect

inter-class changes and not intra-class changes in urban areas.

The second pair is defined based on the image analysis unit, i.e. object-based versus pixel-based approaches. The object-based methods were introduced with the advent of high-resolution sensors. In a recent change detection study, the object-based methods were shown to produce more interesting results than the pixel-based methods on high-resolution data, but their limitations has remained less explored (Desheng Liu & Xia, 2010; Hall & Hay, 2003).

In a study that assessed the advantages and limitations of these methods, Desheng Liu and Xia, (2010) indicated negative impacts of under-segmentation errors and the dependency of the object-based analysis accuracy on the segmentation scale. The first step in each object-based analysis is segmentation, which is negatively affected by poor radiometric and spectral conditions (Fu & Mui, 1981; Dey et al., 2010). In addition, specifying parameters such as shape, compactness and scale for segmentation is challenging and extremely reduces task automation (Jianhua Liu et al., 2017). On the other hand, one of limitations of the pixel-based methods is the salt-and-pepper effect on the results (Campagnolo & Cerdeira, 2006; de Jong et al., 2001; Gao & Mas, 2008), which can be removed by a post-processing operation (Lu et al., 2004). Therefore, the pixel-based method is still suitable for several applications such as change detection (Tewkesbury et al., 2015).

The third pair in the change detection super categories is the unsupervised versus supervised approaches. Unsupervised methods are based on automated computational frameworks in which choosing an initial parameter, such as desired number of classes, is important (Tewkesbury et al., 2015; Karantzalos, 2015; Lu et al., 2004). Most classical and even novel unsupervised change detection methods are used on low/medium-resolution images. For instance, Celik (2010) proposed a novel unsupervised change detection method that minimized the weighted sum of mean square error of changed and unchanged regions by genetic algorithm (GA). This method was tested on Envisat and Landsat images. Similarly, Kusotogullari et al. (2015) proposed a novel unsupervised change detection method based on parallel particle swarm optimization (PSO) and only tested it on Landsat images. The main reason for not using the unsupervised approaches in high-resolution data is that unsupervised methods produce less accurate results in sub-meter spatial resolution images (Hussain et al., 2013). Thus, when using high-resolution imageries, supervised methods get more attention for change detection in complex urban areas.

Other change detection methods for urban areas are based on features, textural and spatial information that are extracted from images. Increase of the intra-class variance and decrease of the inter-class variance in high-resolution images

can be reduced/removed using these information, such as grey-level co-occurrence matrix (GLCM) (Walter, 2004; Erener & Düzgün, 2009; He et al., 2011; Chen & Chen, 2016), edge (Paul et al., 2016), wavelet transform (Celik & Ma, 2011; Raja et al., 2013), Markov chain (Celik, 2009), and morphological profile (Dalla Mura et al., 2008; Volpi et al., 2013). Du et al. (2012) implemented both feature and decision-level fusions for change detection in urban areas through fusing Simple Differencing, Simple Ratio, Absolute Distance, Euclidian Distance, and Chi-Square Transformation algorithms. The experimental results of this study demonstrated that the feature-level fusion could effectively reduce omission errors, while the decision-level fusion was good at restraining commission errors. In addition, the feature-level detection gave significantly higher correctness rates and lower overall errors in building the change detection (Huang et al., 2014).

In summary, the only approaches in the super categories of change detection that are able to deal with the problems incurred in change detection of urban areas using high spatial imageries are those based on features, as well as textural and spatial information. However, considering that so many different types of spatial information and features can be extracted from an image, it is essential to use an optimization algorithm to come up with the best set of features for change detection in urban areas. Among the proposed optimization algorithms, the performance of metaheuristic algorithms, which benefit from both global and local searches, have been proven in numerous studies. These algorithms include evolutionary algorithms such as GA and swarm intelligence algorithms such as PSO (Yang, 2010). While the application and performance of GA and PSO have been evaluated in change detection of urban areas using high resolution images (Chen et al., 2016), adapting and evaluating novel promising optimization approaches need to be evaluated. The present study uses the firefly algorithm (FA), as one of the recently developed optimization algorithms, to increase the accuracy of change detection in urban areas using remote sensing images. The main goal of this study is to assess the ability of the new firefly optimization algorithm in change detection problem of urban areas using high resolution images for feature selection. The performance of this novel optimization algorithm was also compared with those of GA (intrinsically discrete) and PSO (inherently continuous). For this purpose, a comprehensive study was carried out on the features and information that could be extracted in the spatial and

frequency domains to solve the change detection problem. In addition, as there needs to be a measure for evaluating how much a feature contributes to the change detection results, three new indicators were introduced for evaluating the contribution of each feature in change detection (feature efficiency).

## 2. Case study and Methodology

### 2.1. Case study

A pair of multispectral high-resolution images captured over Azadshahr (Tehran province, Iran) on October 2006 by the QuickBird sensor (spatial resolution of 0.6m) and on August 2010 by the GeoEye sensor (spatial resolution of 0.55m) were used to assess the effectiveness of the proposed approach (Figure 1). Also, a pair of multispectral high-resolution images captured over Shiraz (Iran) on June 2011 by the worldview 2 sensor (spatial resolution of 0.4m) and July 2015 by the worldview 3 sensor (spatial resolution of 0.3m) was used (Figure 2). Only red, green, and blue bands of the images were used to implement the proposed method.

### 2.2. Methodology

Considering the poor radiometric conditions of the images and the occurrence of intra-class changes (especially in the first study area), in this paper, a feature-based supervised direct comparison method was used for change detection of the study area through a pixel-based strategy (Figure 3). Figure 4 shows the flowchart of the algorithm used in this study; following the pre-processing step, a set of ten features was extracted from each image to generate a binary change mask and to allow a comprehensive study on the features. Then, an optimum subspace selection process was conducted by the wrapper strategies that used an optimization algorithm for optimum subspace selection through normalizing differential features and parameter setting of the support vector machine (SVM). To reach the final binary mask, the shadows can be removed by the morphological shadow index (MSI), even though the main algorithm still performing desirably in its absence. Also, a connected component analysis was used for noise removal in post-processing. The contribution of each feature to the overall accuracy of the final change mask was calculated by three indices of Effectiveness ( $E$ ), Partial Effectiveness ( $P_E$ ), and Overall Effectiveness ( $O_E$ ). In the following, a brief description of each step is provided.

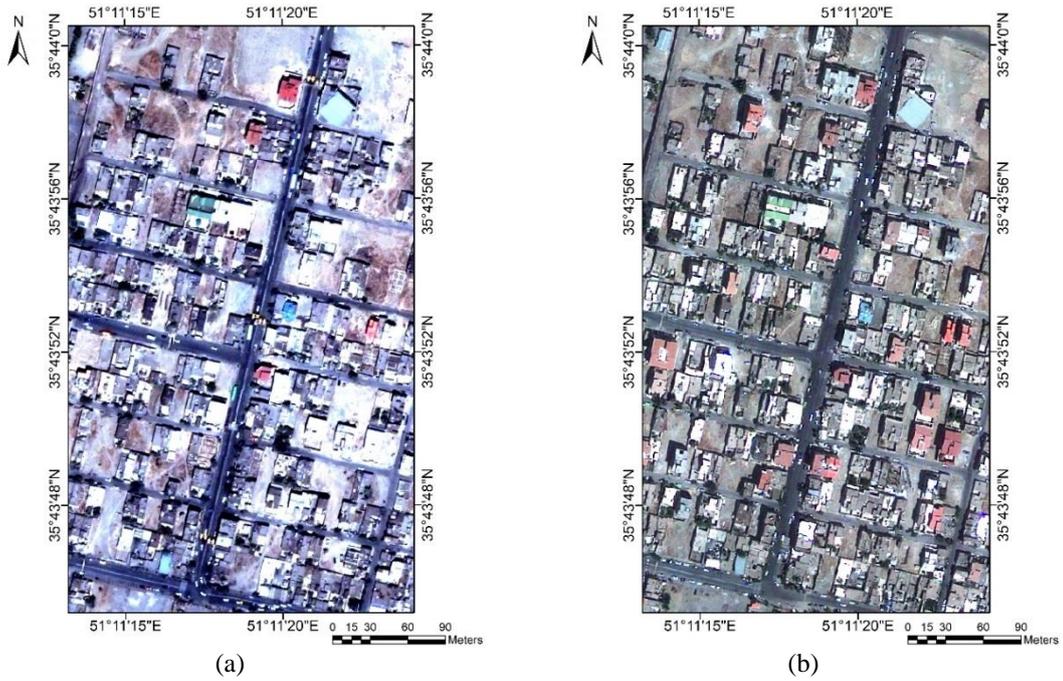


Figure 1. First study area (Azadshahr, Tehran), (a) October 2006 and (b) August 2010

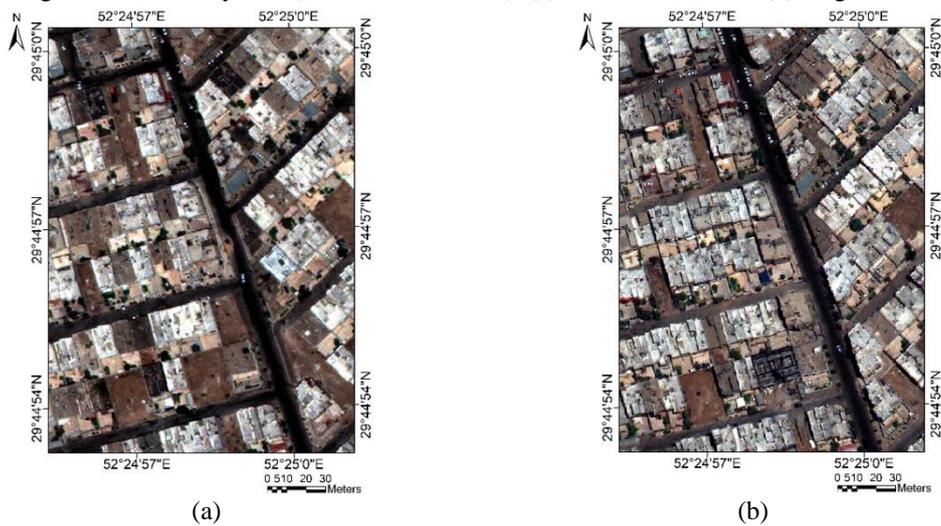


Figure 2. Second study area (Shiraz), (a) WV2 and (b) WV3



Figure 3. Sample of intra-class changes on Tehran, (a) October 2006, and (b) August 2010

### 2.2.1. Pre-processing

The first and one of the most important steps in any remote sensing analysis is the pre-processing step, which includes geometric and radiometric corrections. For the geometric correction, a first-order polynomial method with 11 tie points and the nearest neighbor resampling were used with 0.5 and 0.4 pixel RMS errors for Azadshahr and Shiraz, respectively. The histogram matching was applied for radiometric corrections.

### 2.2.2. Feature extraction

In addition to the spectral space feature, 9 other features are extracted from the spatial and frequency spaces, i.e. 10 features in total. These features include: 1- anomaly, 2- edge, 3- morphological building index (MBI), 4- other color

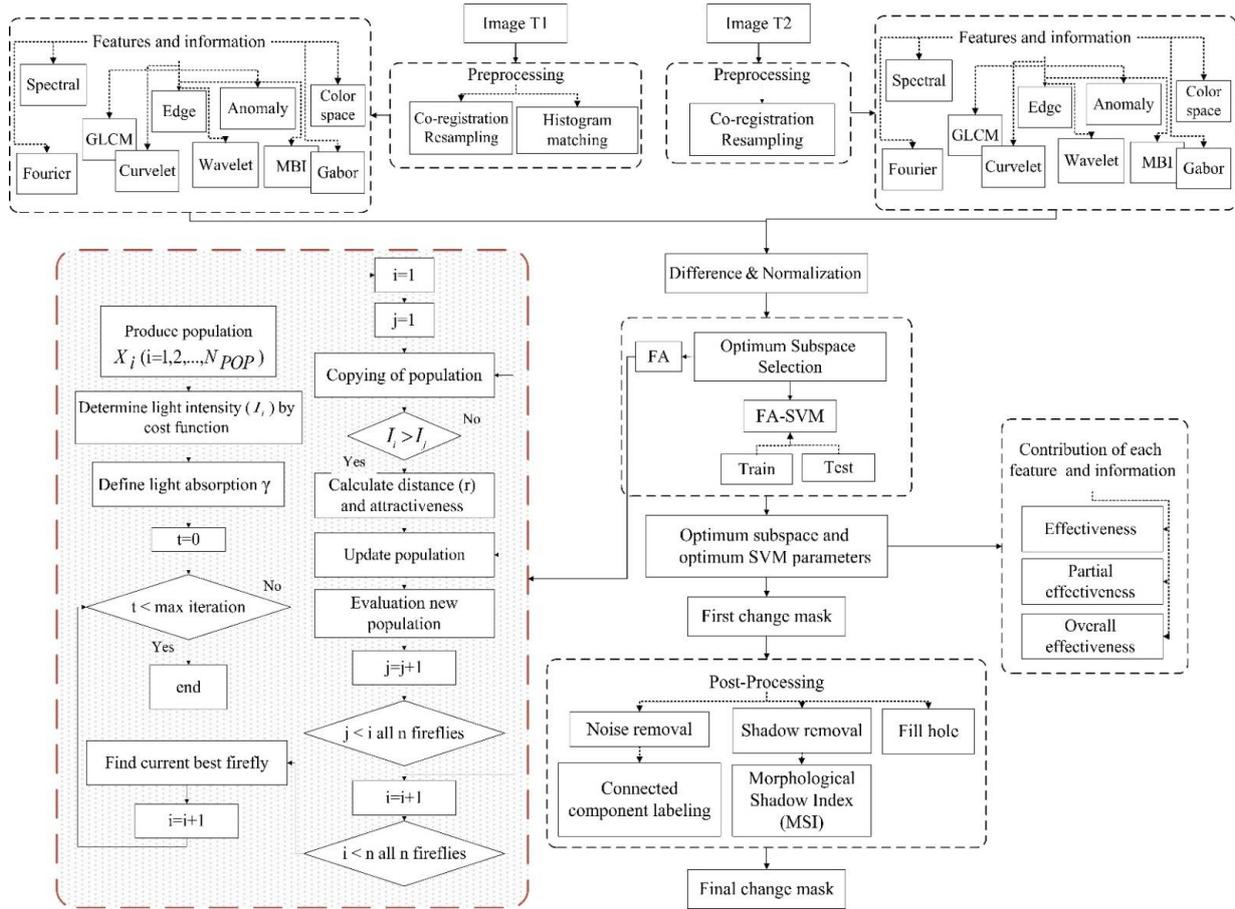


Figure 4. Flowchart of the change detection method

spaces, 5- GLCM, 6- features extracted from wavelet transform, 7- features extracted from Gabor filters, 8- features extracted from Fourier transform, and 9- features extracted from curvelet transform (summarized in Appendix).

### 2.2.3. Feature selection

After extracting features from images and calculating the differences between the features of the two images, an optimum combination of these differential features should be found for creating the binary change mask. This optimum combination can be found using a powerful optimization algorithm. In this paper, FA, one of the recently developed metaheuristic optimization algorithms, was evaluated and compared with PSO and GA.

#### 2.2.3.1. Firefly Algorithm

FA, proposed by Xin-She Yang, is a recent swarm intelligence metaheuristic optimization method that simulates the flashing pattern and characteristics of fireflies (Yang, 2009). In this algorithm, any firefly can be joined to any other brighter one. The light intensity  $I_i$  of the firefly  $i$  at position  $X_i$  in space is set to be associated with the objective function that needs to be optimized. Attraction  $\beta$  between two fireflies  $i$  and  $j$  can be then defined as  $\beta_0$  with  $r$  as the

distance between the two fireflies,  $\beta_0$  as the attractiveness at  $r=0$  and  $\gamma$  as light absorption coefficient, which controls the decrease of the light intensity, the most important parameter in the convergence rate (Łukasik & Żak, 2009). The initial values for  $\gamma$  and  $\beta_0$  and the population size are all set by the user. Eq. (1) formulates the movement of the firefly  $i$  toward the firefly  $j$ , in which  $X_i$  and  $X_j$  are their initial positions, and  $\epsilon_i$  is a vector of random numbers between 0 and 1. The first term in Eq. (1) is the initial position of firefly  $i$ , the second term is the amount of movement due to attraction, and the third term introduces randomization to the whole process, the amount of which is controlled by  $\alpha$ .

$$X'_i = X_i + \beta_0 e^{-\gamma r_{ij}^2} (X_j - X_i) + \alpha (\epsilon_i - \frac{1}{2}) \quad (1)$$

To measure the distance,  $r$ , between any two fireflies in a  $d$  dimensional space, different distance metrics can be used. In the original version of FA, a Cartesian or Euclidean distance is used. The hamming distance (Eq. (2)) that uses XOR operator ( $\oplus$ ) is used in binary FA (Zhang et al., 2016). Following the movement of the fireflies toward new positions based on Eq. (1), the result is probably a real number. To have the outputs conditioned in a binary format, a thresholding function, such as a sigmoid function (Eq. (3)) or a tangent hyperbolic function (Eq. (4)), can be used

(Crawford et al., 2014; Chandrasekaran et al., 2013). Then, the objective function is evaluated based on the new positions. Finally, using Eq. (5), in which rand is a random number uniformly distributed between 0 and 1, a new population is generated and the steps are repeated until a termination criterion is satisfied (for instance maximum number of the new population generation). The best solution is kept after each iteration.

$$r_{hamming} = 1 - \frac{\sum_{k=1}^n |X_i^k \oplus X_j^k|}{n} \quad (2)$$

$$sig = \frac{1}{1 + e^{-X_i}} \quad (3)$$

$$\tanh(|X_i|) = \frac{e^{2 \times X_i} - 1}{e^{2 \times X_i} + 1} \quad (4)$$

$$X_i(t) = \begin{cases} 1 & \text{if } rand < sig \text{ or } \tanh \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

### 2.2.3.2. Particle swarm optimization

PSO is a population-based optimization technique developed by Eberhart and Kennedy (1995). In every iteration, each particle's best position ( $p_{best}$ ) and the global best position of all particles ( $g_{best}$ ) is updated based on the changes in each particle's position ( $X$ ) and velocity ( $V$ ) (Zhen et al., 2008) :

$$V_i^k(t+1) = V_i^k(t) + c_1 r_1 (X_i^{pbest}(t) - X_i^k(t)) + c_2 r_2 (X_i^{gbest}(t) - X_i^k(t)) \quad (6)$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (7)$$

$$S(V_i^k) = 1 / (1 + e^{-V_i^k}) \quad (8)$$

$$X_i(t) = \begin{cases} 1 & \text{if } rand < S(V_i^k) \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

In Eqs. (6)-(9), rand is a random number between 0 and 1 and  $c_1, c_2$  are the two parameters that balance exploration and exploitation.

### 2.2.3.3. Genetic algorithm

GA is the most prominent technique in evolutionary computation and optimization algorithms introduced by John Holland in the 1960s for solving discrete problems. In this algorithm, candidate solutions are considered as chromosomes (strings). Each chromosome consists of genes (bits) with their values being used as unknown parameters. Some of these chromosomes make the first generation and the next generation is created using the current generation and through applying the crossover and mutation operators (Yu & Gen, 2010). The selection procedure, crossover and mutation operators are various and diverse. In this study, the

roulette wheel selection was used as the selection procedure, and the single point crossover, two-point crossover, and uniform crossover were used as the operators with weights of 0.1, 0.2 and 0.7, respectively. In addition, different values were used and evaluated for the probability of crossover ( $P_c$ ), probability of mutation ( $P_m$ ) and mutation rate ( $M_u$ ) of the GA.

### 2.2.3.4. Cost function and encoding

A supervised classification method was implemented for a two-class classification because of the poor and complex radiometric conditions of the images. The SVM is a fast and accurate supervised classification method that maps an input space into a higher dimensional space and is useful for nonlinear classification problems using kernel functions. The radial basis function is the kernel function used in this study due to its good accuracy and having only two main parameters (the penalty factor  $C$  and the kernel parameter  $\gamma$ ) (Lin et al., 2008). A cost function should be defined in a way that increases the overall accuracy of the SVM, while decreases the number of selected features ( $n_f$ ) from the input features ( $N$ ). The cost function, which is the same as the light intensity ( $I$ ) in the FA algorithm, is defined as Eq. (10):

$$I = (0.85 \times (1 - OA)) + (0.15 \times (n_f / N)) \quad (10)$$

The trial and error method was implemented providing a higher weight to the  $OA$  (85%) and a lower weight to the number of selected features (15%) in the cost function, defined in Eq. (10). In addition to the overall accuracy, the number of function evaluations ( $NFE$ ) is an important factor for the comparison of optimization algorithms. To have a fair comparison between all three optimization algorithms, i.e. FA, PSO, and GA, the initial population and its size ( $N_{pop}$ ) are set the same and the stopping criterion is set as a fixed number of iterations ( $It$ ). In terms of encoding, besides the need for increasing the accuracy with the presence or absence of features ( $N$ ) and decreasing the number of features in the subspace, the optimization of the two main parameters of SVM was also required ( $C$  is the SVM parameters and  $\gamma$  is the kernel parameter). Therefore,  $N+2$  decision variables had to be adopted for the encoding process. In this study, a binary version of FA was used and compared with the binary GA and PSO. When using the cost function, it is required to perform a binary to decimal conversion for the two encoded parameters of the SVM. Fireflies, particles, and chromosomes of FA, PSO and GA were all encoded as depicted in Figure 5.

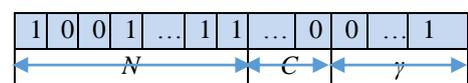


Figure 5. encoding of fireflies, particles and chromosomes

In terms of FA encoding, by encoding each firefly as a three-part binary string like Figure 5, if the  $k^{th}$  cell is one/zero

in the first part of the encoding ( $N$  cells), it means that  $k^{th}$  feature has been/has not been selected as one of the optimal features. After generating random strings based on Figure 5, as the initial population of the fireflies ( $X_i = 1.2. \dots N_{pop}$ ), the light intensity of each firefly ( $I_i$ ) is calculated (Yang, 2009). Following this initialization, each firefly was compared with another firefly and was moved toward brighter ones based on Eq. (1). Then, the brightness of the fireflies in their new positions were calculated based on the objective function. Finally, the algorithm was terminated based on the stop criterion. The final results of this process were the optimal features and SVM parameters for the change detection mask.

#### 2.2.4. Post-processing

One of disturbing factors in the analysis of remote sensing data is the presence of noise, which causes undesirable effects on the results. In this study, the connected component analysis was used to reduce the noise effect. The proposed algorithm in Figure 4 solves the shadow problem using MSI, which calculates differential morphological profiles (DMP) of the black top-hat (B-TH) for different directions ( $d$ ) and scales ( $s$ ) (Huang & Zhang, 2012). MSI is defined by Eq. (11), in which  $D$  and  $S$  are the number of directions and scales, respectively. Also, hole-filling of the image objects was another post-processing operation that was implemented. It should be noted that the shadows were first removed and then, the connected component analyses and hole filling processes were run on the change mask.

$$MSI = \frac{\sum_{d,s} DMP_{B-TH}(d,s)}{D \times S} \quad (11)$$

### 3. Experiments and analysis

The training and testing datasets are uniformly distributed over the images (Table 1).

Table 1. The training and testing datasets' information

	Change (pixels)		No_Change (pixels)	
	Azadshahr	Shiraz	Azadshahr	Shiraz
Training	138	138	148	128
Testing	145	124	145	110

Five common criteria, i.e. recall (sensitivity), precision, f-score, kappa and overall accuracy, were used to evaluate the proposed method (Sokolova et al., 2006). Kappa and overall accuracy are the overall criteria without focusing on individual classes, but recall, precision, and f-score can distinguish the correct classification of labels within different classes. The latter ones are the function of true positives ( $TP$ ) that indicate correctly classified “change”, true negatives ( $TN$ ) that indicate correctly classified “no-change”, false positives ( $FP$ ) that indicate wrongly classified “change”, and

false negatives ( $FN$ ) that indicate wrongly classified “no-change”. A higher recall value indicates a higher ability to correctly identify the relevant class. In contrary, a higher precision value indicates fewer mistakes in identifying the relevant class. The five mentioned criteria are defined as follows:

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$F - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (14)$$

$$OA = \frac{TP + TN}{CP + CN} \quad (15)$$

in which ,  $CN = FP + TN, CP = TP + FN$

$$\hat{k} = \frac{(TP + TN)(CP + CN) - (CP \times RP + CN \times RN)}{(CP + CN)^2 - (CP \times RP + CN \times RN)} \quad (16)$$

in which  $RP = TP + FP, RN = FN + TN$

The proposed algorithm for change detection implemented the GA, PSO, and FA optimizations with different parameter settings for 39, 27 and 48 times, respectively. A summary of the best results achieved in this process is listed in Table 2 and Table 3. The initial and final change masks obtained by the FA are shown in Figure 6. The proposed change detection algorithm was implemented in MATLAB R2015b (MathWorks Inc.) environment on a Windows® operating system with Intel® Core™ i5 CPU and 6 GB of RAM.

Based on the obtained results summarized in Table 3, FA had a better performance than PSO and GA in achieving higher overall accuracies and reasonable costs for creating the change mask. The overall accuracy and kappa coefficient of FA were 95.17% and 0.90, while they were 93.45% and 0.87 for PSO, and 91.03% and 0.82 for GA in the Azadshahr study area. Similarly, the overall accuracy and kappa coefficient of FA were 94.87% and 0.90, while they were 94.44% and 0.89 for PSO, and 93.16% and 0.86 for GA in the Shiraz study area. These results represent a higher performance of FA compared to that of others, while the number of iterations used as the stop criterion and the initial population were the same for all three optimization algorithms. However, the number of function evaluations of FA was more than the  $NFE$  of PSO and GA. This resulted in a dramatic increase of the FA run time (7-8 hours) in comparison to those of GA (65-110 minutes) and PSO (50-70 minutes) for 400 iterations.

Table 2. Summary of the FA-SVM results for Tehran study area

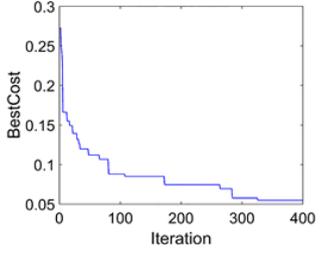
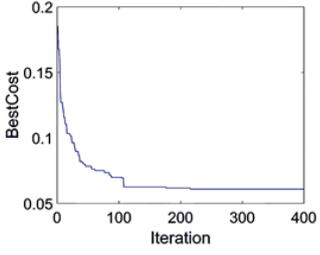
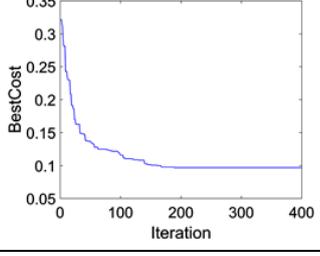
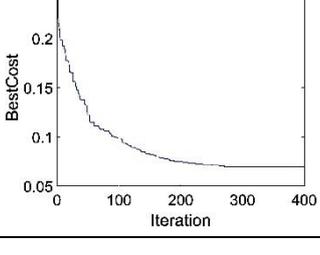
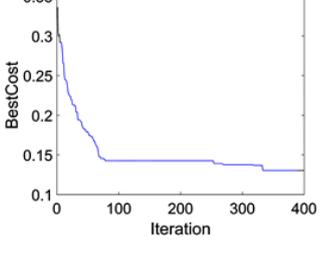
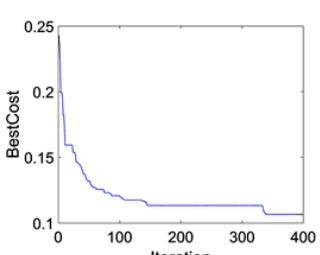
Details	OA (%)	$m$	$\alpha$	$\beta$	$\gamma$	$n_{pop}$	time (H)	It
<b>Sigmoid function and Euclidian distance</b>	77.59	2	0.2	1	0.7	20	31.70	300
	73.10	2	0.5	0.2	1	20	30.63	300
	77.58	2	0.2	1	0.5	20	32.25	300
	72.41	2	0.02	1	1	20	31.04	300
	<b>78.27</b>	<b>2</b>	<b>0.4</b>	<b>1</b>	<b>0.7</b>	<b>20</b>	<b>27.09</b>	<b>300</b>
	66.04	2	0.02	1	1	20	30.10	300
	75.86	2	0.5	1	0.7	20	15.28	200
	75.52	2	0.4	1	0.7	10	5.40	200
	74.13	2	0.4	0.5	0.7	10	6.18	200
	75.52	2	0.7	1	0.7	10	5.46	200
<b>Tanh function and Euclidian distance</b>	89.65	2	0.4	1	0.7	10	1.37	200
	86.89	2	0.7	1	0.7	10	2.03	200
	89.65	2	0.4	1	0.7	10	2.29	300
	89.31	1	0.4	1	1	10	1.76	300
	90	1.5	0.4	1	1	10	1.63	300
	92.35	2	0.4	0.7	1	10	2.09	300
	90	2	0.4	0.5	1	10	2.20	300
	<b>93.44</b>	<b>2</b>	<b>0.4</b>	<b>0.7</b>	<b>1</b>	<b>20</b>	<b>5.91</b>	<b>300</b>
	90.34	2	0.4	1	1	20	6.92	300
	92.07	2	0.4	1	0.7	15	3.87	300
93.10	2	0.2	0.7	1	20	4.52	300	
<b>Tanh function and Hamming distance</b>	90.34	2	0.4	1	0.7	10	1.99	300
	91.38	2	0.4	1	1	20	5.01	300
	94.14	2	0.2	1	1	15	2.66	300
	91.38	2	0.4	0.7	1	20	5.51	300
	93.79	2	0.2	0.7	1	20	4.67	300
	92.76	2	0.2	1	1	20	5.25	300
	93.79	1.5	0.2	1	1	15	2.75	300
	94.48	2	0.2	0.7	1	15	2.77	300
<b>95.17</b>	<b>2</b>	<b>0.2</b>	<b>0.7</b>	<b>1</b>	<b>20</b>	<b>7.06</b>	<b>400</b>	

Based on what already stated, the FA run time improved satisfactorily in Tanh\_Euclidian (4.5-7 hours) mode and Tanh\_Hamming (4.5-6 hours) mode in comparison to Sigmoid\_Euclidian (27-32 hours) mode in the Azadshahr study area for 300 iterations and a population size of 20. Comparison of PSO with GA shows that PSO outperforms GA as GA cannot find a better optimization solution than PSO not only with the same number of function evaluations, but even with a higher number of function evaluations.

In examining various scenarios of FA for feature selection and SVM parameter settings simultaneously, the use of the tangent hyperbolic function with the hamming distance improved the accuracy and the speed of the algorithm compared to the other two combinations, i.e. the sigmoid function with a Euclidean distance and the tangent hyperbolic function with a Euclidean distance. In terms of convergence stability, the convergence plots for the three algorithms for 400 iterations are shown in Figure 7. It is shown that both PSO parameters, i.e.  $p_{best}$  and  $g_{best}$ , fully

converge by iteration 200 and there is no turbulence after that point up to iteration 400 in the Azadshahr study area. However, the best and mean fitness of GA and FA algorithms do not fully converge until iteration 350 and stay there up to iteration 400. The results of the Shiraz study area show that the full convergence of FA and PSO was achieved by iteration 125 and 275, respectively, with no turbulence after those points up to iteration 400. However, the best and the mean fitness of GA do not fully converge until iteration 350 and stay there up to iteration 400. In the GA plot, the algorithm was trapped in a local minimum between iteration 86 and 253 in the Azadshahr study area and between iteration 150 and 355 in the Shiraz study area. This is considered a weakness for an optimization algorithm. The results showed a better performance for FA and PSO in comparison with GA. Table 4 reports the number of produced and selected features for each feature type.

Table 3. Best results of FA-SVM, PSO-SVM and GA-SVM

	Details	Recall (%)		F-score (%)	Precision (%)		OA (%)	NFE	$n_f$	kappa	Convergence plot of iteration-cost
		No.ch	ch	ch	No.ch	ch					
FA (hamming+tanh)	It = 400 $\gamma = 1$ $\beta_0 = 0.7$ $\alpha = 0.2$ Npop = 20 Copt = 68 Gopt = 0.32 Time = 7.06H <b>(Tehran)</b>	97.81	92.81	95.30	97.81	92.81	95.17	141747	17	0.90	
	It = 400 $\gamma = 1$ $\beta_0 = 0.7$ $\alpha = 0.2$ Npop = 20 Copt = 3 Gopt = 0.32 Time = 8.13H <b>(Shiraz)</b>	93.64	95.97	95.20	95.37	94.44	94.87	140165	12	0.90	
PSO	It = 400 C1 = 1.8 C2 = 2.2 Npop = 20 Copt = 70 Gopt = 0.106 Time = 65min <b>(Tehran)</b>	90.34	96.55	93.65	96.32	90.91	93.45	8000	51	0.87	
	It = 400 C1 = 1.8 C2 = 2.2 Npop = 20 Copt = 10 Gopt = 0.08 Time = 96min <b>(Shiraz)</b>	92.73	95.97	94.82	95.33	93.70	94.44	8000	32	0.89	
GA	It = 400 Pc = 0.8 Pm = 0.5 Mu = 0.04 Npop = 20 Copt = 99 Gopt = 0.88 Time = 102min <b>(Tehran)</b>	84.14	97.93	91.61	97.60	86.06	91.03	10400	67	0.82	
	It = 400 Pc = 0.8 Pm = 0.5 Mu = 0.04 Npop = 20 Copt = 113 Gopt = 0.14 Time = 120min <b>(Shiraz)</b>	90.91	95.16	93.65	94.34	92.19	93.16	10400	61	0.86	

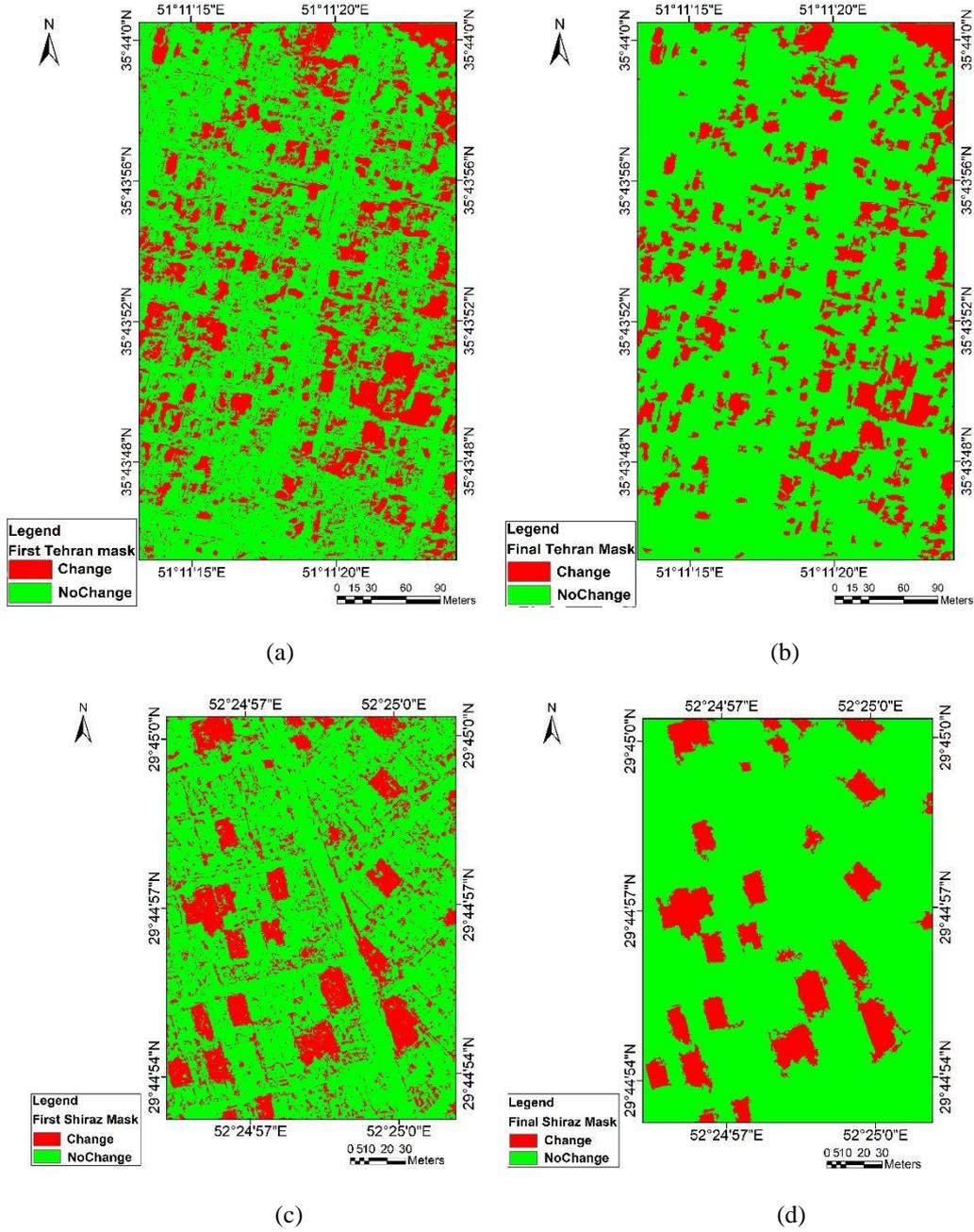


Figure 6. (a) Initial mask of Tehran, (b) final mask of Tehran, (c) initial mask of Shiraz, (d) final mask of Shiraz

As can be seen, FA selected fewer features than the other two optimization algorithms. There were four common features between FA and PSO, six common features between FA and GA, and three common features (two from color space and one from GLCM) between all the algorithms in the Azadshahr study area. In the Shiraz study area, there were eight common features between FA and PSO, ten common features between FA and GA, and five common features (one from color space, two from GLCM, one from Gabor and one from Fourier) between all the algorithms. In terms of the number of parameters a user should set for an algorithm to start, PSO had a lower number of main parameters ( $c_1, c_2$ ) in comparison to FA ( $\gamma, \beta_0$ ) and GA ( $P_c, P_m, M_u$ ). In general, when the combined factor of the run speed and accuracy was

the selection criterion, PSO was the best algorithm, but when the accuracy alone was the selection criterion, FA was the best algorithm for solving the hybrid problem of binary (feature selection) and continuous (SVM parameters) selections.

### 3.1. Feature contribution to change detection

To calculate the contribution of each feature to the change mask based on the overall accuracy, three indices are used: Effectiveness ( $E_{f_i}$ ), Partial Effectiveness ( $PE_{f_i}$ ) and Overall Effectiveness ( $OE_{f_i}$ ). These indices are defined as follows:

$$PE_{f_i} = E_{f_i} / \text{nof}_i \quad (18)$$

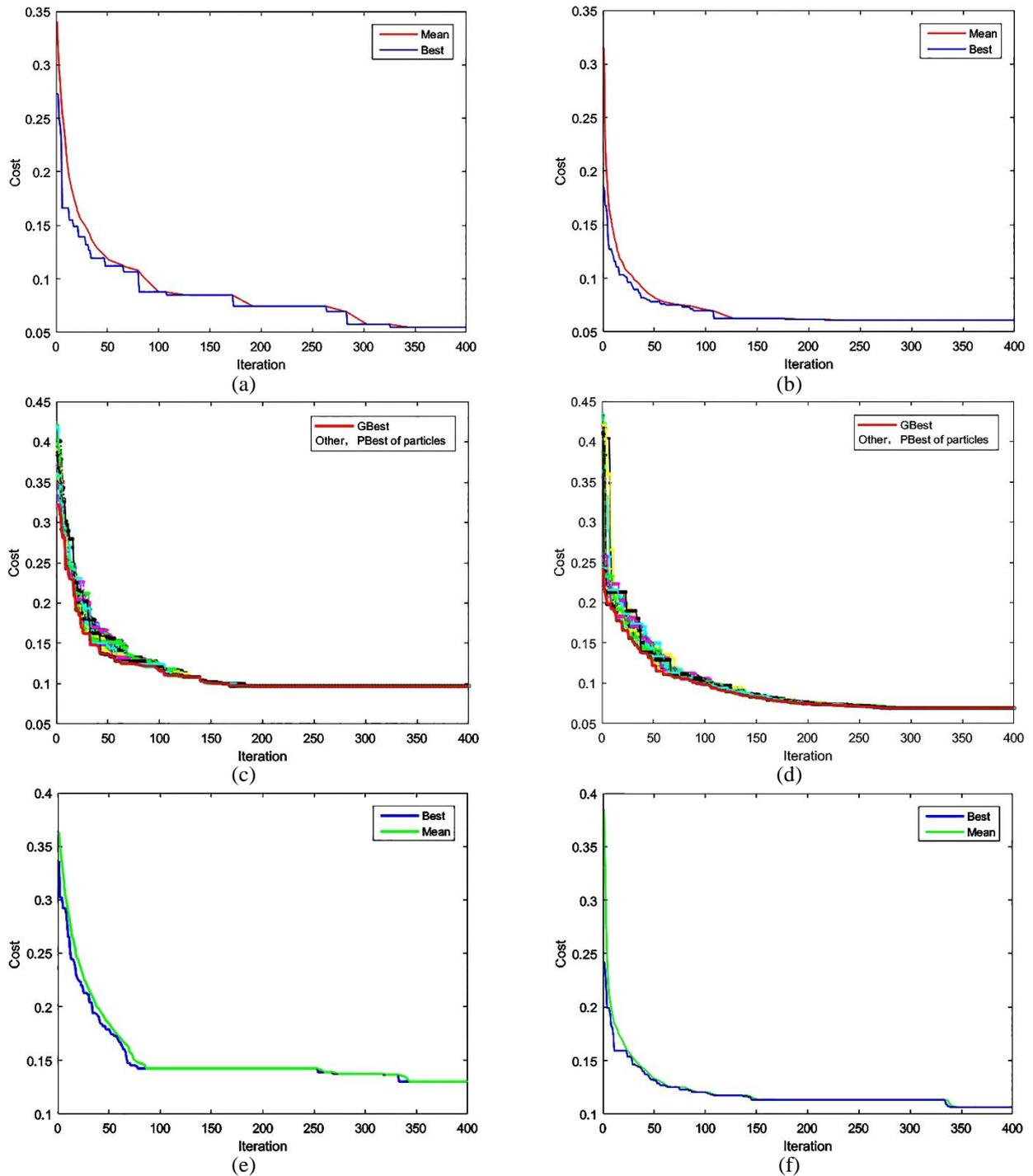


Figure 6. Best and Mean fitness plot of FA for (a) Tehran, and (b) Shiraz, Pbest, and Gbest plot of PSO for (c) Tehran, and (d) Shiraz, Best and Mean fitness plot of GA for (e) Tehran and (f) Shiraz

$$OE_{f_i} = \frac{nof_i}{N_i} \times E_{f_i} \quad (19)$$

$$E_{f_i} = OOA - ROOA \quad (17)$$

in which  $f_i$  is the feature type  $i$ ,  $N_i$  is the number of original features of type  $i$ ,  $nof_i$  is the number of selected features of type  $i$ ,  $OOA$  (Optimum Overall Accuracy) is the overall accuracy using the optimum feature set, and  $ROOA$  (Removed Optimum Overall Accuracy) is the overall

accuracy using the optimum feature set without the feature  $f_i$ . The effectiveness of each feature type was calculated by removing that feature type from the optimum feature set in the best run of the algorithm. The overall process of the effectiveness calculation for feature type  $i$  is as follows:

1. Remove the features of type  $i$ , i.e.  $f_i$ , from the optimum feature set achieved by the optimization

algorithm.

2. Implement the two-class classification of SVM and calculate the overall accuracy (*ROOA*).

Effectiveness is defined as the difference between the

original overall accuracy using the optimum feature set (*OOA*) and the *ROOA* from step 2.

Table 4. The effectiveness of each PSO and FA's feature type

Feature	$E_{f_i}$ (%)				$PE_{f_i}$ (%)				$OE_{f_i}$ (%)			
	Tehran		Shiraz		Tehran		Shiraz		Tehran		Shiraz	
	PSO	FA	PSO	FA	PSO	FA	PSO	FA	PSO	FA	PSO	FA
Anomaly	1.38	2.07	0	0	1.38	2.07	-	-	0.34	0.52	0	0
Curvelet	4.14	5.17	2.28	2.28	0.83	2.58	0.57	2.28	0.83	0.41	0.38	0.10
Edge	7.24	6.55	3.42	1.71	1.44	3.27	1.14	1.71	2.01	0.73	0.57	0.10
Fourier	1.73	6.89	5.98	6.84	1.73	3.44	2.99	6.84	0.29	2.30	1.99	1.14
Gabor	2.77	3.10	4.70	6.41	0.31	1.03	0.94	1.60	0.78	0.29	0.73	0.80
MBI	6.55	0	0	0	3.27	-	-	-	3.27	0	0	0
Other color spaces	14.14	11.38	4.27	4.27	2.83	5.69	2.14	4.27	7.85	2.53	0.95	0.47
spectral	1.38	0	1.71	0	1.38	-	1.71	-	0.46	0	0.57	0
GLCM Texture	8.97	22.76	5.556	6.84	0.56	5.69	0.51	2.28	2.61	1.65	1.02	0.34
Wavelet	10.35	1.38	2.99	7.26	1.72	1.38	0.75	7.26	2.07	0.05	0.40	0.24

In step 2 above, the SVM classification parameters ( $C$  and  $\gamma$ ) are not necessarily the same parameters achieved from the optimization algorithms. These parameters are set by trial and error, in which the user starts with setting the initial parameters and repeats the classification with different parameters several times until there is no improvement in the accuracy. Partial Effectiveness normalizes the effectiveness of each feature type based on the number of selected features for that feature type. However, the overall effectiveness applies a weight to the effectiveness by multiplying the ratio of the number of selected features to all the produced features for each feature type. The results of feature contribution analysis for PSO and FA (hamming+tanh) are shown in Table 5.

The results of PSO algorithm in Table 5 indicate the high effectiveness of other color spaces in the Azadshahr study area. The wavelet transform and the GLCM texture also show a good effectiveness. On the other hand, features extracted from anomaly, Fourier transform, and spectral information have the least effectiveness. The 1.38% effectiveness of the spectral information confirms that it should not be directly used for change detection in urban areas using high-resolution images; however, a combination of spectral data with other features and the spatial information is necessary to improve the accuracy of change detection. The results of FA in Table 5 indicate the high effectiveness of GLCM textures and other color space features in the Azadshahr study area.

The inefficiency of applying the spectral information alone

in high-resolution images in FA is also evident in Table 5. Also, the extracted features by the curvelet transform have effectivenesses of about 4% and 5% in PSO and FA, respectively. This shows the efficiency of this new type of transformation. Each feature selected from other color spaces and textures (GLCM) has a higher efficiency by considering 5.69% partial effectiveness in FA. These two types of features have the maximum value of the overall effectiveness in FA that reflect the ratio of the number of selected features to all produced features for each feature type. The results of the Shiraz study area, which has a relatively better radiometric condition than Azadshahr study area, indicate the proper effectiveness of the Fourier transform, Gabor, other color space and GLCM texture for the PSO and a good effectiveness of the Fourier transform, Gabor, other color space, GLCM texture and Wavelet transform for FA. Also in both algorithms, the effectiveness of the spectral information is low.

Repeatability of an optimization algorithm is an important factor showing how the results can be repeated when the algorithm is run for several times with the same initial parameters. The variability in the result can be quantified using the ratio of the standard deviation to the mean (coefficient of variation or CV); the lower the CV, the higher the repeatability.

This ratio was 0.024, 0.016 and 0.017 for GA, PSO, and FA (hamming+tanh), respectively. Based on the overall accuracy of 11 run times in the Azadshahr study area, GA provided the least repeatable results. Also, the results of the

Shiraz study area showed the better CV in FA than GA and PSO. The results are summarized in Table 6.

**3.2. Comparing the Proposed Methodology with Post-Classification and Thresholding Results**

Post-classification and thresholding methods are the two conventional and widely used techniques in change detection. In the post-classification approach, the classification is separately done on each image and then,

Table 5. The effectiveness of each PSO and FA’s feature type

Feature	$E_{f_i} (%)$				$PE_{f_i} (%)$				$OE_{f_i} (%)$			
	Tehran		Shiraz		Tehran		Shiraz		Tehran		Shiraz	
	PSO	FA	PSO	FA	PSO	FA	PSO	FA	PSO	FA	PSO	FA
Anomaly	1.38	2.07	0	0	1.38	2.07	-	-	0.34	0.52	0	0
Curvelet	4.14	5.17	2.28	2.28	0.83	2.58	0.57	2.28	0.83	0.41	0.38	0.10
Edge	7.24	6.55	3.42	1.71	1.44	3.27	1.14	1.71	2.01	0.73	0.57	0.10
Fourier	1.73	6.89	5.98	6.84	1.73	3.44	2.99	6.84	0.29	2.30	1.99	1.14
Gabor	2.77	3.10	4.70	6.41	0.31	1.03	0.94	1.60	0.78	0.29	0.73	0.80
MBI	6.55	0	0	0	3.27	-	-	-	3.27	0	0	0
Other color spaces	14.14	11.38	4.27	4.27	2.83	5.69	2.14	4.27	7.85	2.53	0.95	0.47
spectral	1.38	0	1.71	0	1.38	-	1.71	-	0.46	0	0.57	0
GLCM Texture	8.97	22.76	5.556	6.84	0.56	5.69	0.51	2.28	2.61	1.65	1.02	0.34
Wavelet	10.35	1.38	2.99	7.26	1.72	1.38	0.75	7.26	2.07	0.05	0.40	0.24

Table 6. Coefficient of variation of different parameters of the GA, PSO and FA

Algorithm	Kappa		OA		Precision (change)		F-score (change)		Recall (change)	
	Tehran	Shiraz	Tehran	Shiraz	Tehran	Shiraz	Tehran	Shiraz	Tehran	Shiraz
GA	0.0572	0.0231	0.0245	0.106	0.0380	0.0106	0.0209	0.0102	0.0204	0.0178
PSO	0.0356	0.0341	0.0160	0.157	0.0309	0.142	0.0139	0.0148	0.0195	0.0178
FA	0.0388	0.0224	0.0179	0.0073	0.0201	0.0071	0.0173	0.0069	0.0283	0.0100

change detection is carried out by comparing the corresponding pixels in the classified images. As mentioned, the post classification approach lacks the ability of detecting the intra-class changes. For instance, if a building changes to another building and there is no sub-class for buildings, there will be three scenarios in change detection using the post-classification approach:

- Classification at each time point has produced correct classes: no-change result is obtained through the comparison of the two classification results (while there is a change).
- Classification at one time point has produced correct classes, but wrong classes at other time point: change result is obtained through the comparison of the two classification results; however, the “from-to” deduction is wrong.
- Classification at each time point has produced wrong classes: this is a complicated case that might result in correct or wrong change detection results depending on the classification results.

In the thresholding method, the change detection is

performed through taking into account the difference between the features. In another study, these two approaches were evaluated over the same study area, i.e. Azadshahr, using the same images. Therefore, we were able to compare the results of the proposed methodology in this manuscript with them. Moghimi (Moghimi, 2015a) used the spatial information, such as GLCM and Gabor filter, to extract the changes in the Azadshahr study area using these two approaches. In the first approach, the author produced an index fused with a discrete wavelet transform after the feature extraction and dimensionality reduction by principal component analysis and kernel principal component analyses. The final change mask was generated using the Markov chains. The author did not consider the shadow. In the second approach, the post-classification was used following the implementation of SVM and GA algorithms for feature selection to classify each image. The first approach achieved an overall accuracy of 78.23% and a kappa of 0.56 and the second approach achieved an overall accuracy of 84.10% and a kappa of 0.68 in change detection; these numbers are much lower than the results obtained in

this study using FA. Figure 8 shows the result of Moghimi (Moghimi, 2015a) using Markov chains and post-classification techniques. The poor performance of post-classification in detection of intra-class changes can be well seen. Saeidzadeh (Saeidzadeh et al., 2016) used a combination of fuzzy thresholding and Otsu thresholding to

produce the change mask of the Azadshahr case study. She showed that a combination of fuzzy and Otsu thresholding approaches has a better performance than fuzzy and Otsu, separately. Their approach achieved an overall accuracy of 71.47% and a kappa of 0.60, indicating a poor result of thresholding method in change detection of complicated

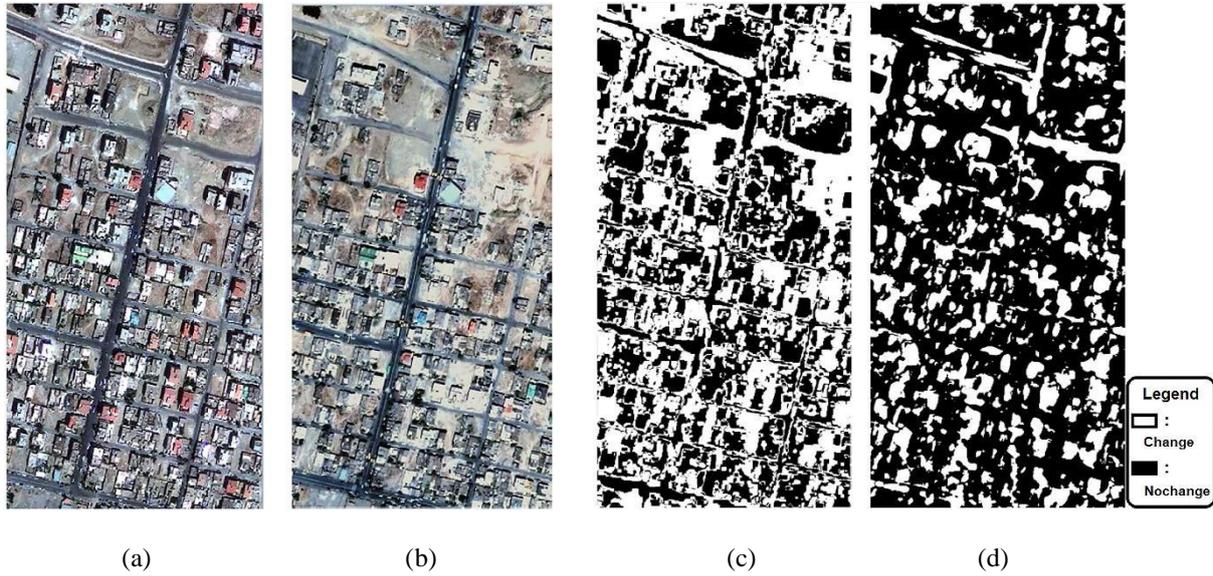


Figure 8. Results of Moghimi (2015b) study. (a), (b) case study images, (c) change mask obtained by post-classification, (d) change mask obtained by Markov chains

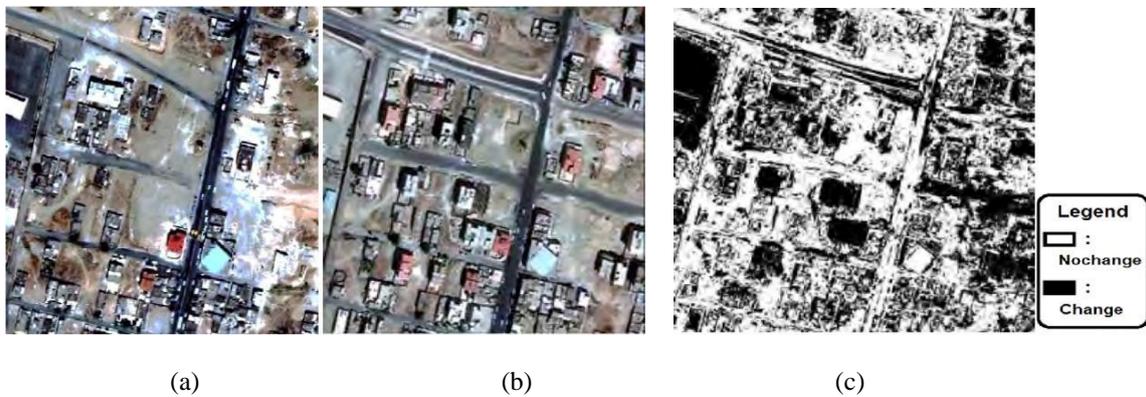


Figure 9. Results of the Saeidzadeh's (Saeidzadeh et al., 2016) study. (a),(b) case study images, (c) change mask of fuzzy-outso

Table 7. Accuracy of change detection results using different techniques

		<i>OA (%)</i>	<i>Kappa</i>
Moghimi (2015b)	Markov chain	78.23	0.56
	Post-classification	84.10	0.68
(Sahebi et al. (2016))	fuzzy	59.16	52.05
	outso	66.53	56.63
	fuzzy-outso	71.47	0.61
change vector analysis (CVA)		68.97	0.38
Fuzzy C-means (FCM)		71.03	0.42
Current study		95.17	0.90

urban areas affected by poor radiometric condition of high-resolution images. Figure 9 shows the results of Saeidzadeh et al., 2016. Note that their study area was a subset of the current Azadshahr study area.

Table 7 summarizes the accuracy of some techniques in creating the change detection mask over the Azadshahr study area. The superiority of the proposed method can be well seen in this table.

#### 4. Conclusions

In the current study, a methodology was developed for change detection in urban areas using high-resolution imagery with successful results. This was achieved using the firefly optimization algorithm for feature selection in the heart of the proposed methodology. Despite the high capacity of the spectral information in high-resolution satellite images, the results of this study highlighted that their application for change detection leads to low-accuracy results. Instead, features such as GLCM texture features and other color spaces along with the features extracted from frequency domain to be the most effective features used in change detection of high-resolution images. This conclusion was achieved by introducing three indices for evaluating the contribution and importance of each feature type in change detection results. Thereby, it is recommended to use these indices as part of standard reporting procedure of future change detection and feature selection studies. Here, it was showed that FA outperforms other well-known approaches. With its great potential, it is recommended to further adapt and exploit this algorithm in other remote sensing applications.

#### Appendix A:

##### Feature Extraction

The following is a summary of the features extracted in the approach proposed in the current study.

##### Anomaly detection

The main goal of anomaly detection is to find the pixels whose spectral properties significantly differ from the background spectral property. A standard approach for anomaly detection is the RX algorithm with two variations of the local and global version. Here, we used both local and global versions of this algorithm as a feature.

##### Edge detection

Edge is an important information in image processing. Canny, Sobel, Prewitt, log, Roberts and Zero crossing are used besides Von Neumann and Moore edge detection by the OR and AND operators.

##### Morphological building index (MBI)

This index was used as one of the features in this study. Buildings are one of the most important elements of urban areas. MBI is a novel method that can extract buildings on high-resolution images. The main idea behind MBI is to build a relationship between implicit characteristics of

buildings, such as brightness, size, contrast, directionality and shape, and the properties of morphological transformation, such as reconstruction, granulometry, and directionality.

##### Other color spaces

One of the techniques used in image processing is the conversion of color spaces. HSL (Hue-Saturation-Luminance/Lightness) is a color space that reduces inter-band correlation. A recent study has taken the advantage of luminance and saturation in this color space for change detection. In addition to HSL, we used two other color spaces of YIQ and YCBR.

##### Texture extracted in spatial domain

The structural arrangement of pixels and their relationship with neighbors is provided by texture analysis. GLCM is a famous and conventional method of extracting second order statistical texture features that have introduced 14 statistical features as texture features. In this study, occurrence texture (mean, variance, entropy), co-occurrence texture (mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, correlation) on the grayscale image by different window sizes (3, 5, 9, 11, and 13) were used as the texture features.

##### Texture extracted from wavelet transformation

By applying the wavelet transform to the image, four sub-images are created. The approximation sub-image contains low-frequency information and the other three sub-images contain high-frequency information at horizontal, vertical and diagonal directions. In this study, the Haar wavelet mother transform method and analysis at levels 1, 2 and 5 were used for extraction of log energy, Shannon's index, entropy and angular second moment features. Also the sub-images that include high-frequency information were used as wavelet features.

##### Gabor filter

Gabor filter is a complex sinusoid harmonic function with two elements: 1-frequency ( $f$ ), 2- orientation ( $\theta$ ). The Gabor filter is formed as the product between an elliptical Gaussian function and a sinusoidal function. This filter consists of real and imaginary parts, in which the real part (cosine function) acts as a low-pass filter, and the imaginary part (sine function) acts as a high-pass filter. In this study, log energy, Shannon's index, entropy and angular second moment of the real part were used by  $f = 0.5\text{Hz}$ , and  $0.8\text{Hz}$  and  $\theta = 0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$  as features extracted from the Gabor filter.

##### Texture extracted from Fourier transformation

Certain frequencies of an image can be analyzed and processed using Fourier transform. For feature extraction, we used Fast Fourier transform (FFT) to convert the image from the spatial domain to the frequency domain, then applying a frequency filter such as circular pass (radius 20), circular cut (radius 3, 5, 10, 15), and band pass (inner radius 7, outer radius 78) on the transformed data and finally transforming

the filtered data back to the spatial domain using inverse FFT.

#### Texture extracted from curvelet transformation

Curvelet transform is a new algorithm in multi-scale transforms developed to improve the wavelet transform. Applying this transform in different angles and different scales is one of the strengths of a curvelet transform. For feature extraction, fast discrete Curvelet transforms (FDCT) is used to convert the image from spatial domain to the Curvelet space. Then, some curvelet coefficients are removed and the inverse Curvelet transform is applied to bring back the data to the spatial domain. Finally, log energy, Shannon's index, entropy and angular second moment are extracted as desired features. It can produce a variety of different features with different combinations of the coarse level (low-frequency), detailed level (mid-frequency), and fine level (high-frequency) coefficients.

#### References

- Bhatt, A., Ghosh, S., & Kumar, A. (2016) 'Spectral Indices Based Change Detection in an Urban Area Using Landsat Data' *Proceedings of Fifth International Conference on Soft Computing for Problem Solving*. Springer, pp. 425-441.
- Campagnolo, M. L., & Cerdeira, J. O. (2006) 'Contextual classification of remotely sensed images with integer linear programming' *CompIMAGE*. pp. 123-128.
- Celik, T. (2009). Multiscale change detection in multitemporal satellite images. *IEEE Geoscience and Remote Sensing Letters*, 6(4), 820-824.
- Celik, T. (2010). Change detection in satellite images using a genetic algorithm approach. *IEEE Geoscience and Remote Sensing Letters*, 7(2), 386-390.
- Celik, T., & Ma, K.-K. (2011). Multitemporal image change detection using undecimated discrete wavelet transform and active contours. *IEEE Transactions on Geoscience and Remote Sensing*, 49(2), 706-716.
- Chandrasekaran, K., Simon, S. P., & Padhy, N. P. (2013). Binary real coded firefly algorithm for solving unit commitment problem. *Information Sciences*, 249, 67-84.
- Chen, Q., & Chen, Y. (2016). Multi-Feature Object-Based Change Detection Using Self-Adaptive Weight Change Vector Analysis. *Remote Sensing*, 8(7), 549.
- Chen, Q., Chen, Y., & Jiang, W. (2016). Genetic particle swarm optimization-based feature selection for very-high-resolution remotely sensed imagery object change detection. *Sensors*, 16(8), 1204.
- Crawford, B., Soto, R., Olivares-Suárez, M., & Paredes, F. (2014). A binary firefly algorithm for the set covering problem. *Modern Trends and Techniques in Computer Science* (pp. 65-73). Springer.
- Dalla Mura, M., Benediktsson, J. A., Bovolo, F., & Bruzzone, L. (2008). An unsupervised technique based on morphological filters for change detection in very high resolution images. *IEEE Geoscience and Remote Sensing Letters*, 5(3), 433-437.
- de Jong, S. M., Hornstra, T., & Maas, H.-G. (2001). An integrated spatial and spectral approach to the classification of Mediterranean land cover types: the SSC method. *International Journal of Applied Earth Observation and Geoinformation*, 3(2), 176-183.
- Devi, R. N., & Jij, G. (2015). Change Detection Techniques- A Survey. *International Journal on Computational Sciences & Applications (IJCSA)*, 5(2).
- Dey, V., Zhang, Y., & Zhong, M. (2010). A review on image segmentation techniques with remote sensing perspective. na.
- Du, P., Liu, S., Gamba, P., Tan, K., & Xia, J. (2012). Fusion of difference images for change detection over urban areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(4), 1076-1086. doi:10.1109/JSTARS.2012.2200879.
- El-Hattab, M. M. (2016). Applying post classification change detection technique to monitor an Egyptian coastal zone (Abu Qir Bay). *The Egyptian Journal of Remote Sensing and Space Science*, 19(1), 23-36.
- Erener, A., & Düzgün, H. S. (2009). A methodology for land use change detection of high resolution pan images based on texture analysis. *Italian Journal of Remote Sensing*, 41(2), 47-59.
- Fu, K.-S., & Mui, J. (1981). A survey on image segmentation. *Pattern recognition*, 13(1), 3-16.
- Galdavi, S., Mohammadzadeh, M., Salmanmahiny, A., & Nejad, A. N. (2013). Urban Change Detection Using Multi-temporal Remotely Sensed Imagery (Case Study: Gorgan Area, Northern Iran). *Environment and Urbanization Asia*, 4(2), 339-348.
- Gao, Y., & Mas, J. F. (2008). A comparison of the performance of pixel-based and object-based classifications over images with various spatial resolutions. *Online journal of earth sciences*, 2(1), 27-35.
- Hall, O., & Hay, G. J. (2003). A multiscale object-specific approach to digital change detection. *International Journal of Applied Earth Observation and Geoinformation*, 4(4), 311-327.
- He, C., Wei, A., Shi, P., Zhang, Q., & Zhao, Y. (2011). Detecting land-use/land-cover change in rural-urban fringe areas using extended change-vector analysis. *International Journal of Applied Earth Observation and Geoinformation*, 13(4), 572-585.
- Huang, X., & Zhang, L. (2012). Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(1), 161-172.
- Huang, X., Zhang, L., & Zhu, T. (2014). Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1), 105-115.
- Hussain, M., Chen, D., Cheng, A., Wei, H., & Stanley, D. (2013). Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80, 91-106.
- Karantzalos, K. (2015). Recent advances on 2D and 3D change detection in urban environments from remote sensing data. *Computational Approaches for Urban Environments* (pp. 237-272). Springer.

- Kusetogullari, H., Yavariabdi, A., & Celik, T. (2015). Unsupervised change detection in multitemporal multispectral satellite images using parallel particle swarm optimization. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(5), 2151-2164.
- Li, W., Lu, M., & Chen, X. (2015) 'Automatic change detection of urban land-cover based on SVM classification' *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*. IEEE, pp. 1686-1689.
- Lin, S.-W., Ying, K.-C., Chen, S.-C., & Lee, Z.-J. (2008). Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert systems with applications*, 35(4), 1817-1824.
- Liu, D., & Xia, F. (2010). Assessing object-based classification: advantages and limitations. *Remote Sensing Letters*, 1(4), 187-194.
- Liu, J., Du, M., & Mao, Z. (2017). Scale computation on high spatial resolution remotely sensed imagery multi-scale segmentation. *International Journal of Remote Sensing*, 38(18), 5186-5214.
- Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004). Change detection techniques. *International journal of remote sensing*, 25(12), 2365-2401.
- Łukasik, S., & Żak, S. (2009) 'Firefly algorithm for continuous constrained optimization tasks' *International Conference on Computational Collective Intelligence*. Springer, pp. 97-106.
- Mehrotra, A., Singh, K. K., & Khandelwal, P. (2014) 'An unsupervised change detection technique based on Ant colony Optimization' *Computing for Sustainable Global Development (INDIACom), 2014 International Conference on*. IEEE, pp. 408-411.
- Moghimi, A. (2015a). Integrating textural and spectral information from satellite images for change detection in urban areas using direct and post classification comparison methods. *M.Sc Thesis, K.N. Toosi University of Technology*.
- Moghimi, A. (2015b) Integrating textural and spectral information from satellite images for change detection in urban areas using direct and post classification comparison methods. *M.Sc Thesis, K.N. Toosi University of Technology*
- Mosammam, H. M., Nia, J. T., Khani, H., Teymouri, A., & Kazemi, M. (2016). Monitoring land use change and measuring urban sprawl based on its spatial forms: The case of Qom city. *The Egyptian Journal of Remote Sensing and Space Science*.
- Paul, A., Chowdary, V., Srivastava, Y., Dutta, D., & Sharma, J. (2016). Change detection of linear features in temporally spaced remotely sensed images using edge-based grid analysis. *Geocarto International*, 1-15.
- Phalke, S. (2006). Change detection of man-made objects using very high resolution images. *Library and Archives Canada= Bibliothèque et Archives Canada*.
- Raja, R. A., Anand, V., Kumar, A. S., Maithani, S., & Kumar, V. A. (2013). Wavelet based post classification change detection technique for urban growth monitoring. *Journal of the Indian Society of Remote Sensing*, 41(1), 35-43.
- Saeidzadeh, F., Sahebi, M., Ebadi, H., & Sadeghi, V. (2016). Change Detection of Multitemporal Sattelite Images by Comparison of Binary Mask and Most Classification Comparison Methods. *Journal of Geomatics Science and Technology*, 5(3), 111-128.
- Sahebi, M., Ebadi, H., & Sadeghi, V. (2016). Change Detection of Multitemporal Sattelite Images by Comparison of Binary Mask and Most Classification Comparison Methods. *Journal of Geomatics Science and Technology*, 5(3), 111-128.
- Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006) 'Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation' *Australasian Joint Conference on Artificial Intelligence*. Springer, pp. 1015-1021.
- Stow, D., Tinney, L., & Estes, J. (1980). Deriving land use/land cover change statistics from Landsat-A study of prime agricultural land.
- Sui, H., Zhou, Q., Gong, J., & Ma, G. 7 (2008) 'Processing of multi-temporal data and change detection' *Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences: 2008 ISPRS Congress Book*. CRC Press, p. 227.
- Tewkesbury, A. P., Comber, A. J., Tate, N. J., Lamb, A., & Fisher, P. F. (2015). A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, 160, 1-14.
- Van Oort, P. (2007). Interpreting the change detection error matrix. *Remote Sensing of Environment*, 108(1), 1-8.
- Volpi, M., Tuia, D., Bovolo, F., Kanevski, M., & Bruzzone, L. (2013). Supervised change detection in VHR images using contextual information and support vector machines. *International Journal of Applied Earth Observation and Geoinformation*, 20, 77-85.
- Walter, V. (2004). Object-based classification of remote sensing data for change detection. *ISPRS Journal of photogrammetry and remote sensing*, 58(3), 225-238.
- Wen, D., Huang, X., Zhang, L., & Benediktsson, J. A. (2016). A novel automatic change detection method for urban high-resolution remotely sensed imagery based on multiindex scene representation. *IEEE Transactions on Geoscience and Remote Sensing*, 54(1), 609-625.
- Yang, X.-S. (2009) 'Firefly algorithms for multimodal optimization' *International symposium on stochastic algorithms*. Springer, pp. 169-178.
- Yang, X.-S. (2010). Engineering optimization: an introduction with metaheuristic applications. John Wiley & Sons.
- Yu, X., & Gen, M. (2010). Introduction to evolutionary algorithms. Springer Science & Business Media.
- Yuan, F., Sawaya, K. E., Loeffelholz, B. C., & Bauer, M. E. (2005). Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. *Remote sensing of Environment*, 98(2), 317-328.
- Zhang, J., Gao, B., Chai, H., Ma, Z., & Yang, G. (2016). Identification of DNA-binding proteins using multi-features fusion and binary firefly optimization algorithm. *BMC bioinformatics*, 17(1), 323.
- Zhen, L., Wang, L., Wang, X., & Huang, Z. 2 (2008) 'A novel PSO-inspired probability-based binary

*Moradi et al, 2018*

optimization algorithm' *Information Science and Engineering, 2008. ISISE'08. International Symposium on*. IEEE, pp. 248-251.