Information Sciences xxx (2010) xxx–xxx

Fuzzy clustering algorithms for unsupervised change detection in remote sensing images

Ashish Ghosh [a,*], Niladri Shekhar Mishra [b], Susmita Ghosh [c]

[a] Machine Intelligence Unit and Center for Soft Computing Research, Indian Statistical Institute, 203 B.T. Road, Kolkata 700 108, India
[b] Department of Electronics and Communication Engineering, Netaji Subhash Engineering College, Kolkata 700 152, India
[c] Department of Computer Science and Engineering, Jadavpur University, Kolkata 700 032, India

ARTICLE INFO

Article history:
Received 5 March 2009
Received in revised form 8 October 2010
Accepted 13 October 2010
Available online xxxx

Keywords:
Remote sensing
Change detection
Multi-temporal images
Fuzzy clustering
Fuzzy *c*-means clustering
Gustafson–Kessel clustering
Genetic algorithms
Simulated annealing
Xie–Beni validity measure

ABSTRACT

In this paper, we propose a context-sensitive technique for unsupervised change detection in multitemporal remote sensing images. The technique is based on fuzzy clustering approach and takes care of spatial correlation between neighboring pixels of the difference image produced by comparing two images acquired on the same geographical area at different times. Since the ranges of pixel values of the difference image belonging to the two clusters (*changed* and *unchanged*) generally have overlap, fuzzy clustering techniques seem to be an appropriate and realistic choice to identify them (as we already know from pattern recognition literatures that fuzzy set can handle this type of situation very well). Two fuzzy clustering algorithms, namely fuzzy *c*-means (FCM) and Gustafson–Kessel clustering (GKC) algorithms have been used for this task in the proposed work. For clustering purpose various image features are extracted using the neighborhood information of pixels. Hybridization of FCM and GKC with two other optimization techniques, genetic algorithm (GA) and simulated annealing (SA), is made to further enhance the performance. To show the effectiveness of the proposed technique, experiments are conducted on two multispectral and multitemporal remote sensing images. A fuzzy cluster validity index (Xie–Beni) is used to quantitatively evaluate the performance. Results are compared with those of existing Markov random field (MRF) and neural network based algorithms and found to be superior. The proposed technique is less time consuming and unlike MRF does not require any a priori knowledge of distributions of changed and unchanged pixels.

1. Introduction

In remote sensing applications, *change detection* is the process aimed at identifying differences in the state of a land cover by analyzing a pair of images acquired on the same geographical area at different times [58,61]. Such a problem plays an important role in different domains like studies on land use/land cover dynamic [10], monitoring shifting cultivations [6], burned area identification [4], analysis of deforestation processes [24,30], assessment of vegetation changes [8], monitoring of urban growth [47] and oceanography [43]. Since all these applications usually require an analysis of large areas, development of completely automatic and unsupervised change detection techniques is of high relevance to reduce the time required by manual image analysis.

Change detection in remotely sensed data may be done either in supervised or in unsupervised manner [1,3,5,7,8,13,17,21,24,32,33,36,46,47]. In supervised techniques, a set of *training patterns* is required for learning the

classifier. In real-life, it is difficult to have data containing spectral signatures of changes from which *training patterns* can be generated. In unsupervised techniques, there is no need of *training data*. Thus the usefulness of unsupervised techniques is more than supervised ones for this problem. We may think of unsupervised change detection problem as a clustering one where the task is to discriminate the data into two groups *changed* and *unchanged*.

When an user has two multitemporal images in hand to detect *changes* between them, it may happen that the images are not consistent or comparable because of misregistration [11,65] (*i.e.* the two images may not be coregistered [7]) and radiometric and geometric errors [66]. Moreover, some noise may be present. Hence before analyzing the images for detecting changes, a certain degree of (pre) processing is needed [7]. Thus, in literature [58,61], three steps are suggested to be performed sequentially for unsupervised *change detection*. They are (i) pre-processing, (ii) image comparison and (iii) image analysis. In step (i), operations like coregistration, radiometric and geometric corrections and noise reduction are done to make the two multitemporal images compatible. To remove the effects of sensor errors and environmental factors, radiometric corrections are needed [7,66].

In step (ii), the two pre-processed images are taken as input and compared pixel by pixel and thereafter another image is generated, called the difference image (DI). To generate the DI, we may consider: (a) only one spectral band (*i.e.*, Univariate Image Differencing - UID) [61], (b) multiple spectral bands (*i.e.* Change Vector Analysis - CVA) [5,61], (c) vegetation indices (Vegetation Index Differencing - VID) [61,64] etc. Tasseled Cap Transformation [18] is also a popular method. The most popular of these is the CVA and used in our study. We have chosen CVA because by using this technique reflectance properties of various land cover types can be combined.

After performing the above two steps (pre-processing and image comparison and thereby generating the DI) *change detection* is done on the DI. Either context-insensitive or context-sensitive procedure is adopted [21] for this. Histogram thresholding [5,46] is of the first kind. The threshold value may be detected by manual trial-and-error (MTET) process or by automatic techniques by analyzing the statistical distribution of the DI. In these cases spatial correlation between the neighboring pixels is not taken into account. Most of the context-sensitive techniques [1,3,5,36] are based on MRF, require the selection or estimation of a model for the statistical distributions of *changed* and *unchanged* classes, and can overcome the drawbacks of context-insensitive approaches mentioned earlier. Algorithms (like Expectation–Maximization (EM) [14]) are required for estimating the class distributions assuming different standard distributions e.g., Gaussian [5], generalized Gaussian [1] and mixture of Gaussians [3]. A few context-sensitive techniques using neural networks are also suggested recently [21,22,24,54].

To the best of the authors' knowledge, the use of fuzzy clustering for change detection in remotely sensed images has not been reported in the literature. So, in order to overcome the limitations imposed by the need of selecting or estimating a statistical model for *changed* and *unchanged* class distributions, we propose unsupervised, distribution free and context-sensitive change detection techniques based on fuzzy clustering [2] approach. Our attempt here is to recover the *changed* and *unchanged* regions of the DI by constructing two clusters. Normally the pixels of the DI belonging to two clusters *changed* and *unchanged* are not separable by sharp boundaries (as they are highly overlapped). As fuzzy clustering technique is more appropriate and realistic to separate overlapping clusters [66], we have chosen fuzzy clustering techniques to have a better judgement of the two groups. From our results it is also noticed that fuzzy clustering is a better choice than crisp clustering (as the crisp version yields worse results). In this regard we have used two fuzzy clustering algorithms namely fuzzy c-means [2] and Gustafson Kessel [26]. There are several fuzzy cluster validity indices available in the literature to evaluate fuzzy clustering results. We have used one of the popular measures proposed by Xie and Beni [69] which considers both intra cluster compactness and inter cluster separation. We have used Mahalanobis norm while evaluating the outcome of GK type clustering as in [39].

As clustering processes are sensitive to their initializations and thus have a tendency to get stuck to local optima, we have combined them with two well known other optimization techniques namely genetic algorithms (GAs) [23] and simulated annealing (SA) [67] with a hope to have improved performance.

In this work we emphasis on the physical interpretation of the utility of these two fuzzy clustering algorithms for change detection without going into deep mathematical details. We have stated some elementary formulae in Section 2.2 to describe the fuzzy clusterings without deviating from this very goal. We also tried to analyze how and to what extent efficiencies of the used techniques could be enhanced by adjusting the parameters of the fuzzy algorithms or by hybridizing them with other optimization techniques. Our main concern here is to deal with an environment having overlapping irregular shaped and unstructured clusters.

To assess the proposed technique, experiments are carried out on two real world data sets and compared the results with those obtained by already published techniques [5,21] for solving the same problem of change detection on the same data sets. The proposed technique is found to be superior with respect to both error and time requirement.

This paper is organized as follows: Section 2 provides a brief description of a few crisp and fuzzy clustering algorithms. Section 3 is about genetic algorithms and simulated annealing as a search procedure. Section 4 is about validity measures for fuzzy clustering. Section 5 describes the proposed *change detection* technique. The data sets used in the experiments and the results obtained are described in Sections 6 and 7, respectively. Finally, in Section 8, conclusions are drawn.

2. Clustering

Cluster analysis [31,34] partitions a data set into a reasonable number of disjoint groups, where each group contains similar patterns. The partitions should be such that patterns are "homogeneous" within the clusters and "heterogeneous" be-

tween the clusters. Each datum in the data set is a point in the feature space having the dimensionality equal to the number of attributes of it. The clustering algorithms (both fuzzy and non-fuzzy) used in the present investigation are described here in brief.

2.1. Hard c-means (HCM) clustering

One of the simplest clustering techniques is HCM clustering [16,62]. In this method, from a set of patterns, c number of patterns are randomly chosen as initial cluster centers of the c clusters. In each iteration the patterns are assigned to the cluster having the nearest center; and the centers of the clusters are updated after assignment of all patterns. The center of a cluster is the arithmetic mean of the patterns assigned to it at the previous iteration. Thus if $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c]$, includes c number of vectors $\mathbf{v}_i$ ($\mathbf{v}_i$, $1 \leqslant i \leqslant c$) of cluster prototypes, then after the first iteration $\mathbf{v}_1$ becomes the arithmetic mean of the patterns assigned to the first cluster at this iteration, $\mathbf{v}_2$ the arithmetic mean of the patterns assigned to second cluster and so on. It means that the different components of the vector $\mathbf{v}_i$ are the arithmetic means of corresponding component values of the patterns belonging to the ith cluster. This process stops when the centers become fixed *i.e. no changes* occur from the *partitioning* point of view.

This process is "partitive" because at every iteration the partitions or boundaries between the clusters are changed and some patterns from one cluster are moved to some other clusters. The boundaries are crisp or hard as each pattern is assigned to only one cluster. The algorithm is put in a tabular form (in Table 1).

The HCM algorithm basically minimizes the following objective function

$$J(\mathbf{X};\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{n} D_{ik}, \tag{1}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ is the set of n patterns, $\mathbf{x}_k$ is the kth pattern $\in \mathbf{X}$ and $D_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|^2$ (Euclidean norm) is the dissimilarity measure between the sample x_k and the ith cluster center.

2.2. Fuzzy clustering

Fuzzy sets were introduced in 1965 by Lotfi Zadeh [70] as a new way to represent vagueness in everyday life. They are generalizations of conventional (crisp) set theory. Conventional sets contain objects that satisfy precise properties required for membership. Fuzzy sets, on the other hand, contain objects that satisfy imprecisely defined properties to varying degrees. A fuzzy set A in a space of points $Univ = \{u\}$ is a class of events with a continuum of grades of membership and is characterized by a membership function $\mu_A(u)$ which associates with each point in $Univ$ a real number in the interval $[0,1]$ with the value of $\mu_A(u)$ at u representing the grade of membership of u in A. Formally, a fuzzy subset A of the universe $Univ$ is defined as a collection of ordered pairs

$$A = \{(\mu_A(u), u), \forall u \in Univ\}, \tag{2}$$

where the support of A is the subset of $Univ$ and is defined as:

$$S(A) = \{u | u \in Univ \ \& \ \mu_A(u) > 0\}. \tag{3}$$

The characteristic function $\mu_A(u)$, $(0 \leqslant \mu_A(u) \leqslant 1)$, in fact, can be viewed as a weighting coefficient which reflects the ambiguity in making a decision about an event, and as it approaches unity, the grade of membership of an event in A becomes higher. For example, $\mu_A(u) = 1$ indicates a strict containment of the event u in A. If on the other hand, u does not belong to A, $\mu_A(u) = 0$. Any intermediate value would represent the degree to which u could be a member of A.

Relevance of fuzzy set theoretic methods in pattern recognition and image analysis problems has adequately been addressed in the literature [2,12,15,25,37,45,48,50,53,55,56,63,68]. Fuzzy set theories are reputed to handle uncertainties [42], to a reasonable extent, arising from deficiencies of information available from a situation (the deficiency may result from incomplete, ill-defined, not fully reliable, vague and contradictory information). Fuzzy clustering is very useful in

Table 1
HCM algorithm.

Input: Unlabeled data set of n patterns.
Output: c clusters.
Step 1: Select c patterns randomly from the data set as initial cluster centers $\mathbf{v}_i$; $i = 1,2,\ldots,c$.
Step 2: Assign each (unassigned) pattern to the cluster having the nearest cluster center.
Step 3: Compute the mean of assigned patterns to the ith cluster and assign it to the corresponding $\mathbf{v}_i$ ($\forall$ i).
Step 5: Compute $\Delta = \|(\mathbf{V}_t - \mathbf{V}_{t-1})\|$; t denotes tth iteration.
Step 6: Check if $\Delta < \epsilon$; where ϵ is a predefined small positive constant.
Step 7: If the above condition is not true goto Step 2.
Step 8: Stop.

discovering clusters/groups from data with overlapping between them. Contrary to the classical clustering (where one element can be assigned to exactly one group), in fuzzy clustering the elements are assigned not to any one group but to all the groups with certain degree of belonging. This degree of belonging is represented by a numeric value between 0 to 1 and is called "membership grade" which indicates how much an element belongs to a particular cluster. This membership value in hard clustering can be either 1 or 0 (1 is for "belonging" and 0 is for "not belonging").

The basic concepts of image analysis, e.g., the concept of a region, or a boundary or a corner or a relation between regions do not lend themselves to precise definition [52,57,59]. This realization motivated the researchers to develop fuzzy image processing and recognition systems [2,9,20,38,52,53]. A gray tone image possesses ambiguity within each pixel because of the possible multi-valued levels of brightness. If the gray levels are scaled to lie in the range [0,1], we can regard the gray level of a pixel as its degree of membership in the set of high-valued *bright pixels* - thus a gray tone image can be viewed as a fuzzy set. Regions, features, primitives, properties, and relations among them that are not crisply defined can similarly be regarded as fuzzy subsets of images. In our problem we assume that a pixel can belong to both *changed* and *unchanged* classes with certain degrees of membership. The situation is inherently ambiguous as there is no clear boundary between changed and unchanged regions, and we try to solve it using fuzzy clustering [2,62].

Amongst various fuzzy clustering algorithms, fuzzy *c*-means (FCM) [2] is the basic one. Since it has some drawbacks, several algorithms have been developed to improve the performance. Here we will have a comprehensive study on the basic FCM as well as on an improvement done by Gustafson and Kessel [26].

2.2.1. Fuzzy c-means clustering (FCM)

One of the commonly used families of clustering algorithms is the scheme based on function optimization. FCM algorithm belongs to that family. It attempts to find fuzzy partitioning of a given pattern-set by minimizing the objective functional

$$J_m(\mathbf{X};U,\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik})^m D_{ik}, \tag{4}$$

where $U=[\mu_{ik}] \in M_{fcn}$, fuzzy partition matrix of $\mathbf{X}$, and

$$\mathbf{v}_i = \frac{\sum_{k=1}^{n}(\mu_{ik})^m x_k}{\sum_{k=1}^{n}(\mu_{ik})^m} \tag{5}$$

with μ_{ik} (degree of belonging of pattern x_k to the ith cluster) is expressed as

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{(m-1)}}}, \tag{6}$$

where $d_{ik} = \sqrt{D_{ik}}$ and $m(>1)$ is a parameter, called fuzzifier, which controls the fuzziness of the algorithm. Efficiency of FCM is highly dependent on the proper selection of the fuzzifier. For $m \gg 1$ the process is more fuzzy. During optimization of the functional $J_m(\mathbf{X};U,\mathbf{V})$, following two constraints must be satisfied: (i) $\sum_{i=1}^{c}\mu_{ik} = 1$ and (ii) $\mu_{ik} \in [0,1]$. (Thus HCM is a special case of FCM where $\mu_{ik} = 0$ or 1, $\forall$ i and k.)

In FCM we generally normalize the partition U. If we shift the algorithm stated above by one-half cycle so that initialization is done on $\mathbf{V}$ rather than on U, then the type of iteration is called alternating optimization (AO). Here the stopping criterion becomes $\|(\mathbf{V}_t - \mathbf{V}_{t-1})\| < \epsilon$. The resulting algorithm is called FCM-AO. Bezdek [51] proposed that the procedure terminating on U is much more stringent, since many more parameters become closer before termination occurs. The FCM algorithm is put in a tabular form in Table 2.

Singularity occurs in FCM when one or more of the distances $\|(\mathbf{x}_k - \mathbf{v}_i)\| = 0$ at any iteration. This is rare in practice, but if this happens then (6) cannot be evaluated. In this case, it has been suggested in [51] to assign zeros to each nonsingular class and distribute membership to the singular classes arbitrarily subject to the constraint $\sum_{i=1}^{c}\mu_{ik} = 1$.

Though the FCM algorithm is popular for its simplicity and less computation time, it tends to recover clusters with similar sizes and densities and circular shapes. When clusters have non-spherical (ellipsoidal/linear varieties) shapes, FCM fails to

Table 2
FCM algorithm.

Input: Unlabeled data set consisting of n patterns.
Output: Prototypes $\mathbf{V}$ and fuzzy partition matrix U.
Step 1: Set fuzzifier m and initialize U randomly.
Step 2: Compute each cluster center using (5).
Step 3: Compute all possible distances d_{ik}: $i = 1,2,\ldots,c$; $k = 1,2,\ldots,n$.
Step 4: Update fuzzy partition matrix using (6).
Step 5: Compute $\Delta = \|(U_t - U_{t-1})\|$; t denotes tth iteration.
Step 6: Check if $\Delta < \epsilon$; where ϵ is a predefined small positive constant.
Step 7: If the above condition is not true goto Step 2.
Step 8: Stop.

Table 3
GKC algorithm.

Input: Unlabeled data set consisting of n patterns.
Output: Prototypes **V** and fuzzy partition matrix U.
Step 1: Set fuzzifier m, initialize U randomly and ρ_i reasonably.
Step 2: Compute each cluster center using Eq. (5).
Step 3: Compute all the distances d_{ikA_i} using Eq. (8); $i = 1,2,\ldots,c$; $k = 1,2,\ldots,n$.
Step 4: Update fuzzy partition matrix using Eq. (9).
Step 5: Compute $\Delta = \|(U_t - U_{t-1})\|$; t denotes tth iteration.
Step 6: Check if $\Delta < \epsilon$; ϵ is a predefined small positive constant.
Step 7: If the above condition is not true goto Step 2.
Step 8: Stop.

provide good performance. Suitable value of m may improve the performance. In literature [51] suggestion has been given to select the value of this parameter in between 1.5 and 2.5, but it is not applicable for all the data sets [51].

2.2.2. Gustafson–Kessel clustering (GKC)

It is already stated that since FCM employs Euclidean norm to measure the dissimilarity between patterns and cluster centers, only spherical clusters can be detected properly using it. Gustafson and Kessel introduced [26] adaptive distance norm to measure the distance between clusters using fuzzy covariance matrix (a fuzzy equivalent of the classical covariance) - a representation of cluster centers along with data points. Using GKC [26] ellipsoidal clusters can be detected. Each cluster has its own norm-inducing matrix A_i, a positive definite symmetric one, for automatically adapting its shape. The fuzzy covariance matrix F_i of the ith cluster is expressed as:

$$F_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^m (\mathbf{x}_k - \mathbf{v}_i)(\mathbf{x}_k - \mathbf{v}_i)^T}{\sum_{k=1}^{n} (\mu_{ik})^m}. \tag{7}$$

The distance d_{ikA_i} is computed as:

$$d_{ikA_i} = \sqrt{(\mathbf{x}_k - \mathbf{v}_i)^T A_i (\mathbf{x}_k - \mathbf{v}_i)}, \tag{8}$$

where the norm inducing matrix $A_i = [\rho_i det(F_i)]^{1/\eta} F_i^{-1}$, η is the dimension of input patterns. ρ_i is a predefined constant which controls the shape of the i^{th} cluster. Thus

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ikA_i}}{d_{jkA_i}} \right)^{\frac{2}{(m-1)}}}. \tag{9}$$

The objective function J_m will now be of the form

$$J_m(\mathbf{X}; U, \mathbf{V}, A) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m D_{ikA_i}, \tag{10}$$

where $D_{ikA_i} = d_{ikA_i}^2$.

It can be noticed that in the first step of the algorithm, though U is initialized randomly, ρ_i has to be set properly to detect the proper shapes of the clusters. The role of the parameter m is the same as that used in FCM. The algorithm is put in a tabular form (in Table 3).

3. Genetic algorithms and simulated annealing

Genetic algorithms (GAs) [23,49] and Simulated annealing (SA) [40] are two widely used optimization techniques for dealing with objective functions with high non-linearity or having multiple peaks and valleys in their landscapes. Fuzzy clustering algorithms may get stuck to local optima because of random initial configurations. To get rid of this problem *i.e.* to reach to a global optimum with higher probability we have combined fuzzy clustering algorithms with these optimization techniques (GAs and SA). Let us briefly discuss them here.

3.1. Genetic algorithm

When exploring a problem domain with GAs one must specify a set (known as *population*) of proper representation of possible solutions (called individuals or *chromosomes*). It is common to represent a chromosome as a string of 1s and 0s (known as *binary coded* GA) [49]. Initially the *chromosomes* are randomly initialized associated with a fitness value based on a fitness function (*fit_func*). *Crossover* (recombination) and *mutation* are two common operators of GAs. A selection mechanism (*sel_mech*) is adopted to select two chromosomes randomly from the population as a parent pair. Then crossover is

Table 4
Genetic algorithm.

Input:	A set of possible solutions.
Output:	A set of good solutions.
Step 1:	Define *fit_func*, *sel_mech*, Set $cross_{prob}$, mut_{prob}, Initialize *pop* of a reasonable size.
Step 2:	Do the following evolutionary operations on the chromosomes: • *selection* • *crossover* • *mutation*
Step 3:	Check *termin_crtn*. If not met go to Step 2.
Step 4:	Stop.

Table 5
Simulated annealing.

Input:	A randomly chosen configuration.
Output:	A globally optimum configuration.
Step 1:	Randomly select (initialize) a configuration ($\mathbf{Conf}_{current}$) and set temperature T_0 to a high value; 0 denotes 0th iteration.
Step 2:	Compute the cost ($\mathbf{Cost}_{current}$) of the ($\mathbf{Conf}_{current}$) by some procedure.
Step 3:	Perturb ($\mathbf{Conf}_{current}$) by some rule and obtain perturbed configuration ($\mathbf{Conf}_{prtrbd}$).
Step 4:	Compute the cost of $\mathbf{Conf}_{prtrbd}$, $\mathbf{Cost}_{prtrbd}$.
Step 5:	Compute $\Delta C = (\mathbf{Cost}_{\mathbf{prtrbd}} - \mathbf{Cost}_{current})$.

Step 6:	If ($\mathbf{Cost}_{prtrbd}$ <= $\mathbf{Cost}_{current}$)
	$\mathbf{Conf}_{current} = \mathbf{Conf}_{prtrbd}$
	else
	$\mathbf{Conf}_{current} = \mathbf{Conf}_{prtrbd}$ with a probability **exp** $(-\Delta \mathbf{C}/\ T_t)$.

Step 7:	Decrement T_t slightly by some rule.
Step 8:	Check some termination criterion.
Step 9:	If the criterion is not met goto Step 3.
Step 10:	Stop.

done on the selected pair. This leads to exchange of segments of the parents from randomly chosen point(s) (called crossover point(s)) and by this process two offsprings are generated. Crossover is performed depending upon a probability called *crossover probability* ($cross_{prob}$). After crossover the two offsprings are subjected to *mutation*, where the bits of the offsprings are flipped from 0 to 1 or vice versa. Mutation is performed depending upon (a very low) a probability called *mutation probability* (mut_{prob}). The process continues until all the parent pairs undergo the same set of operations and the total process is called a generation. This process ends after some termination criterion (*termin_crtn*) is satisfied. The basic GA stated above is put in a tabular form (in Table 4).

Clusterings (both hard and fuzzy) combined with GAs is a popular approach in research [35,60]. Several techniques and algorithms are suggested in the last two decades (*e.g.* [19,28,29]) to enhance the clustering processes by hybridizing them with GAs.

3.2. Simulated annealing

Simulated annealing (SA) is one of the stochastic relaxation algorithms which simulates the physical annealing procedure [40,44]. It is a global-search procedure. The basic SA algorithm is put in Table 5.

Clusterings combined with SA is also a popular and well-discussed approach in literature [41,67]. A set of randomly initialized cluster representatives is input to the process as candidate solutions. The solution represents the configuration (***Conf***). The ***Cost*** is the value of the corresponding objective function of the clustering. The aim is to get the global optimum solution at the end of process.

4. Validity measures for fuzzy clustering

Validity indices for any clustering can be used from several angles to measure the correctness about the partitioning. It may tell us about the number of natural groups in some data *i.e.* the number of the clusters or it may be used to determine the value of some parameters by which clustering may be affected. A popular index to validate the outcome of a fuzzy clustering proposed by Xie and Beni [69], known as Xie–Beni index, is used widely by many researchers and is described as:

$$v_{XB} = \frac{\sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik})^2 ||\mathbf{x}_k - \mathbf{v}_i||^2}{n\left(\underset{i\neq j}{min}||\mathbf{v}_i - \mathbf{v}_j||^2\right)}. \tag{11}$$

For a good partitioning the index stated here should be minimum. In Eq. (11) the numerator represents the compactness of the fuzzy clusters whereas the denominator is the distance of the two least separated fuzzy clusters. Thus minimizing the numerator (increasing the compactness) and maximizing the denominator (increasing the minimal separation) would suffice the basic intention of clustering. It is less sensitive to the value of m compared to other validity indices like partition coefficient (v_{PC}) of Bezdek and that given by Fukuyama and Sugeno (v_{FS}). Unlike v_{PC} it can consider the data structures of the clusters.

The index proposed by Xie–Beni uses Euclidean norm in its numerator. So, we have used this to evaluate the outcome of FCM (or FCM combined with GA/SA) as FCM employs Euclidean norm for clustering purpose. For GK type we have changed this norm to a scaled Mahalanobis one as in [39] to evaluate the outcome of the process and this is denoted by v_{XBe}. The subscript '*e*' is used to indicate ellipsoidal nature of clusters.

5. Proposed change detection technique

Remotely sensed images are normally stored in *gray scale* format. It is common to represent them in 8-bit gray scale, having 256 (2^8) shades of gray levels. In our experiment also we have used two data sets of the described format where every pixel is of a gray shade between 0 to 255 (0 represents black and 255 white).

Let us consider two coregistered and radiometrically corrected multispectral images X_1 and X_2 of size $p \times q$, acquired over the same geographical area at two different time instants t_1 and t_2, and let $DI = \{l_{(m,n)}, 1 \leqslant m \leqslant p, 1 \leqslant n \leqslant q\}$ be the difference image obtained by applying the CVA technique as follows:

$$P_{DI_{(m,n)}} = \sqrt{\sum_{i=1}^{num}\left(P_{\mathbf{X_1}_{(m,n)_{b_i}}} - P_{\mathbf{X_2}_{(m,n)_{b_i}}}\right)^2}, \tag{12}$$

where $P_{DI_{(m,n)}}$ is the gray value of the (m,n)th pixel in the difference image generated from corresponding pixels of the images $\mathbf{X}_1$ and $\mathbf{X}_2$ having *num* bands $b_1, b_2, \ldots, b_{num}$. A specific section of the electromagnetic spectrum (of the order of micrometer in our case) is called a band. Generating a difference image by using several bands allows us to combine the information about reflectance properties of the land cover types (soil, vegetation, water *etc.*) at different wavelengths of light.

To exploit the spatio-contextual information, input patterns are generated corresponding to each pixel in the difference image *DI*, considering its spatial neighborhood system N^d of order d. For a given spatial position (m,n), $N^d(m,n)$ is defined as

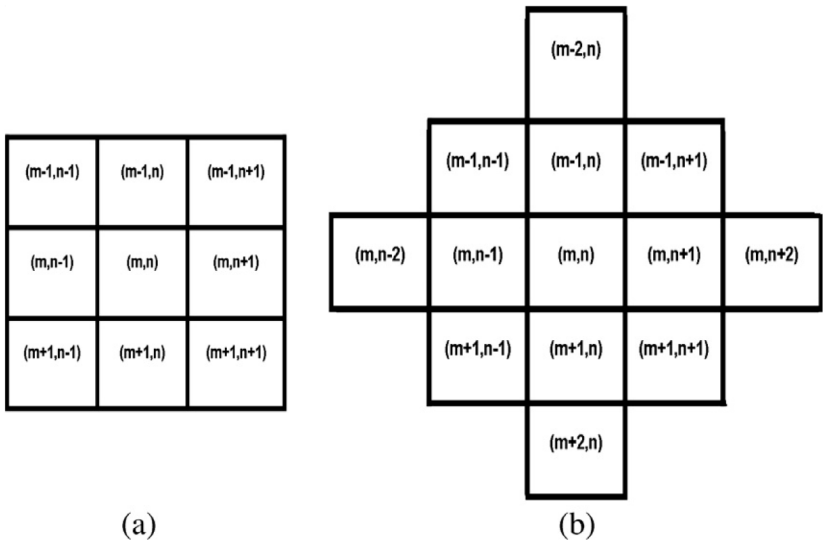

Fig. 1. Neighborhood of the pixel at position (m,n). (a) $\mathbf{N}^2$ and (b) $\mathbf{N}^3$.

follows: $N^d(m,n) = \{(m,n) + (i,j), (i,j) \in N^d\}$. Figs. 1(a) and (b) depict the second (N^2) and third (N^3) order neighborhood systems of a pixel at position (m,n).

In the present work we have generated a two dimensional pattern corresponding to the pixel at position (m,n) of *DI* by considering the gray value of that pixel as one feature and the average of the gray value of the 8 neighbors in N^2 (neighbors are within a distance of 2 from the center pixel of the system) as another feature. We may work upon some higher order ($d > 2$) neighborhood system (like N^3) to generate this feature also; but in those cases computational cost will be more. Since each pixel generates a pattern, the total number of patterns will be $p \times q$.

After generating the pattern-set we have used two fuzzy clustering algorithms discussed earlier namely FCM and GKC to generate two clusters. We have used HCM also to establish the effectiveness of using fuzzy clustering algorithms. FCM assumes that the clusters are spheroidal, while GKC tries to extract the exact shape of clusters by varying some parameters. To overcome the pitfalls of clusterings (hard/fuzzy) like getting stuck at some suboptimal points in their search spaces due to initial configurations, we have hybridized them with GA and SA. After generating two clusters by a clustering model, one has to be marked as *changed* and the other as *unchanged*. For this purpose we have calculated the mean values of the two clusters and the cluster whose center is closer to the origin (of the feature space) is labeled as *unchanged* and the other one as *changed*. The pixels corresponding to *changed* one are marked as black (*gray level* 0) and the *unchanged* ones are marked as white (*gray level* 255) in the generated *change detection map*.

Combination of HCM with GA is termed as G_HCM (Genetic HCM) and that with SA as SA_HCM (simulated annealing HCM) in the rest of this paper. For fuzzy clustering also we have named the processes in a similar way, combining FCM with GA results in G_FCM and with SA, SA_FCM. For GKC we only worked upon combining with SA (SA_GKC) and not with GA. Section 7 will highlight the practical difficulties behind it.

6. Description of the data sets

In order to carry out the experimental analysis aimed to assess the effectiveness of the proposed approach, we considered two multitemporal remote sensing data sets corresponding to geographical areas of Mexico and Island of Sardinia, Italy. The spatial resolution of the sensors (ETM+ of LANDSAT-7 and TM of LANDSAT-5) is 30 meters. Each pixel thus represents an area of 30 m $\times$ 30 m. The information regarding the north direction of the data could not be made available from the authority. A detailed description of each of the data sets is given below.

6.1. Data set related to mexico area

The first data set used in our experiment is made up of two multispectral images acquired by the Landsat Enhanced Thematic Mapper Plus (ETM+) sensor of the Landsat - 7 satellite in an area of Mexico on 18th April 2000 and 20th May 2002. From the entire available Landsat scene, a section of 512 $\times$ 512 pixels has been selected as test site. Between the two aforementioned acquisition dates, fire destroyed a large portion of the vegetation in the considered region.

Figs. 2(a) and (b) show channel 4 of the 2000 and 2002 images, respectively. In order to make a quantitative evaluation of the effectiveness of the proposed approach, a reference map was manually defined (see Fig. 2(d)) according to a detailed visual analysis of both the available multitemporal images and the difference image (see Fig. 2(c)). Different color composites of the above mentioned images were used to highlight all the portions of the changed area in the best possible way. This procedure resulted in a reference map containing 25,599 changed and 2,36,545 unchanged pixels. Experiments were carried out to produce, in an automatic way, a change detection map as similar as possible to the reference map that represents the best result obtainable with a time consuming procedure.

Analysis of the behavior of the histograms of multitemporal images did not reveal any significant difference due to light and atmospheric conditions at the acquisition dates. Therefore, no radiometric correction algorithm was applied. The 2002 image was registered on the 2000 one using 12 ground control points. The procedure led to a residual average misregistration error on ground control points of about 0.3 pixels.

6.2. Data set related to Sardinia island, Italy

The second data set used in our experiment is composed of two multispectral images acquired by the Landsat Thematic Mapper (TM) sensor of the Landsat - 5 satellite in September 1995 and July 1996. The test site is a section of 412 $\times$ 300 pixels of a scene including lake Mulargia on the Island of Sardinia (Italy).

Between the two aforementioned acquisition dates, water level in the lake increased (see the lower central part of the image). Figs. 3(a) and (b) show channel 4 of the 1995 and 1996 images, respectively. As in the case of Mexico data set, in this case also a reference map was manually defined (see Fig. 3(d)) according to a detailed visual analysis of both the available multitemporal images and the difference image (see Fig. 3(c)). At the end, 7480 changed and 1,16,120 unchanged pixels were identified. As the histograms did not show any significant difference, no radiometric correction algorithms were applied to the multitemporal images. The images were coregistered with 12 ground control points resulting in an average residual misregistration error of about 0.2 pixels on the ground control points.

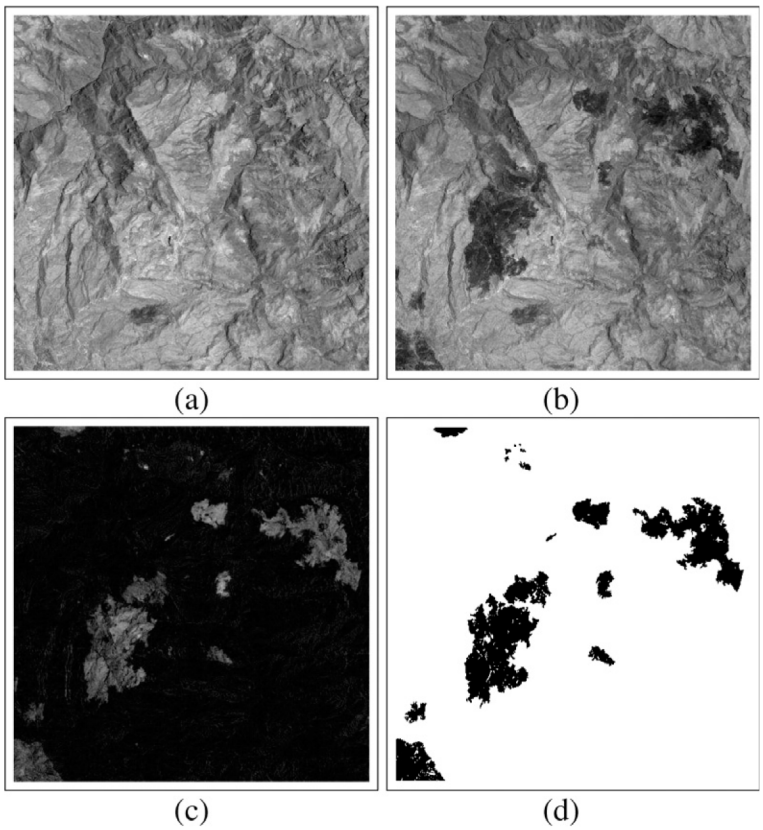

Fig. 2. Image of Mexico area. (a) Band 4 of the Landsat ETM+image acquired in April 2000, (b) band 4 of the Landsat ETM+ image acquired in May 2002, (c) corresponding difference image generated by CVA technique, and (d) reference map of the changed area.

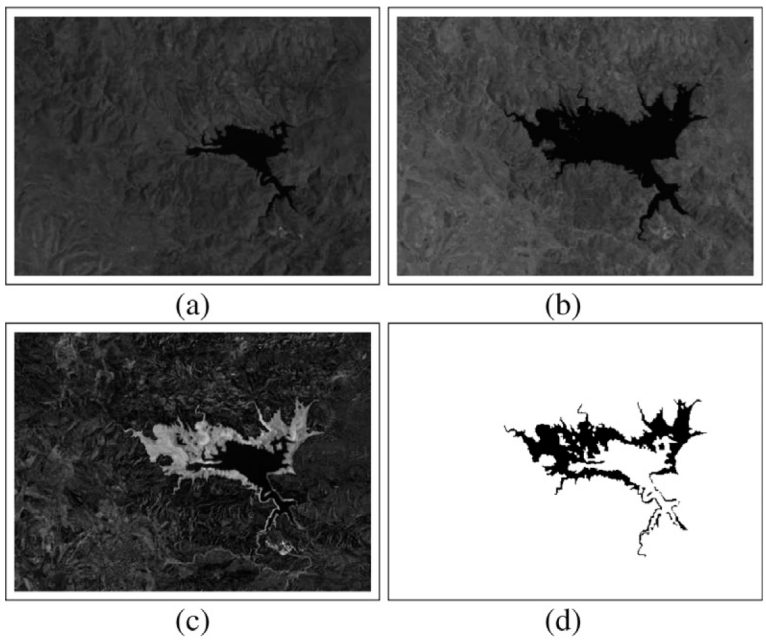

Fig. 3. Image of Sardinia island, Italy. (a) Band 4 of the Landsat TM image acquired in September 1995, (b) band 4 of the Landsat TM image acquired in July 1996, (c) difference image generated by CVA technique using bands 1, 2, 4, and 5; and (d) reference map of the changed area.

7. Experiments and results

To assess the effectiveness of the proposed approach, we have made both qualitative (visual) and quantitative analysis of the experimental results. In visual analysis we have compared *change detection map* (a binary image) with the ground truth image (again, a binary one). We then presented a quantitative analysis with respect to *overall error*. We made a comparative study of the performances of our proposed algorithms with those of one context-insensitive technique and two context-sensitive techniques. The context-insensitive technique (MTET) as mentioned earlier produces a minimum error change detection map by finding an optimal decision threshold for DI assuming pixels to be independent in spatial domain. Context-sensitive techniques compared are: (i) the technique presented in [5] where EM is combined with MRF and will be referred as EM+MRF and (ii) a technique based on "Hopfield-Type Neural Networks" [21] (will be referred as HTNN). In [21] four different Hopfield-type network models were presented. We have compared with 2nd order continuous model as we also have used 2nd order neighborhood information in the present work.

For Mexico data set, the DI is generated using the CVA algorithm by considering band 4 as it is reported to be very effective to locate burned area. For Sardinia data set, the DI is generated by the CVA algorithm using spectral bands 1, 2, 4 and 5. As mentioned earlier, for both the fuzzy clustering techniques the value of fuzzifier (m) affects the results. For GKC algorithm ρ_i also affects the results. Here we have presented the best results obtained by varying these parameters. We have varied m starting from 1.1 taking an incremental step size of 0.1 in case of FCM and stopped at that point where the performance of the algorithm started degrading. In case of GKC maintaining the same criterion for varying m, we have fixed one ρ_i to 1 and varied the other one from 1 taking an incremental step size of 0.1 and vice versa up to that extent where we have achieved the best result. For all the experiments ϵ is set to 0.0000001.

7.1. Hybridization with GA

As mentioned above the fuzzy clustering algorithms function better for some values of the parameters (m for FCM and m and ρ_i for GKC). While combining them with GA or SA we have considered those values of the parameters only.

When dealing with clusterings combined with GA it is common to represent a chromosome as a candidate solution that consists of the representatives of the clusters. Clustering algorithms are associated with an objective function. As better clustering promises the corresponding functional to be minimized, we have chosen the fitness function as the functional itself as in [29].

In this work the chromosomes are binary representations of the two features of the two cluster centers (*changed* and *unchanged*) in some specific order. Our remotely sensed images are of 8-bit. So, the first feature of a generated pattern can be represented by a minimum of 8 bits. The second feature of the pattern is a real valued number (Ref. Section 5). We have represented this by 11 bits where first 8 bits are to represent the integer part and last 3 bits are for floating part of the same for a better precision. So a chromosome is of 38 bits (19 bits representing a cluster center).

A population size of 30 has been taken in our experiments. We have used 2-fold tournament selection [23] in our experiments to select the two parent chromosomes to reproduce two offsprings. Here two chromosomes are randomly sampled from the current population. The one having higher fitness is chosen as the first parent. The same procedure is repeated to select the second one. For better exploration two point crossover (as two class problem is involved) is performed as reported in literatures [27]. $cross_{prob}$ and mut_{prob} are set to 0.9 and 0.01 respectively. An elitist strategy [49] is adopted to preserve the best solution; and the process terminates when the fitness of the best individual of a population is not changed for consecutive 50 generations.

7.2. Hybridization with SA

In SA only one candidate solution is considered for optimization purpose, unlike GA. A set of randomly chosen cluster representatives (means for HCM and fuzzy means for FCM and GKC) is determined such that the corresponding clusters become compact. The set represents the configuration ***Conf*** and the corresponding value of the associated objective functional (of the concerned clustering model) is the ***Cost*** (Ref. Section 3.2) at that moment. To perturb the $\boldsymbol{Conf}_{\boldsymbol{current}}$ a randomly chosen cluster representative (one $\mathbf{v}_i$) is changed slightly by adding a random number from a normal distribution with mean 0 and standard deviation 1.

If infinite time is allotted to SA it can find the global or near global minimum without getting trapped to a local one. This requires starting from a very high temperature (T) or control parameter ($\rightarrow\infty$), a slow reduction of the temperature and a very low ($\rightarrow 0$) final temperature. This is the sufficient condition for this technique to reach to a global optimum. In [41,67] detailed discussions suggest that it may not be the necessary condition for better results. Several authors have proposed many heuristics to overcome this computational burden. We have chosen one of them. We have followed the concept of acceptance ratio (χ_0) which is defined as the ratio of number of accepted transitions divided by the proposed transitions [67] to set the initial temperature. To cool the system we have chosen the *linear cooling scheme* where $T_{t+1} = (T_t - lin_c)$ (t denotes tth iteration). It has been reported in literatures [41] that SA is very much sensitive to the random numbers generated (responsible for the nature of moves) as well as its success highly depends upon the scaling of the function it tries to optimize and the current temperature T_t (having dependency on the T_0) during iterations [67] because all the factors control the probability of acceptance when an uphill movement is made. The approach we have followed is that for some certain set of random numbers, choosing lin_c in between 0.001 to 0.01 and minimum T_0 as 10. If χ_0 falls below 0.8 we doubled our initial

temperature T_0 to start the process. Number of iterations is chosen depending on the value of lin_c to prevent temperature from being negative.

Now we will explore the reason why it was not possible to hybridize GKC with GA. In both the optimizations (GA and SA) we have to evaluate $J_m(\mathbf{X}; U, \mathbf{V}, A)$ at every iteration. For this purpose we need U and $\mathbf{V}$ both and they must correspond to each other. In SA we have generated random numbers of small variation (from a normal distribution with mean 0 and standard deviation 1) and only one fuzzy mean has been perturbed. Thus new $\mathbf{v}_i$ and the previous μ_{ik} are highly correlated. So, in SA_GKC, we have evaluated $J_m(\mathbf{X}; U, \mathbf{V}, A)$ by taking the new $\mathbf{v}_i$ and previous values of μ_{ik}. As in GA crossover operation is involved, so $\mathbf{v}_i$ can get changed drastically and the above mentioned correspondence between μ_{ik} and $\mathbf{v}_i$ may get completely lost. Therefore we can not use the evolved $\mathbf{v}_i$ to calculate the new d_{ikA_i} as the scaled Mahalanobis distance (d_{ikA_i}) is a function of the present membership values (μ_{ik}) and corresponding fuzzy means ($\mathbf{v}_i$) (can be verified from Eqs. (7) and (8)) and so the new μ_{ik}. Also we can not use the previous μ_{ik} to compute the value of $J_m(\mathbf{X}; U, \mathbf{V}, A)$ directly. So we have ignored this hybridization.

7.3. Visual analysis

The *change detection maps* obtained for Mexico data set by the HCM, FCM and GKC are shown in Figs. 4 and 5(a) and (b), respectively. Figs. 6 and 7(a) and (b) are the change detection maps obtained for Sardinia island data by using HCM, FCM and GKC respectively.

One can visually compare the *change detection map* generated by the proposed algorithms with the corresponding ground truth. This gives a rough idea about the quality of the generated change detection map.

The effectiveness of the proposed *change detection technique* is evaluated globally by analyzing the *change detection map*. The *change detection map* obtained by HCM seems to be better than FCM for Mexico data set because of the *false alarms* (discussed later in "Quantitative Analysis" section) generated over wider portions by the FCM. But careful inspection reveals that the extreme bottom-left corner of the scene is detected in a much better way by FCM, where HCM failed. For Sardinia data set visual inspection of the *change detection map* shows that FCM is doing better than HCM. Similarly GKC is doing better than FCM. This directs us to judge the effectiveness of the proposed technique quantitatively, which obviously is better than visual inspection and is presented in the following section.

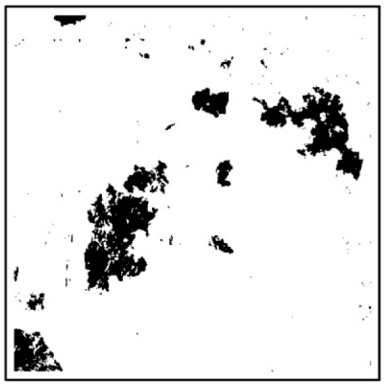

Fig. 4. Change detection map obtained for Mexico data set by HCM.

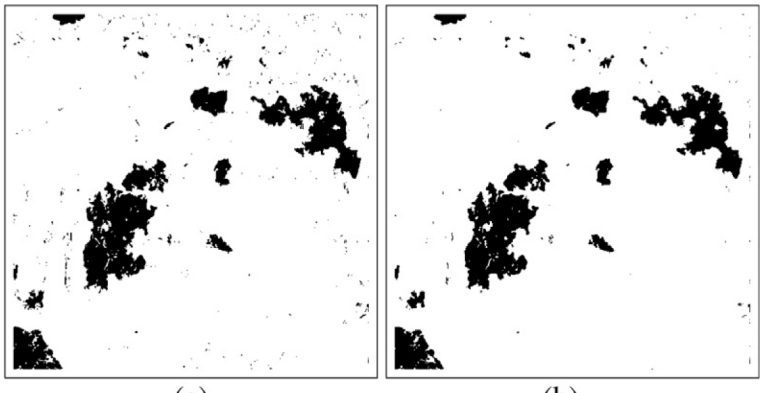

Fig. 5. Change detection maps obtained for Mexico data set by (a) FCM ($m = 14.0$) and (b) GKC ($m = 15.5$ and $\rho_1 = 1$ and $\rho_2 = 2.2$).

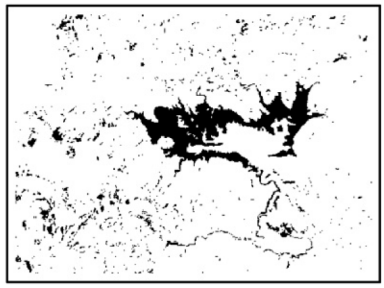

Fig. 6. Change detection map obtained for Sardinia data set by HCM.

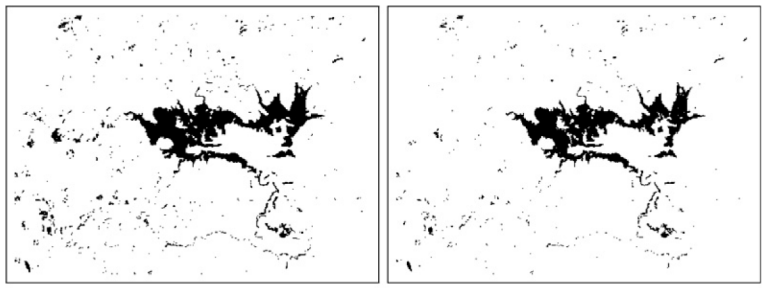

Fig. 7. Change detection maps obtained for Sardinia data set by (a) FCM ($m = 1.1$) and (b) GKC ($m = 2$ and $\rho_1 = 3.5$ and $\rho_2 = 1$).

7.4. Quantitative analysis

Quantitative analysis is carried out in terms of both *overall error* (*OE*), *false alarms* (*i.e.*, unchanged pixels identified as changed ones - *FA*) and *missed alarms* (*i.e.*, changed pixels categorized as unchanged ones - *MA*). It is better to have less *MA* because it denotes the actual changes that the algorithm failed to detect. Also *OE* should be as small as possible. We have compared the results (different types of alarms) generated by crisp clustering (HCM) and by fuzzy clustering algorithms (FCM and GKC) using the "ground truth" image as the reference map. In Tables 6, 9 and 12 we put the overall change-detection error obtained by the HCM, FCM and GKC techniques respectively (we have put the best results obtained by varying the parameters).

While combining FCM and GKC with GA and SA we have used those parameter values for which they were giving better results. Results produced by the hybridization processes (of FCM and GKC with GA or SA) are put in Tables 10, 11 and 13. For comparison between hard and fuzzy clustering, we have optimized the functional of HCM and put the results in Tables 7 and 8. From the results it is clear that using the hybrid procedures, results are improved. It is observed from the results that for our data sets SA is working better than GA to detect the changes.

From the results it is obvious that for *change detection* purpose it is better to use fuzzy clusterings (FCM or GKC) rather than crisp clustering (HCM). Overall error incurred by FCM is less than that obtained by HCM in all cases; and GKC incurred

Table 6
Missed alarms, false alarms and overall error obtained by HCM.

Data set	*MA*	*FA*	*OE*
Mexico	3425	747	4172
Sardinia	275	4133	4408

Table 7
Missed alarms, false alarms and overall error obtained by G_HCM.

Data set	*MA*	*FA*	*OE*
Mexico	2827	785	3612
Sardinia	3006	132	3138

Table 8
Missed alarms, false alarms and overall error obtained by SA_HCM.

Data set	*MA*	*FA*	*OE*
Mexico	1965	1095	3060
Sardinia	425	2727	3152

Table 9
Missed alarms, false alarms, overall error and validity index obtained by FCM.

Data set	*MA*	*FA*	*OE*	m	v_{XB}
Mexico	1164	1764	2928	14.0	0.08
Sardinia	494	2246	2740	1.1	0.30

the least error. It is noticed that the results are improved in all cases for GKC. Since GKC uses Mahalanobis distance and can extract even non-convex clusters, it produced better results for the used data sets (since the data mainly contain irregular shaped *changed* regions). Again by varying ρ_i (instead of fixing to 1) in GKC, shapes of the corresponding clusters can be approximated more accurately. The reason being, by varying the parameter (ρ_i) with respect to each other (*i.e.* fixing one to 1 and varying the other) the assumed shapes of the clusters could be varied. This is one of the major factors for GKC to work well for this problem. This suggests that incorporation of the knowledge that the *changed* and *unchanged* classes are very unstructured and are of different shapes, may be helpful while solving *change detection* problems. It has been seen that the proposed technique works well within a range of values of the parameters (m, ρ_1 and ρ_2).

From Tables 9–13 we can see the behavior of v_{XB} and v_{XBe}. For Sardinia data set we can see that when changes are detected properly it is reflected on the value of the index v_{XB} *i.e.* it is decreasing (except the case of G_FCM). On the other hand for Mexico data set the v_{XB} index is constant (0.8 for FCM, G_FCM and SA_FCM) but v_{XBe} is showing better results. The value of m used for Mexico data is large whereas for the Sardinia it is small. As high value of m suggests more vagueness (and overlapping of the two clusters) in data, so it would not be incorrect to infer that the two clusters are more unstructured in Mexico data set than the other one. Thus, v_{XBe} is a better choice for our purpose to extract the clusters by assuming them as elliptical as this index employs Mahalanobis norm.

As mentioned earlier we compared the results obtained by the proposed fuzzy techniques with those obtained using EM+MRF [5], MTET [5] and HTNN [21]. To have a comparative study among these and the proposed techniques we have

Table 10
Missed alarms, false alarms, overall error and validity index obtained by G_FCM.

Data set	*MA*	*FA*	*OE*	m	v_{XB}
Mexico	1148	1721	2869	14.0	0.08
Sardinia	627	1983	2610	1.1	0.30

Table 11
Missed alarms, false alarms, overall error and validity index obtained by SA_FCM.

Data set	*MA*	*FA*	*OE*	m	v_{XB}
Mexico	1095	1725	2790	14.0	0.08
Sardinia	915	975	1890	1.1	0.26

Table 12
Missed alarms, false alarms, overall error and validity index obtained by GKC.

Data set	*MA*	*FA*	*OE*	m	$\boldsymbol{\rho}_1$	$\boldsymbol{\rho}_2$	v_{XBe}
Mexico	1453	1130	2583	15.5	1	2.2	0.029355
Sardinia	826	1017	1843	2	3.5	1	0.210041

Table 13
Missed alarms, false alarms, overall error and validity index obtained by SA_GKC.

Data set	*MA*	*FA*	*OE*	m	$\boldsymbol{\rho}_1$	$\boldsymbol{\rho}_2$	v_{XBe}
Mexico	947	1587	2534	15.5	1	2.2	0.028852
Sardinia	1139	558	1697	2	3.5	1	0.196665

Table 14
Missed alarms, false alarms and overall error for Mexico data set.

Techniques used	*MA*	*FA*	*OE*
MTET	2404	2187	4591
HCM	3425	747	4172
G_HCM	2827	785	3612
SA_HCM	1965	1095	3060
HTNN	558	2707	3265
EM+MRF (β = **1.5**)	946	2257	3203
FCM (m = **14.0**)	1164	1764	2928
G_FCM (m = **14.0**)	1148	1721	2869
SA_FCM(m = **14.0**)	1065	1725	2790
GKC (m = **15.5**, ρ_1 = 1, ρ_2 = 2.2)	1453	1130	2583
SA_GKC (m = **15.5**, ρ_1 = **1**, ρ_2 = **2.2**)	**947**	**1587**	**2534**

Table 15
Missed alarms, false alarms and overall error for Sardinia data set.

Techniques used	*MA*	*FA*	*OE*
MTET	1015	875	1890
HCM	275	4133	4408
G_HCM	3006	132	3138
SA_HCM	425	2727	3152
HTNN	1187	722	1909
EM+MRF (β = **2.2**)	592	1108	1700
FCM (m = **1.1**)	494	2246	2740
G_FCM (m = **1.1**)	627	1983	2610
SA_FCM (m = **1.1**)	915	975	1890
GKC (m = 2, ρ_1 = **3.5**, ρ_2 = **1**)	826	1017	1843
SA_GKC (m = 2, ρ_1 = **3.5**, ρ_2 = **1**)	**1139**	**558**	**1697**

put the results in Tables 14 and 15. From these tables it is seen that for the two data sets (Mexico and Sardinia) the SA_GKC technique performed the best (OE = 2534 for Mexico and OE = 1697 for Sardinia). Both showed higher performance than EM+MRF which is a computationally intensive procedure.

From the overall analysis it is felt that GK type clusterings (GKC and SA_GKC) can be a reasonable choice for generating change detection maps. The existing techniques produced similar performance with the proposed ones, but require either the assumption of distributions of classes and is very time consuming (EM+MRF) or needs more time (HTNN). On the other hand the proposed technique does not require any *a priori* knowledge of the data distributions and is very fast.

7.5. Effect of parameters

To analyze the effect of the parameters in a better way we have plotted OE against m in Fig. 8 for both the algorithms for Mexico data set. For GKC the plot is for constant ρ_i (taking ρ_1 = 1 and ρ_2 = 2.2). Fig. 9 shows the plot of OE against ρ_2 for

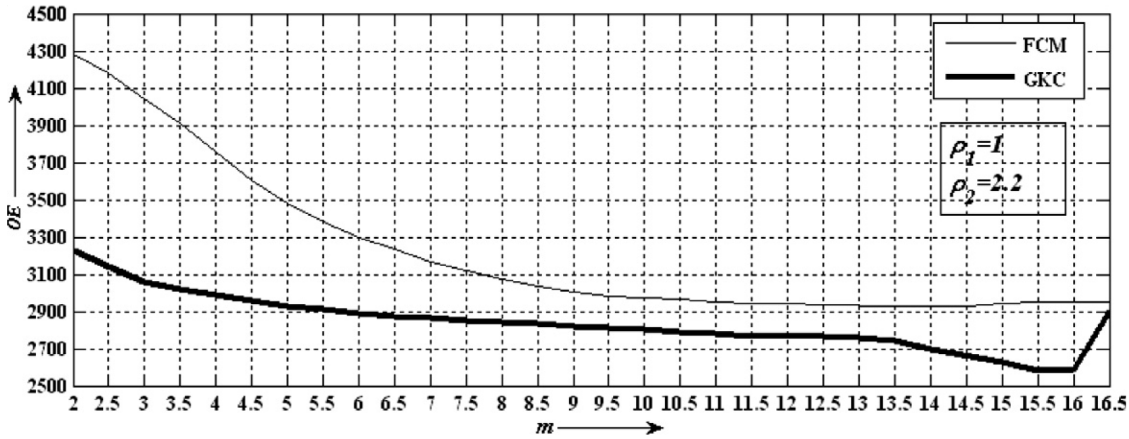

Fig. 8. Plot of OE as a function of m.

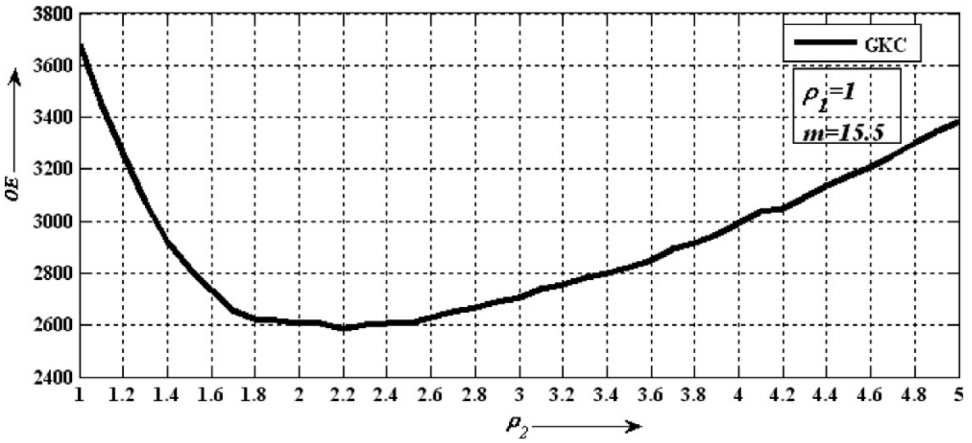

Fig. 9. Plot of *OE* as a function of ρ_2.

constant ρ_1(=1) and m(=15.5). We have tested our technique taking the parameter values also outside the ranges shown. The graph is shown for a reasonable range for which the *overall error* is less.

Let us explain the graphs one by one. Fig. 8 tells us how the value of the fuzzifier m influences the result. Looking at the graph of FCM we can say that when m is increased beyond 2, *OE* reduces gradually. Up to $m = 14.0$ *OE* reduces and at that point FCM gives the least *OE*. Beyond the mentioned point *OE* starts increasing. As seen form the figure, variations of *OE* is negligible for m in the range 11–15; and thus one can select any value of m to have a reasonable performance. The range showing less variation may be interpreted as linear region of the graph. The similar behavior can be observed for GKC also. One can see that if proper values of ρ_i are chosen then GKC performs in a well manner. As in FCM, here also we can notice that from $m = 2$ onwards the performance is increasing slowly (*i.e. OE* is decreasing) in a continuous manner. For this algorithm *OE* continually reduced up to $m = 15.5$ where GKC showed its best performance. Beyond this *OE* started increasing. From the graph, corresponding to GKC, we can suggest that the constant *OE* region is from 11 up to 13.5.

Though the value of the fuzzifier affects the result in the same manner for both the algorithms FCM and GKC, and for both the cases we are having some linear section, GKC shows better performance than FCM over the entire range. In case of FCM we are having a larger linear range compared to GKC, with lesser performance.

Fig. 9 tells us about the effect of ρ_i for some constant value of m, the fuzzifier. We have set $m = 15.5$ as it gave us the least *OE*. We have plotted *OE* versus ρ_2 by fixing $\rho_1 = 1$. The reverse *i.e. OE* versus ρ_1 can also be done. Looking at the graph one can notice that when ρ_2 is increased beyond 1, *OE* reduces continuously up to 2.2; and if further increased then *OE* increases. Minute observations reveal that for ρ_2 in the range 2 to 2.5 the *OE* value is order of 2600. Thus one can accept any value of ρ_2 in this range to have a reasonable performance.

Now the question arises why are we interested about the stable regions of the parameters! The answer is one can use these ranges of values for similar types of images. Similar findings can be observed for other data sets also.

8. Discussion and conclusion

Unsupervised context-sensitive techniques using fuzzy clustering approach for detecting changes in multitemporal, multispectral remote sensing images have been proposed in this paper. Since the pixels of the difference image belonging to the two clusters (*changed* and *unchanged*) are not separable by sharp boundaries (as they are highly overlapped), fuzzy clustering techniques seem to be more appropriate and realistic choice to separate them. From the results also it is noticed that fuzzy clustering is a better option than crisp clustering. Among the two fuzzy clustering algorithms, only fuzzy *c*-means and Gustafson–Kessel (GK) are used in the present experiment. To prevent from getting stuck to local optima the fuzzy clustering algorithms are combined with genetic algorithm and simulated annealing. As GK type clustering can extract clusters with different (including non-spherical) shapes it is found to be more effective. A fuzzy cluster validity index namely Xie–Beni index has been used to validate the results of the change detection problem.

The proposed fuzzy clustering techniques have advantages over the context-sensitive process (EM+MRF) presented in [5] as they are distribution free (need not require any explicit assumption about the underlying two classes, *changed* and *unchanged*) as well as they are less computation intensive. Compared to another context-sensitive technique proposed in [21], the fuzzy techniques proposed here are very simple, less costly and showed improved performance.

Though it is seen from visual as well as quantitative analysis, that methods based on fuzzy clustering are well suited for change detection of remotely sensed data, there are some unavoidable problems also - proper selection of the values of fuzzifiers m and ρ_i. Previous domain knowledge may be useful in fixing up the parameter values. In future we hope to explore this issue. Other fuzzy clustering algorithms, like fuzzy c-varieties/elliptotypes which extract linear substructures in data *i.e.*, line-shaped clusters (elongated shapes) may be worth exploring for this problem also.

Acknowledgements

The authors thank the anonymous referees for their constructive criticism and valuable suggestions. They also like to thank the Department of Science and Technology, Government of India and University of Trento, Italy, the sponsors of the ITPAR programs and Prof. L. Bruzzone, the Italian collaborator of the project, for providing the data.

References

[1] Y. Bazi, L. Bruzzone, F. Melgani, An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images, IEEE Transactions on Geoscience and Remote Sensing 43 (4) (2005) 874–887.
[2] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function, Plenum Press, New York, 1981.
[3] L. Bruzzone, D.F. Prieto, An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images, IEEE Transactions on Geoscience and Remote Sensing 11 (4) (2002) 452–466.
[4] L. Bruzzone, D. Fernàndez Prieto, An adaptive parcel-based technique for unsupervised change detection, International Journal of Remote Sensing 21 (4) (2000) 817–822.
[5] L. Bruzzone, D. Fernàndez Prieto, Automatic analysis of the difference image for unsupervised change detection, IEEE Transactions on Geoscience and Remote Sensing 38 (3) (2000) 1171–1182.
[6] L. Bruzzone, S.B. Serpico, An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images, IEEE Transactions on Geoscience and Remote Sensin 35 (4) (1997) 858–867.
[7] M.J. Canty, Image Analysis, Classification and Change Detection in Remote Sensing, CRC Press, Taylor & Francis, 2006.
[8] P.S. Chavez Jr., D.J. MacKinnon, Automatic detection of vegetation changes in the southwestern United States using remotely sensed images, Photogrammetric Engineering and Remote Sensing 60 (5) (1994) 1285–1294.
[9] H.D. Cheng, H. Xu, A novel fuzzy logic approach to mammogram contrast enhancement, Information Sciences 148 (1-4) (2002) 167–184.
[10] J. Cihlar, T.J. Pultz, A.L. Gray, Change detection with synthetic aperture radar, International Journal of Remote Sensing 13 (3) (1992) 401–414.
[11] X. Dai, S. Khorram, The effects of image misregistration on the accuracy of remotely sensed change detection, IEEE Transactions on Geoscience and Remote Sensing 36 (5) (1998) 1566–1577.
[12] J.V. de Oliveira, W. Pedrycz (Eds.), Advances in Fuzzy Clustering and its Applications, Wiley, 2007.
[13] P. Deer and P. Eklund, Values for the fuzzy c-means classifier in change detection for remote sensing, in: Proceedings of the Ninth International Conference on Information Processing and Management of Uncertainty, 2002, pp. 187–194.
[14] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society 39 (1) (1977) 1–38.
[15] D. Dubois, H. Prade, Fuzzy Sets and Systems: Theory and Applications, Academic Press, New York, 1980.
[16] R. Duda, P. Hart, D. Stork, Pattern Classification, Wiley & Sons (Asia) Pte. Ltd, Singapore, 2006.
[17] M.T. Eismann, J. Meola, R.C. Hardie, Hyperspectral change detection in the presence of diurnal and seasonal variations, IEEE Transactions on Geoscience and Remote Sensing 46 (1) (2008) 237–249.
[18] T. Fung, An assessment of TM imagery for land-cover change detection, IEEE Transactions on Geoscience and Remote Sensing 28 (4) (1990) 681–684.
[19] A. Ghosh, S. Dehuri, S. Ghosh (Eds.), Multi-Objective Evolutionary Algorithms for Knowledge Discovery from Databases, Springer, Berlin, 2008.
[20] A. Ghosh, N.R. Pal, S.K. Pal, Self-organization for object extraction using multilayer neural network and fuzziness measures, IEEE Transactions on Fuzzy Systems 1 (1) (1993) 54–68.
[21] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo, A. Ghosh, A context-sensitive technique for unsupervised change detection based on Hopfield-type neural networks, IEEE Transactions on Geoscience and Remote Sensing 45 (3) (2007) 778–789.
[22] S. Ghosh, S. Patra, A. Ghosh, An unsupervised context-sensitive change detection technique based on modified self-organizing feature map neural network, International Journal of Approximate Reasoning 50 (2009) 37–50.
[23] D.E. Goldberbg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, 1989.
[24] S. Gopal, C. Woodcock, Remote sensing of forest change using artificial neural networks, IEEE Transactions on Geoscience and Remote Sensing 34 (2) (1996) 398–404.
[25] P. Guo, Fuzzy data envelopment analysis and its application to location problems, Information Sciences 179 (6) (2009) 820–829.
[26] D.E. Gustafson, W.C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, in: IEEE Conference on Decision and Control, San Diego, CA, 1979, pp. 761–766.
[27] L.O. Hall, J.C. Bezdek, S. Boggavarpu, and A. Bensaid. Genetic fuzzy clustering, in: NAFIS '94: Proceedings of the First International Joint Conference of the North American Fuzzy Information Processing Society Biannual Conference, 1995, pp. 411–415.
[28] L.O. Hall and B. Ozyurt, Scaling genetically guided fuzzy clustering, in: ISUMA-NAFIS '95: Proceedings of the 3rd International Symposium on Uncertainty Modelling and Analysis, 1995, pp. 328–332.
[29] L.O. Hall, B. Ozyurt, J.C. Bezdek, Clustering with a genetically optimized approach, IEEE Transactions on Evolutionary Computation 3 (1999) 103–112.
[30] T. Hame, I. Heiler, J.S. Miguel-Ayanz, An unsupervised change detection and recognition system for forestry, International Journal of Remote Sensing 19 (6) (1998) 1079–1099.
[31] J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kauffann, India, 2008.
[32] T. Han, M.A. Wulder, J.C. White, N.C. Coops, M.F. Alvarez, C. Butson, An efficient protocol to process landsat images for change detection with tasselled cap transformation, IEEE Transactions on Geoscience and Remote Sensing 4 (1) (2007) 147–151.
[33] Z. Hong, Q. Jiang, H. Dong, S. Wang, J. Li, An improved FCM-based model for urban change detection using high-resolution remotely sensed images, Environmental Informatics Archives 3 (ISEIS Publication Series Number P002) (2005) 352–359.
[34] A.K. Jain, R.C. Dubes, Algorithms for Clustering Data, Prentice-Hall, Englewood Cliffs, NJ, 1988.
[35] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM Computing Surveys 31 (3) (1999) 264–323.
[36] T. Kasetkasem, P.K. Varshney, An image change detection algorithm based on Markov random field models, IEEE Transactions on Geoscience and Remote Sensing 40 (8) (2002) 1815–1823.
[37] A. Kaufmann, Fuzzy Subsets - Fundamental Theoretical Elements, Academic Press, New York, 1980.
[38] J.M. Keller, C.L. Carpenter, Image segmentation in the presence of uncertainty, International Journal of Intelligent Systems 5 (1990) 193–208.
[39] Y.I. Kim, D.W. Kim, D. Lee, K.H. Lee, A cluster validation index for GK cluster analysis based on relative degree of sharing, Information Sciences 168 (1-4) (2004) 225–242.
[40] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671–680.
[41] R.W. Klein, R.C. Dubes, Experiments in projection and clustering by simulated annealing, Pattern Recognition 22 (2) (1989) 213–220.
[42] G.J. Klir, T. Folger, Fuzzy Sets, Uncertainty and Information, Prentice-Hall, Englewood, Cliffs, 1988.
[43] L.G. Leu, H.W. Chang, Remotely sensing in detecting the water depths and bed load of shallow waters and their changes, Ocean Engineering 32 (10) (2005) 1174–1198.
[44] S.Z. Li, Markov Random Field Modeling in Image Analysis, Springer-Verlag, Japan, 2001.
[45] F.D. Martino, V. Loia, I. Perfilieva, S. Sessa, An image coding/decoding method based on direct and inverse fuzzy transforms, International Journal of Approximate Reasoning 48 (1) (2008) 110–131.

[46] F. Melgani, G. Moser, S.B. Serpico, Unsupervised change-detection methods for remote-sensing data, Optical Engineering 41 (2002) 3288–3297.
[47] K.R. Merril, L. Jiajun, A comparison of four algorithms for change detection in an urban environment, Remote Sensing of Environment 63 (2) (1998) 95–100.
[48] R. Min, H.D. Cheng, Effective image retrieval using dominant color descriptor and fuzzy support vector machine, Pattern Recognition 42 (1) (2009) 147–157.
[49] M. Mitchell, An introduction to genetic algorithms, Prentice-Hall of India, New Delhi, 2005.
[50] I. Ozkan, I.B. Türkşen, N. Canpolat, A currency crisis and its perception with fuzzy *c*-means, Information Sciences 178 (8) (2008) 1923–1934.
[51] N.R. Pal, J.C. Bezdek, On cluster validity for the fuzzy *c*-means model, IEEE Transactions on Fuzzy Systems 3 (3) (1995) 370–379.
[52] S.K. pal, A. Ghosh, Fuzzy geometry in image analysis, Fuzzy Sets and Systems 48 (2) (1992) 23–40.
[53] S.K. Pal, D. Dutta Majumder, Fuzzy Mathematical Approach to Pattern Recognition, John Wiley, New York, 1986.
[54] S. Patra, S. Ghosh, A. Ghosh, Change detection of remote sensing images with semi-supervised multilayer perceptron, Fundamenta Informaticae 84 (3-4) (2008) 429–442.
[55] W. Pedrycz, Fuzzy sets in pattern recognition: methodology and methods, Pattern Recognition 23 (1/2) (1990) 121–146.
[56] W. Pedrycz, F. Gomide, Fuzzy Systems Engineering: Toward Human-Centric Computing, Wiley-IEEE, 2007.
[57] J.M.S. Prewitt, Object enhancement and extraction, in: B.S. Lipkin, A. Rosenfeld (Eds.), Picture Processing and Psycho-Pictorics, Academic Press, New York, 1970.
[58] J.A. Richards, X. Jia, Remote Sensing Digital Image Analysis, 4th ed., Springer-Verlag, Berlin, 2006.
[59] A. Rosenfeld, Fuzzy geometry of image subsets, Pattern Recognition Letters 2 (1984) 311–317.
[60] R.H. Sheikh, M.M. Raghuwanshi, A.N. Jaiswal, Genetic algorithm based clustering: A survey, in: ICETET '08: Proceedings of the 2008 First International Conference on Emerging Trends in Engineering and Technology, 2008, pp. 314–319.
[61] A. Singh, Digital change detection techniques using remotely sensed data, International Journal of Remote Sensing 10 (6) (1989) 989–1003.
[62] S. Theodoridis, K. Koutroumbas, Pattern Recognition, Elsevier Academic Press, USA, 2008.
[63] H. Tong, M. Zhao, X. Li, Applications of computational intelligence in remote sensing image analysis, in: ISICA '09: Proceedings of the Fourth International Symposium on Intelligence Computation and Applications, Springer, 2009, pp. 171–179.
[64] J.R.G. Townshend, C.O. Justice, Spatial variability of images and the monitoring of changes in the normalized difference vegetation index, International Journal of Remote Sensing 16 (12) (1995) 2187–2195.
[65] J.R.G. Townshend, C.O. Justice, C. Gurney, J. McManus, The impact of misregistration on change detection, IEEE Transactions on Geoscience and Remote Sensing 30 (1992) 1054–1060.
[66] B. Tso, P.M. Mather, Classification Methods for Remotely Sensed Data, Taylor & Francis, London, 2001.
[67] P.J.M. van Laarhoven, E.H.L. Aarts, Simulated Annealing: Theory and Applications, Kluwer Academic Publisher, 1986.
[68] H. Wang, B. Fei, A modified fuzzy *c*-means classification method using a multiscale diffusion filtering scheme, Medical Image Analysis 13 (2) (2009) 193–202.
[69] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (8) (1991) 841–847.
[70] L.A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–353.