Applied Soft Computing 12 (2012) 2683–2692

Fuzzy clustering algorithms incorporating local information for change detection in remotely sensed images

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

[a] Department of Electronics and Communication Engineering, Netaji Subhash Engineering College, Kolkata 700152, India
[b] Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India
[c] Center for Soft Computing Research, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India

ARTICLE INFO

Article history:
Received 13 April 2011
Received in revised form 23 January 2012
Accepted 5 March 2012
Available online 24 April 2012

Keywords:
Remote sensing
Change detection
Multitemporal images
Local information
Fuzzy *c*-means clustering
Gustafson–Kessel clustering
Genetic algorithms
Simulated annealing
Xie–Beni and fuzzy hypervolume validity measures

ABSTRACT

In this paper we have used two fuzzy clustering algorithms, namely fuzzy *c*-means (FCM) and Gustafson–Kessel clustering (GKC) along with local information for unsupervised change detection in multitemporal remote sensing images. In conventional FCM and GKC no spatio-contextual information is taken into account and thus the result is not so much robust to small changes. Since the pixels are highly correlated with their neighbors in image space (spatial domain), incorporation of local information enhances the performance of the algorithms. In this work we have introduced a new technique for incorporation of local information. Change detection maps are obtained by separating the pixel-patterns of the difference image into two groups. 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. Two fuzzy cluster validity measures (Xie–Beni and fuzzy hypervolume) have been used to quantitatively evaluate the performance. Results are compared with those of existing state of the art 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 a process aimed at identifying the differences in the state of a land cover by analyzing a pair of images acquired on the same geographical area at different times [1,2]. Such a problem plays an important role in different domains like studies on land use/land cover dynamic [3], monitoring shifting cultivations, burned area identification [4], analysis of deforestation processes [5,6], assessment of vegetation changes [7], monitoring of urban growth [8] and oceanography [9]. 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 (effort) required by manual image analysis.

Change detection in remotely sensed data may be done either in supervised or in unsupervised manner [4,5,7,8,10–16]. 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 partition the data into two groups *changed* and *unchanged*.

Before performing *change detection* between two multitemporal images, a certain degree of (pre)processing is needed [11] because of co-registration error [11,17,18], radiometric and geometric errors [19]. Thus, in literature [1,2], 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 [11,19].

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) [2], (b) multiple spectral bands (i.e. Change Vector Analysis – CVA) [2], (c) vegetation indices (Vegetation Index

* Corresponding author. Tel.: +91 33 2575 3110/3100; fax: +91 33 2578 3357.
E-mail addresses: niladrimishra@gmail.com (N.S. Mishra), susmitaghoshju@gmail.com (S. Ghosh), ash@isical.ac.in (A. Ghosh).

Differencing – VID) [2,20], etc. Tasseled Cap Transformation [21] is also a popular method. The most popular of these is the CVA and is 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* (step (iii)) is done on the *DI*. Either context-insensitive or context-sensitive procedure is adopted [4] for this. Histogram thresholding [16] 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 [10,15] 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) [22]) are required for estimating the class distributions assuming different standard distributions e.g. Gaussian, generalized Gaussian [10] and mixture of Gaussians. A few context-sensitive techniques using neural networks are also suggested recently [4,5,23,24].

Relevance of fuzzy set theoretic methods in pattern recognition and image analysis problems has adequately been addressed in the literature [25–35]. Fuzzy clustering incorporating local information 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 [26] approach. 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 techniques are more appropriate and realistic to separate overlapping clusters [19], we have chosen fuzzy clustering techniques to have a better judgement of the two groups. In this regard we have used two fuzzy clustering algorithms namely fuzzy *c*-means [26] and Gustafson–Kessel [36]. There are several fuzzy cluster validity indexes available in the literature to evaluate fuzzy clustering results. We have used two of them. The first one is proposed by Xie and Beni [37] and the second one is by Gath and Geva [38]. They consider both intra cluster compactness and inter cluster separation. While evaluating the outcome of GK-type clustering using Xie–Beni validity index we have used Mahalanobis norm as in [39].

In image clustering applications FCM or GKC can be treated differently from data clustering. Pixels are normally highly correlated to their neighbors in the image space. This should be exploited for more efficiency. Also the homogeneous and nonhomogeneous regions (in context with gray values of pixels) in one image do not bear the same information. So instinct suggests that the amount of local information should vary from zone to zone, and better if varied from pixel to pixel. A pre-computation is done to incorporate the local neighborhood information in a *variable* fashion to the pixels of the *DI*. After generating the patterns they are subjected to clustering for identifying their class labels (*changed* or *unchanged*). Local information is incorporated here in such a way that its amount can vary from pixel to pixel automatically depending on the degree of homogeneity of its surrounding pixels (over a fixed window). It makes this method more robust for small changes and experimental results show that this technique is very efficient than the existing ones.

Two well known other optimization techniques namely genetic algorithms (GAs) [40] and simulated annealing (SA) [41] have been used to minimize the objective functions of the above mentioned clustering techniques to yield better results.

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 [4] for solving the same problem of change detection on the same data sets. We also compared the result of hard clusterings to show the effectiveness of the fuzzy ones. The proposed techniques have an edge with respect to both error and time requirements.

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

2. Clustering

The clustering algorithms [42] (both fuzzy and non-fuzzy) used in the present investigation are described here in brief.

2.1. Hard c-means (HCM) clustering

The HCM [42] algorithm minimizes the following objective function to divide the data set into *c* clusters.

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

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 $\mathbf{x}_k$ and the ith cluster center $\mathbf{v}_i$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c]$.

2.2. Fuzzy clustering

2.2.1. Fuzzy c-means (FCM) clustering

FCM [26] attempts to find fuzzy partitioning of a given data set by minimizing the objective functional

$$J_m(\mathbf{X};U,\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{n} (\mu_{ik})^m D_{ik}, \tag{2}$$

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

$$\mathbf{v}_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^m \mathbf{x}_k}{\sum_{k=1}^{n} (\mu_{ik})^m} \tag{3}$$

with μ_{ik} (degree of belonging of pattern $\mathbf{x}_k$ to the ith cluster) is expressed as

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} (d_{ik}/d_{jk})^{2/(m-1)}}, \tag{4}$$

where $d_{ik} = \sqrt{D_{ik}}$ and m (>1) is a parameter, called fuzzifier, which controls the fuzziness of the patterns. 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) $\mu_{ik} \in [0,1]$.

2.2.2. Gustafson–Kessel clustering (GKC)

Gustafson and Kessel introduced [36] adaptive distance norm to measure the distance between clusters using fuzzy covariance matrix (F_i). Each cluster has its own norm-inducing matrix A_i, a positive definite symmetric one, for automatically adapting its shape. F_i for the ith cluster is expressed as

$$F_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^m (\mathbf{x}_k - \mathbf{v}_i)(\mathbf{x}_k - \mathbf{v}_i)^T}{\sum_{k=1}^{n} (\mu_{ik})^m}. \tag{5}$$

The distance d_{ikA_i} is computed as

$$d_{ikA_i} = \sqrt{(\mathbf{x}_k - \mathbf{v}_i)^T A_i (\mathbf{x}_k - \mathbf{v}_i)}, \quad (6)$$

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

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} (d_{ikA_i}/d_{jkA_i})^{2/(m-1)}}. \quad (7)$$

The objective function J_m will now be of the form

$$J_m(\mathbf{X}; U, \mathbf{V}, A) = \sum_{i=1}^{c}\sum_{k=1}^{n} (\mu_{ik})^m D_{ikA_i}, \quad (8)$$

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

3. Validity measures for fuzzy clustering

3.1. Xie–Beni validity measure

A popular index to validate the outcome of a fuzzy clustering proposed by Xie and Beni [37], known as Xie–Beni index, is used widely by many researchers and is described as

$$v_{XB} = \frac{\sum_{i=1}^{c}\sum_{k=1}^{n} (\mu_{ik})^2 ||\mathbf{x}_k - \mathbf{v}_i||^2}{n(\underbrace{min}_{i \neq j} ||\mathbf{v}_i - \mathbf{v}_j||^2)}. \quad (9)$$

For a good partitioning, the index stated here should be minimum. The index uses Euclidean norm in its numerator. So, we have used this to evaluate the outcome of FCM-type clusterings (FCM and RFCM, discussed later in Section 4.1) as FCM employs Euclidean norm for clustering purpose. For GK-type (GKC and RGKC, discussed later in Section 4.1) clustering 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.

3.2. Fuzzy hypervolume validity measure

Based on the hypervolume and density of clusters, Gath and Geva proposed the fuzzy hypervolume validity [38] measure. The index is defined as

$$v_{FHV} = \sum_{i=1}^{c} [determinant(F_i)]^{1/2}. \quad (10)$$

A fuzzy partition is expected to have a low v_{FHV} value if the partition is tight. We have used this to validate all FCM-type and GK-type clusterings for our application.

4. Proposed change detection technique

In our experiments we have used two data sets of the described format where every pixel is of a gray shade between 0 and 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 $DI = \{l_{(m,n)}, 1 \le m \le p, 1 \le n \le q\}$ be the difference image obtained by applying the CVA technique as follows:

$$P_{DI_{(m,n)}} = \sqrt{\sum_{i=1}^{num} (P_{X_{1(m,n)b_i}} - P_{X_{2(m,n)b_i}})^2}, \quad (11)$$

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 X_1 and 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.

To exploit the spatio-contextual information, the pixel-pattern values (at the time of clustering) in the difference image (DI) are modified by considering the influence of its immediate spatial neighborhood N^d of order d. For a given spatial position (m,n), $N^d(m,n)$ is defined as follows: $N^d(m,n) = \{(m,n) + (i,j), (i,j) \in N^d\}$. Fig. 1(a) and (b) depicts the second (N^2) and third (N^3) order neighborhood systems of a pixel at position (m,n).

4.1. Incorporation of neighborhood information

Incorporation of local information can be done in various ways and several methods are suggested to do so [43,44]. The present work is inspired by the method proposed by Cai et al. [44]. A local similarity measure (S_{kr}, described later) has been introduced, which is varied from pixel to pixel to incorporate the local statistics of the image. Using this measure the input image is converted to a linearly transformed image as follows:

$$\xi_k = \frac{\sum_{r \in N_k} S_{kr} x_r}{\sum_{r \in N_k} S_{kr}}, \quad (12)$$

where ξ_k is the processed gray value incorporating the local constraint corresponding to the input gray value of pixel x_k (window center).

Cai et al. have defined S_{kr} in such a way that it can adopt the variation in shape of window. For square window (of fixed size for all the pixels) we can ignore this constraint and can consider only what they have called as gray level relationship (i.e. the spatio-contextual relation) which takes care of intensity inhomogeneity within the local window. In this case S_{kr} can be expressed as

$$S_{kr} = exp\left[\frac{-||x_k - x_r||^2}{\lambda_g \times \sigma_{g\_k}^2}\right] \quad (13)$$

where σ_{g_k} as

$$\sigma_{g\_k} = \sqrt{\frac{\sum_{r \in N_k} ||x_r - x_k||^2}{N_R}}; \quad (14)$$

λ_g denotes the global scale factor of the spread of S_{kj}. Parameter σ_{g_k} is a function of the local density surrounding the central pixel and its value reflects the gray value homogeneity of local window [44].

After generating patterns using Eq. (12) (instead of using gray levels) we have clustered them by taking $c = 2$ by HCM, FCM, and GKC and named the processes as RHCM, RFCM and RGKC respectively, where 'R' denotes "Robust". The pixel corresponding to the position (m,n) of DI is thought as the center pixel of the 8 neighbors in N^2 i.e. neighbors those are within a distance of 2 from the center pixel of the system. We may work upon some higher order ($d > 2$) neighborhood system (like N^3) to do this 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 two clusters by a clustering algorithm, one cluster 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

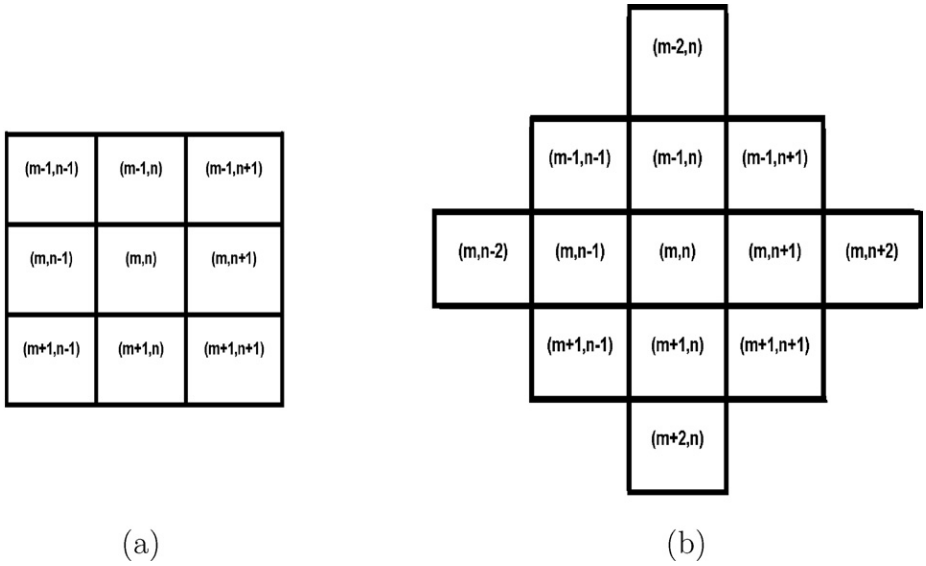

Fig. 1. Neighborhood of the pixel at position (m, n). (a) N^2 and (b) N^3.

corresponding to *changed* ones are marked as black (*gray level* 0) and the *unchanged* ones are marked as white (*gray level* 255) in the generated *change detection map*. For comparison purpose and to establish the effectiveness and applicability of the incorporation of the local neighborhood information, in the present context we have applied all the basic clustering models (HCM, FCM and GKC) on the pixels of *DI* treating them as patterns (i.e. detecting changes by clustering in a context-insensitive manner) and have called them as *context-insensitive clusterings*.

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. The used fuzzy clustering algorithms show better performance for some values of the parameters (m for FCM-type, m and ρ_i for GK-type and λ_g for the robust clusterings). While combining them with GA or SA, we have considered those values of the parameters only. This attempt is to study how GA or SA can enhance the performances of the fuzzy clustering algorithms under the same environment to detect changes for remotely sensed data. While combining HCM with GA we call it G_HCM (Genetic HCM) and that with SA as SA_HCM (Simulated Annealing HCM) in the rest of the work. 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 GK-type we only worked upon combining with SA and not with GA. Section 6 will highlight the practical difficulties behind it. While combining our proposed methods (context-sensitive clustering models) with the optimization methods, the hybridization is named in a similar way as in the case of context-insensitive clusterings with the optimization methods (e.g. RHCM with GA will be called as G_RHCM or with SA as SA_RHCM and so on).

When dealing with clusterings combined with GA the chromosomes are binary representations of the two cluster centers (*changed* and *unchanged*) in some specific order. Our remotely sensed images are of 8-bit. So, when dealing with context-insensitive clusterings a center is represented by 8 bits and thus a chromosome is of 16 bits. When working with our proposed model the same is a real number. 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 22 bits (11 bits representing a cluster center). As better clustering promises the corresponding functional to be minimized, we have chosen the fitness function as the functional itself.

Similarly, using SA a set of (initially randomly chosen) good cluster representatives (means for HCM-type and fuzzy means for FCM-type and GK-type) is determined such that the corresponding clusters become compact. The set of cluster centers represents the *configuration* [41] and the corresponding value of the associated objective functional (of the concerned clustering model) is the *cost* [41] at that moment. To perturb the present configuration 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.

5. 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 m. Each pixel thus represents an area of 30 m × 30 m. A detailed description of each of the data sets is given below.

5.1. Data set related to Mexico area

The first data set used in the experiments 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 × 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.

Fig. 2(a) and (b) shows 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 236,545 unchanged pixels. Experiments were carried out to produce, in an automatic way, a change detection map as similar as possible to the reference map

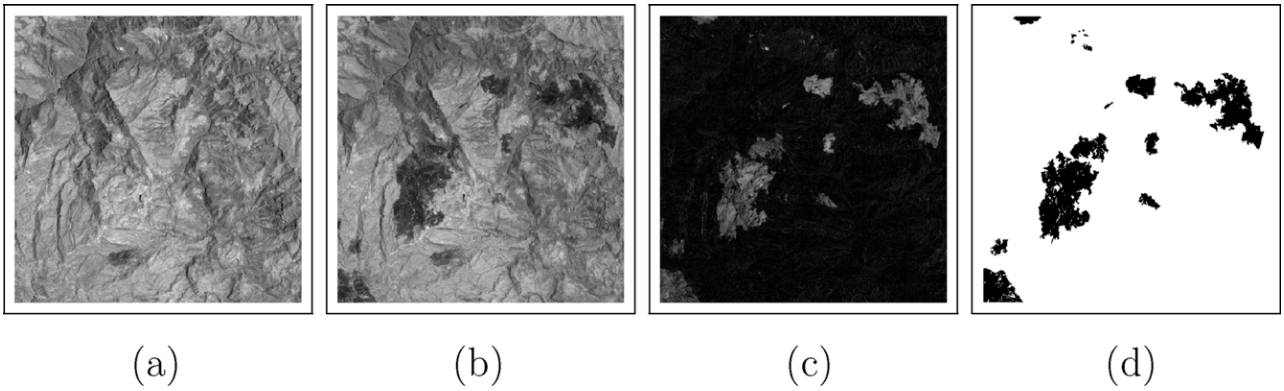

Fig. 2. Image of Mexico area. (a) Band 4 of the Landsat ETM+image acquired in April 2000, (b) band 4 of the Landsat ETM+image acquired in May 2002, (c) corresponding difference image generated by CVA technique, and (d) reference map of the changed area.

that represents the best result obtainable with a time consuming procedure.

5.2. 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 412×300 pixels of a scene including lake Mulargia on the Island of Sardinia (Italy).

Between the two aforementioned acquisition dates, the water level in the lake increased (see the lower central part of the image). Fig. 3(a) and (b) shows 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 116,120 unchanged pixels were identified.

6. Experiments and results

To assess the effectiveness of the proposed approach, we have made both qualitative (visual) and quantitative analyses of the experimental results. In visual analysis we have compared *change detection map* (a binary image) with the ground truth image (a binary one). We then presented a quantitative analysis with respect to *overall error*. We have made a comparative study of the performances of our proposed algorithms with two context-insensitive techniques and two context-sensitive techniques. The first context-insensitive technique is MTET [10]. It produces a minimum error change detection map by finding an optimal decision threshold for *DI*. The second one is *context-insensitive clusterings* as stated previously (refer Section 4). Both the context-insensitive techniques assume the pixels to be independent in spatial domain. Context-sensitive techniques compared with are: (i) the technique presented in [10] where EM is combined with MRF and will be referred as EM + MRF and (ii) a technique based on "Hopfield Type Neural Networks" [4] (will be referred as HTNN). In [4] four different Hopfield-type network models were used. We have compared with 2nd order continuous model as we also have used 2nd order neighborhood information in the present work.

As already discussed, our process considers the local homogeneity around the pixels while incorporating local information. In literature [44] it has been reported that this type of local-information-incorporation is more "robust" to noise and outliers than others [43]. In [43], modification of the objective function of the standard FCM is done. The re-definition of the objective function is such that incorporation of local information from its neighborhood region (within a fixed window) is *constant* for the whole image space. We have tested this concept for hard as well as for fuzzy clustering (both for FCM and GK-type) for our application also. It has been seen that this type is not fruitful for the present purpose (i.e. change detection) and so is not reported here.

(a)–(c) and those for Sardinia data set are shown in Fig. 6(a)–(c) respectively. The same by RHCM, RFCM and RGKC for Mexico data set are shown in Fig. 5

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 areas. For Sardinia data set, the *DI* is generated by the CVA algorithm using spectral bands 1, 2, 4 and 5. As mentioned earlier, for all the fuzzy clustering techniques the value of fuzzifier (m) affects the results. For GK-type clusterings 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-type clustering (FCM and RFCM) and stopped at that point where the performance of the algorithm started degrading. In case of GK-type (GKC and RGKC), maintaining the same criterion for varying m, we have fixed one

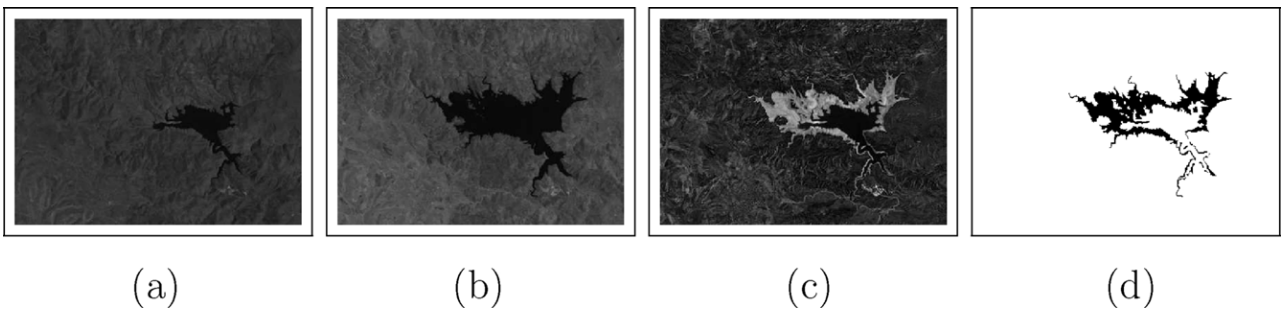

Fig. 3. Image of Mexico area. (a) Band 4 of the Landsat ETM+image acquired in April 2000, (b) band 4 of the Landsat ETM+image acquired in May 2002, (c) corresponding difference image generated by CVA technique, and (d) reference map of the changed area.

ρ_i to 1 and varied the other from 1 taking an incremental step size of 0.1 and vice versa up to that extent where we have achieved the best result. Regarding the local-information sensitive processes (proposed method) we have set the λ_g also showing best results. For all the experiments, ϵ is set to 0.0000001.

While working with GA, a population size of 30 was taken. We have used 2-fold tournament selection [40] 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 [45]. *Crossover probability* and *mutation probability* are set to 0.9 and 0.01 respectively. Elitist strategy [40] 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.

In SA only one candidate solution is considered for optimization purpose, unlike GA. 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 ($T \to \infty$), a slow reduction of the temperature and a very low ($\to 0$) final temperature. This is the sufficient condition for this technique to reach to a global optimum [41,46]. To overcome the computational burden of SA 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 number of transitions [41] to set the initial temperature. To cool the system a *linear cooling scheme* [41] is adopted. It has been reported in literature [46] that SA is very much sensitive to the random numbers generated 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 initial temperature T_0) during iterations [41] 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 (and minimum T_0 as 10) if χ_0 falls below 0.8 we doubled the initial temperature T_0 to start the process.

Now we will explore the reason why it was not possible to hybridize GKC and RGKC 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 and SA_RGKC, 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 cannot 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 the corresponding fuzzy means ($\mathbf{v}_i$) (can be verified from Eqs. (5) and (6)) and so the new μ_{ik}. Also we cannot 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.

6.1. Visual analysis

The *change detection maps* obtained for Mexico data set by the HCM, FCM and GKC are shown in Fig. 4(a)–(c) and those for Sardinia data set are shown in Fig. 7(a)–(c) respectively.

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

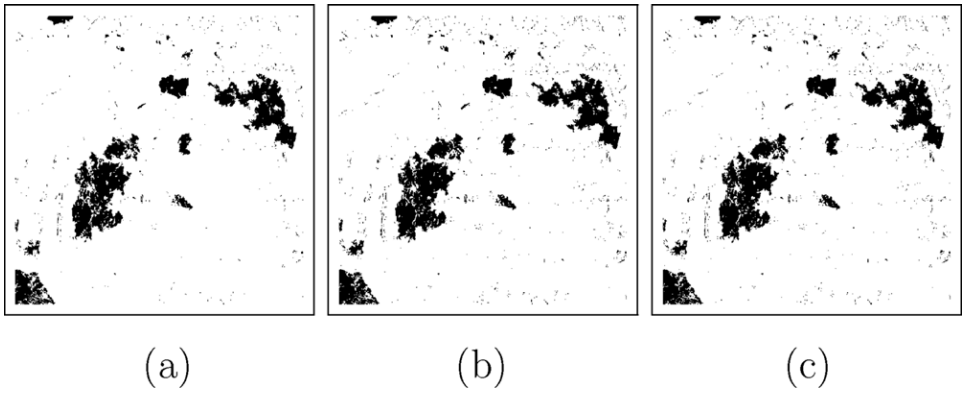

Fig. 4. Change detection maps obtained for Mexico data set by (a) HCM, (b) FCM ($m = 5.7$) and (c) GKC ($m = 3$ and $\rho_1 = 1$ and $\rho_2 = 2$).

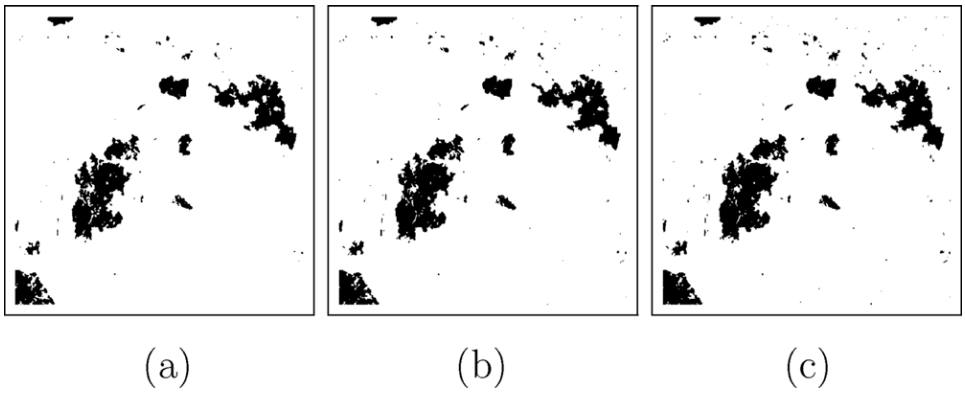

Fig. 5. Change detection maps obtained for Mexico data set by (a) RHCM ($\lambda_g = 6$), (b) RFCM ($m = 13.5$ and $\lambda_g = 6$) and (c) RGKC ($m = 10.5$, $\lambda_g = 6$ and $\rho_1 = 1$ and $\rho_2 = 3.4$).

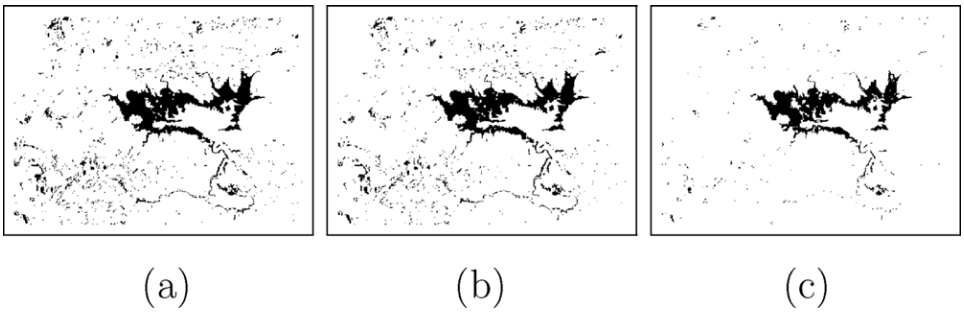

Fig. 6. Change detection maps obtained for Sardinia data set by (a) HCM, (b) FCM ($m = 1.1$) and (c) GKC ($m = 2$ and $\rho_1 = 3.5$ and $\rho_2 = 1$).

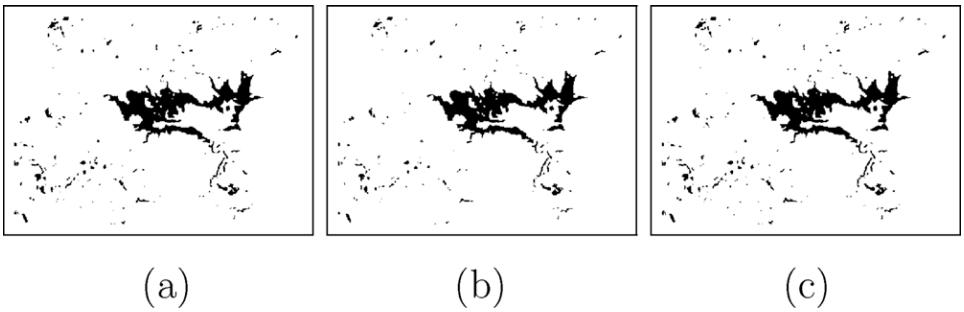

Fig. 7. Change detection maps obtained for Sardinia data set by (a)RHCM ($\lambda_g = 6$), (b) RFCM ($m = 1.2$ and $\lambda_g = 6$) and (c) RGKC ($m = 3.1$, $\lambda_g = 6$ and $\rho_1 = 3.5$ and $\rho_2 = 1$).

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 Section 6.2) generated at larger portions by the FCM. But careful inspection reveals that the extreme bottom-left corner of the scene is detected much better by FCM, where HCM failed. The same inference can be drawn for this data set while comparing RHCM and RFCM. It is very difficult to say visually whether both FCM and GKC yielded the same result or not. The inference is the same while comparing RFCM and RGKC. For Sardinia data set visual inspection of the *change detection map* shows that one can hardly find any difference between the performance of HCM and FCM. But it is clear that GKC is doing better than FCM. The observation is similar for our proposed method, RGKC is doing better than RFCM and similarly RFCM is doing better than RHCM for the same data set. This leads us to judge the effectiveness of our proposed technique quantitatively, which obviously is better than visual inspection and is presented in the following section.

6.2. Quantitative analysis

6.2.1. Quantitative evaluation using error measures

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 the basic clustering models (i.e. context-insensitive clusterings), and our proposed methods (RHCM, RFCM and RGKC) using the "ground truth" image as the reference map.

Results obtained by all the techniques (existing, context-insensitive clusterings and proposed methods) are shown in Tables 1 and 2 for a comparative and thorough discussion. The tables are divided into three subparts to represent the existing techniques (MTET, HTNN and EM + MRF), context-insensitive clusterings and our proposed methods, respectively.

It is seen from the results that the context-insensitive clusterings are not well-suited for our purpose. Though GKC/SA_GKC performed little better in case of our second data set (comparable with MTET and HTNN and worse than EM + MRF); but for the first one it showed inferior performance. Thus in general this type (context-insensitiveness) is not useful for the present problem.

Now let us concentrate on our proposed *change detection technique*. Let us have a look at the outcome of our proposed methods (the last subpart of both the tables). From Tables 1 and 2 it is seen that *crisp set based* method (RHCM) did not perform well for both the data sets. Also hybridization of the same with GA and SA did not show better performance (only SA_RHCM performed better than the existing for the first data set). On the contrary all the *fuzzy set based* methods (from RFCM to SA_RGKC) are found to be superior than other techniques.

It is worth noting that with local information GK-type (RGKC) showed its superiority over FCM-type (RFCM). On the contrary without local information GK-type (GKC) could not provide reasonable performance. This suggests that the proposed method of

Table 1
Missed alarms, false alarms and overall error for Mexico data set.

Techniques used	*MA*	*FA*	*OE*
MTET	2404	2187	4591
HTNN	558	2707	3265
EM + MRF ($\beta = 1.5$)	946	2257	3203
HCM	3860	1251	5111
G_HCM	2868	1808	4676
SA_HCM	2868	1808	4676
FCM ($m = 5.7$)	2404	2186	4590
G_FCM ($m = 5.7$)	2404	2186	4590
SA_FCM ($m = 5.7$)	2404	2186	4590
GKC ($m = 3$, $\rho_1 = 1$, $\rho_2 = 2$)	2404	2186	4590
SA_GKC ($m = 3$, $\rho_1 = 1$, $\rho_2 = 2$)	2404	2186	4590
RHCM ($\lambda_g = 6$)	3198	665	3863
G_RHCM ($\lambda_g = 6$)	2283	873	3156
SA_RHCM ($\lambda_g = 6$)	685	2246	2931
RFCM ($m = 13.5$, $\lambda_g = 6$)	1742	1104	2846
G_RFCM ($m = 13.5$, $\lambda_g = 6$)	1476	1271	2747
SA_RFCM ($m = 13.5$, $\lambda_g = 6$)	1075	1557	2632
RGKC ($m = 10.5$, $\rho_1 = 1$, $\rho_2 = 3.4$, $\lambda_g = 6$)	1095	1531	2626
SA_RGKC ($m = 10.5$, $\rho_1 = 1$, $\rho_2 = 3.4$, $\lambda_g = 6$)	1099	1520	2619

Table 2
Missed alarms, false alarms and overall error for Sardinia data set.

Techniques used	*MA*	*FA*	*OE*
MTET	1015	875	1890
HTNN	1187	722	1909
EM + MRF ($\beta = 2.2$)	592	1108	1700
HCM	304	3877	4181
G_HCM	304	3877	4181
SA_HCM	458	2593	3051
FCM ($m = 1.1$)	382	3143	3525
G_FCM ($m = 1.1$)	382	3143	3525
SA_FCM ($m = 1.1$)	545	2100	2645
GKC ($m = 2$, $\rho_1 = 3.5$, $\rho_2 = 1$)	1015	874	1889
SA_GKC ($m = 2$, $\rho_1 = 3.5$, $\rho_2 = 1$)	1015	874	1889
RHCM ($\lambda_g = 6$)	530	1973	2503
G_RHCM ($\lambda_g = 6$)	1993	161	2154
SA_RHCM ($\lambda_g = 6$)	634	1503	2137
RFCM ($m = 1.2$, $\lambda_g = 6$)	659	1414	2073
G_RFCM ($m = 1.2$, $\lambda_g = 6$)	659	1410	2069
SA_RFCM ($m = 1.2$, $\lambda_g = 6$)	881	790	1671
RGKC ($m = 3.1$, $\rho_1 = 3.5$, $\rho_2 = 1$, $\lambda_g = 6$)	1089	489	1578
SA_RGKC ($m = 3.1$, $\rho_1 = 3.5$, $\rho_2 = 1$,$\lambda_g = 6$)	1096	474	1570

incorporating local information will make GK-type more powerful than others.

6.2.2. Evaluation in terms of fuzzy cluster validity index

In Section 6.2.1, an analysis was carried in terms of *OE*. It was conveyed that less the value of *OE*, better is the technique. Also we have shown that incorporation of local information makes the proposed *fuzzy set based* techniques more effective than the pre-suggested techniques. Here we will try to validate the proposed *fuzzy set based* techniques in terms of *validity indexes* as stated in Section 3. To do so we calculated validity measures proposed by Xie–Beni (v_{XB} for FCM-type and v_{XBe} for GK-type respectively) and that by Gath–Geva (v_{FHV}) and put the results along with the corresponding *OE* in Tables 3–6.

One can see the values of v_{XB} with corresponding *OE* for all the processes (from FCM to SA_RFCM) from Tables 3 and 4. From both the tables it is observed that less *OE* ensures lesser or constant

Table 3
Missed alarms, false alarms and overall error for Mexico data set by FCM-type clustering.

Techniques used	*OE*	v_{XB}	v_{FHV}
FCM ($m = 5.7$)	4590	0.09	28.11
G_FCM ($m = 5.7$)	4590	0.09	28.11
SA_FCM ($m = 5.7$)	4590	0.09	28.11
RFCM ($m = 13.5$, $\lambda_g = 6$)	2846	0.07	21.22
G_RFCM ($m = 13.5$, $\lambda_g = 6$)	2747	0.07	21.22
SA_RFCM ($m = 13.5$, $\lambda_g = 6$)	2632	0.07	21.22

Table 4
Missed alarms, false alarms and overall error for Sardinia data set by FCM-type clustering.

Techniques used	*OE*	v_{XB}	v_{FHV}
FCM ($m = 1.1$)	3525	0.11	43.20
G_FCM ($m = 1.1$)	3525	0.11	43.20
SA_FCM ($m = 1.1$)	2645	0.10	42.97
RFCM ($m = 1.2$, $\lambda_g = 6$)	2073	0.07	38.58
G_RFCM ($m = 1.2$, $\lambda_g = 6$)	2069	0.07	38.57
SA_RFCM ($m = 1.2$, $\lambda_g = 6$)	1671	0.07	38.14

Table 5
Missed alarms, false alarms and overall error for Mexico data set by GK-type clustering.

Techniques used	*OE*	v_{XBe}	v_{FHV}
GKC ($m = 3$, $\rho_1 = 1$, $\rho_2 = 2$)	4590	0.044411	28.11
SA_GKC ($m = 3$, $\rho_1 = 1$, $\rho_2 = 2$)	4590	0.044411	28.11
RGKC ($m = 10.5$, $\rho_1 = 1$, $\rho_2 = 3.4$, $\lambda_g = 6$)	2626	0.034798	18.49
SA_RGKC ($m = 10.5$, $\rho_1 = 1$, $\rho_2 = 3.4$, $\lambda_g = 6$)	2619	0.034619	18.15

Table 6
Missed alarms, false alarms and overall error for Sardinia data set by GK-type clustering.

Techniques used	*OE*	v_{XBe}	v_{FHV}
GKC ($m = 2$, $\rho_1 = 3.5$, $\rho_2 = 1$)	1889	0.041605	38.08
SA_GKC ($m = 2$, $\rho_1 = 3.5$, $\rho_2 = 1$)	1889	0.041605	38.03
RGKC ($m = 3.1$, $\rho_1 = 3.5$, $\rho_2 = 1$, $\lambda_g = 6$)	1578	0.032058	35.03
SA_RGKC ($m = 3.1$, $\rho_1 = 3.5$, $\rho_2 = 1$, $\lambda_g = 6$)	1570	0.031982	34.95

v_{XB} (i.e. the index did not increase while *OE* decreased). This is a good signature. From Tables 5 and 6 one can notice that movement from GKC to SA_RGKC, v_{XBe} value always decreased for less *OE*. Thus v_{XBe} showed better results for our purpose to extract the clusters by assuming them as elliptical as this index employs Mahalanobis norm.

Similar findings can be corroborated from the response of v_{FHV} also (see Tables 3–6). For GK-type the index shows lesser value than FCM-type and for SA_RGKC it shows the least value.

From an overall analysis it is felt that while solving the change detection problem, if we incorporate the knowledge from neighborhood of the pixels with fuzzy clustering the output is improved (than the pre-suggested techniques and crisp clustering). The existing techniques require either the assumption of distributions of classes and are very time consuming (EM + MRF) or need 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. Since GK-type uses Mahalanobis distance and can extract even non-convex clusters, it produces better results for the used data sets (since the data mainly contain irregular shaped *changed* regions). The value of *m* used for Mexico data is large whereas for Sardinia data set it is less. 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. Again by varying ρ_i (instead of fixing to 1) in GK-type, 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 parameter to 1 and varying the other) the assumed shapes of the clusters could be varied. This is one of the major factors for GK-type to work well for this problem. This suggests that incorporation of the knowledge about the structures of the *changed* and *unchanged* classes may be helpful while solving *change detection* problems. It has been seen that the proposed technique works well within a reasonable range of values of the parameters (m, ρ_1, ρ_2 and λ_g).

6.3. Effect of parameters

From the above analysis it is evident that the "robust" versions of fuzzy clusterings can be more useful for change detection of remote sensing images. But it is seen experimentally that the performances

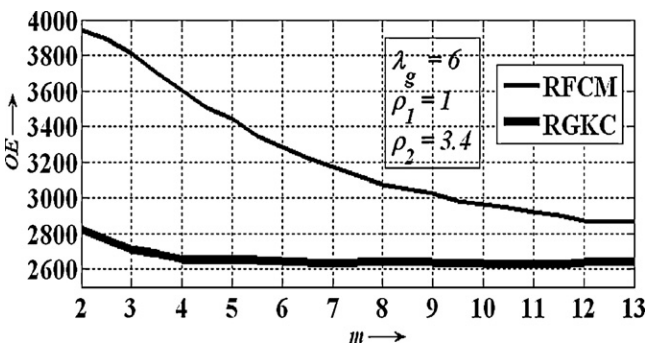

Fig. 8. Plot of *OE* as a function of *m*.

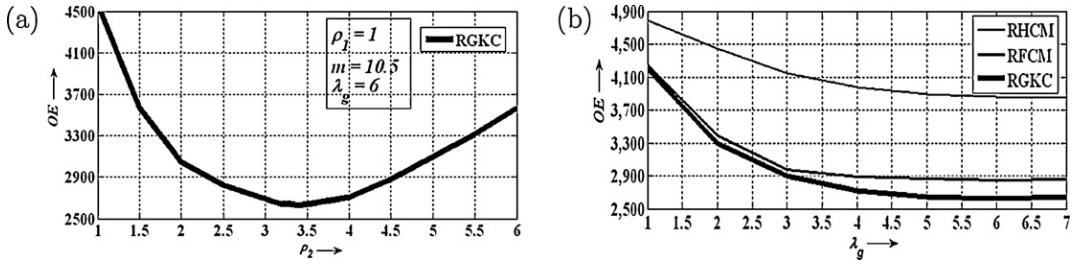

Fig. 9. Plot of *OE* as a function of (a) ρ_2 and (b) λ_g.

of them are highly influenced by the *parameters* which include the fuzzifier (m), the variable to control the clusters' shapes (ρ_i) and the variable-local-information incorporation parameter (λ_g). For a certain set of these parameter values the performance of a process may get enhanced. How one can choose the parameter-values? This section is to deal with the issue of finding out some general solution of this query.

To analyze the effect of parameters in a better way we have plotted *OE* against m in Fig. 8 for both RFCM and RGKC for Mexico data set. For RGKC the plot is for constant ρ_i (taking $\rho_1 = 1$ and $\rho_2 = 3.4$). Fig. 9(a) shows the plot of *OE* against ρ_2 for constant ρ_1 (=1) and m (=10.5) for the same data set. Finally we have plotted the *OE* against λ_g for RHCM, RFCM and RGKC taking the other parameter values (m for RFCM and RGKC and ρ_i for RGKC) constant in Fig. 9(b). We have tested our technique taking the parameter values also outside the ranges shown. The graphs are shown for a reasonable range for which *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 RFCM we can say that when m is increased beyond 2, *OE* falls exponentially. Up to $m = 12$ *OE* reduces at a high rate and from that point the falling rate is reduced. As seen form the figure, variations of *OE* is negligible for m in the range 12–13; 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. A similar behavior can be observed for RGKC also. One can see that if proper values of ρ_i are chosen then RGKC performs in a well manner. Unlike RFCM, here we can notice that the linear region is more. From $m = 2$ to 4, *OE* falls slowly and then the performance is sustained throughout the range shown i.e. *OE* is almost constant (of the order of 2650).

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

From the above analysis it is obvious that RGKC is performing the best for our data sets. The main reason behind this is we can vary the relative shapes of the clusters (corresponding to the changed and unchanged regions) by tuning the parameter ρ_i. It is a vital issue for this process to perform well. Thus ρ_i should not be chosen randomly. Let us analyze the behavior of this parameter for Mexico data set. Fig. 9(a) tells us about the effect of ρ_i for some constant value of m and λ_g. We have set $m = 10.5$ and $\lambda_g = 6$ as these values 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 rapidly. In the interval 3.2–3.5 we have a stable region (as *OE* is constant, 2650). Then if ρ_2 is further increased, *OE* increases slowly. So, the choices of this parameter should be from this linear region and one can accept any value of ρ_2 in this range to have a reasonable performance.

Fig. 9(b) shows the effect of λ_g after fixing the other parameters at those values where they were giving the best (e.g. for RFCM m at 13.5 and for RGKC m at 10.5, ρ_1 at 1 and ρ_2 at 3.4). It is worth noting that the nature of all the graphs is similar; which means all the "robust" techniques are having a resemblance in their performances.

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

It was observed experimentally that for the context-insensitive clusterings (and also for constant-local-information incorporation [43]) it is very hard to find out the linear regions of the parameters. In Tables 1 and 2 we have put the best results for our proposed method by tuning the parameter values.

7. Conclusion and discussion

Unsupervised context-sensitive techniques using fuzzy clustering and local information 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 a more appropriate and realistic choice to separate them. Incorporation of local information further enhances the performance of the algorithms and makes the algorithms "robust". Among the fuzzy clustering algorithms, only fuzzy c-means and Gustafson–Kessel (GK) are used in the present experiment. The fuzzy clustering techniques are combined with genetic algorithm and simulated annealing to yield better results. As GK-type clustering can extract clusters with different (including non-spherical) shapes, it is found to be more effective. Two fuzzy cluster validity indexes namely Xie–Beni and fuzzy hypervolume have been used to validate the results of the change detection problem.

The proposed fuzzy clustering technique has advantages over the context-sensitive process (EM + MRF) presented in [10] as they are distribution free (do 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 [4], 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 clusterings are well suited for change detection of remotely sensed data, there are some unavoidable problems also – proper selection of the values of fuzzifiers m, ρ_i and (for the robust versions) λ_g. 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 wish to 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] J.A. Richards, X. Jia, Remote Sensing Digital Image Analysis, 4th ed., Springer-Verlag, Berlin, 2006.
[2] A. Singh, Digital change detection techniques using remotely sensed data, International Journal of Remote Sensing 10 (6) (1989) 989–1003.
[3] J. Cihlar, T.J. Pultz, A.L. Gray, Change detection with synthetic aperture radar, International Journal of Remote Sensing 13 (3) (1992) 401–414.
[4] 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.
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] M.J. Canty, Image Analysis, Classification and Change Detection in Remote Sensing, CRC Press, Taylor & Francis, 2006.
[12] 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.
[13] 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.
[14] 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 (2005) 352–359, ISEIS Publication Series Number P002.
[15] 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.
[16] F. Melgani, G. Moser, S.B. Serpico, Unsupervised change-detection methods for remote-sensing data, Optical Engineering 41 (2002) 3288–3297.
[17] 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.
[18] 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 (5) (1992) 1054–1060.
[19] B. Tso, P.M. Mather, Classification Methods for Remotely Sensed Data, Taylor & Francis, London, 2001.
[20] 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.
[21] T. Fung, An assessment of TM imagery for land-cover change detection, IEEE Transactions on Geoscience and Remote Sensing 28 (4) (1990) 681–684.
[22] 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.
[23] 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.
[24] 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.
[25] E. Barrenechea, H. Bustince, B. De Baets, C. Lopez-Molina, Construction of interval-valued fuzzy relations with application to the generation of fuzzy edge images, IEEE Transactions on Fuzzy Systems 19 (5) (2011) 819–830.
[26] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function., Plenum Press, New York, 1981.
[27] B. Caldairou, N. Passat, P.A. Habas, C. Studholme, F. Rousseau, A non-local fuzzy segmentation method: application to brain MRI, Pattern Recognition 44 (9) (2011) 1916–1927.
[28] P.J. Herrera, G. Pajares, M. Guijarro, A segmentation method using Otsu and fuzzy *k*-means for stereovision matching in hemispherical images from forest environments, Applied Soft Computing 11 (8) (2011) 4738–4747.
[29] Y.-C. Hu, Pattern classification by multi-layer perceptron using fuzzy integral-based activation function, Applied Soft Computing 10 (3) (2010) 813–819.
[30] S. Sadi-Nezhad, K.K. Damghani, Application of a fuzzy TOPSIS method base on modified preference ratio and fuzzy distance measurement in assessment of traffic police centers performance, Applied Soft Computing 10 (4) (2010) 1028–1039.
[31] 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.
[32] Z. Yang, F.-L. Chung, W. Shitong, Robust fuzzy clustering-based image segmentation, Applied Soft Computing 9 (1) (2009) 80–84.
[33] G.J. Klir, T. Folger, Fuzzy Sets, Uncertainty and Information, Prentice-Hall, Englewood Cliffs, 1988.
[34] A. Ghosh, S.K. Meher, B. Uma Shankar, A novel fuzzy classifier based on product aggregation operator, Pattern Recognition 41 (3) (2008) 961–971.
[35] A. Ghosh, B. Uma Shankar, S.K. Meher, A novel approach to neuro-fuzzy classification, Neural Networks 22 (1) (2009) 100–109.
[36] 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.
[37] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (8) (1991) 841–847.
[38] I. Gath, A.B. Geva, Unsupervised optimal fuzzy clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (7) (1989) 773–781.
[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] D.E. Goldberbg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, 1989.
[41] P.J.M. van Laarhoven, E.H.L. Aarts, Simulated Annealing: Theory and Applications, Kluwer Academic Publisher, 1986.
[42] S. Theodoridis, K. Koutroumbas, Pattern Recognition, Elsevier Academic Press, USA, 2008.
[43] M.N. Ahmed, S.M. Yamany, N. Mohamed, A.A. Farag, T. Moriarty, A modified fuzzy *c*-means algorithm for bias field estimation and segmentation of MRI data, IEEE Transactions on Medical Imaging 21 (3) (2002) 193–199.
[44] W. Cai, S. Chen, D. Zhang, Fast and robust fuzzy *c*-means clustering algorithms incorporating local information for image segmentation, Pattern Recognition 40 (3) (2007) 825–838.
[45] L.O. Hall, J.C. Bezdek, S. Boggavarpu, 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.
[46] R.W. Klein, R.C. Dubes, Experiments in projection and clustering by simulated annealing, Pattern Recognition 22 (2) (1989) 213–220.