Change detection in remotely sensed images using semi-supervised clustering algorithms

Article *in* International Journal of Knowledge Engineering and Soft Data Paradigms · January 2013

DOI: 10.1504/IJKESDP.2013.058127

3 authors:

Dr. Moumita Roy
Indian Institute of Information Technology Guwahati

Susmita Ghosh
Jadavpur University

Ashish Ghosh
Indian Statistical Institute

Change detection in remotely sensed images using semi-supervised clustering algorithms

Moumita Roy and Susmita Ghosh

Department of Computer Science and Engineering,
Jadavpur University,
Kolkata 700032, India
E-mail: moumita2009.roy@gmail.com
E-mail: susmitaghoshju@gmail.com

Ashish Ghosh*

Center for Soft Computing Research,
Indian Statistical Institute,
203 B.T. Road, Kolkata 700108, India
E-mail: ash@isical.ac.in
*Corresponding author

Abstract: Scarcity of sufficient ground truth information is the primary bottleneck for adopting any supervised methodology in change detection domain and hence, unsupervised approaches are mostly used for this task. But, with a few labelled patterns in hand, semi-supervised methods can be chosen instead of unsupervised ones to utilise both the labelled and unlabelled patterns completely. Work on semi-supervised learning (both in the areas of clustering and classification) is now being explored. In this article, a detailed study has been made by applying some of the semi-supervised clustering techniques for change detection. In present investigation, five semi-supervised clustering techniques, namely COP-KMeans, seeded-KMeans, constrained-KMeans, semi-supervised-HMRF-KMeans and semi-supervised-kernel-KMeans algorithms are used. A comparative analysis has been made among these algorithms and standard K-Means algorithm, using two multi-temporal remotely sensed images and are also statistically validated using paired t-test. Experimental results conclude that constrained-KMeans for both the datasets is more applicable for change detection than COP-KMeans and seeded-KMeans. Semi-supervised-HMRF-KMeans and semi-supervised-kernel-KMeans algorithms are found not to be robust for all the datasets because these algorithms outperform constrained-KMeans in case of only one dataset.

Keywords: multi-temporal images; semi-supervised clustering; change detection.

Reference to this paper should be made as follows: Roy, M., Ghosh, S. and Ghosh, A. (2013) 'Change detection in remotely sensed images using semi-supervised clustering algorithms', *Int. J. Knowledge Engineering and Soft Data Paradigms*, Vol. 4, No. 2, pp.118–137.

Biographical notes: Moumita Roy received her BE in Computer Science and Engineering from University Institute of Technology, Burdwan University and ME in Computer Science and Engineering from Bengal Engineering and Science University, Shibpur in 2006 and 2009, respectively. She is currently pursuing her PhD in Computer Science and Engineering Department at Jadavpur University, Kolkata as a Senior Research Fellow under Council of Scientific and Industrial Research (CSIR), India. From 2006 to 2007, she also served as a Lecturer in Computer Science and Engineering Department of IMPS College of Engineering and Technology, Malda. She has selected for pursuing PhD Internship under Erasmus Mundus programme during the session 2013–1014 at University of Trento, Italy. Her research interests include pattern recognition, neural networks, and remote sensing image analysis.

Susmita Ghosh received her BSc (with honors) in Physics and BTech in Computer Science and Engineering from the University of Calcutta, India in 1988 and 1991, respectively. She received her MTech in Computer Science and Engineering from the Indian Institute of Technology, Bombay in 1993, and PhD (Engineering) from Jadavpur University, Kolkata, in 2004. She has visited various universities/academic institutes for collaborative research/international projects and delivered lectures in different countries including Australia, China, Italy, Japan and the USA. She has published many research papers in internationally reputed journals and refereed conferences and has edited three books. She has been a Lecturer and then a Reader with the Department of Computer Science and Engineering, Jadavpur University since 1997. She is currently an Associate Professor at the same department. Her research interests include genetic algorithms, neural networks, image processing, pattern recognition, soft computing, and remotely sensed image analysis.

Ashish Ghosh is Head and Professor of the Machine Intelligence Unit at Indian Statistical Institute, Kolkata, India. He received his BE in Electronics and Telecommunication from the Jadavpur University, Kolkata in 1987, and MTech and PhD in Computer Science from the Indian Statistical Institute, Kolkata in 1989 and 1993, respectively. His research interests include pattern recognition and machine learning, remotely sensed image analysis, and bioinformatics. He has already published more than 120 research papers in internationally reputed journals and refereed conferences, has edited eight books and is acting as a member of the editorial board of various international journals.

1 Introduction

Change detection, one of the most challenging tasks in the field of pattern recognition and machine learning (Bishop, 2006), is a process of finding out the temporal effects of multi-temporal images (Singh, 1989; Canty, 2007; Radke et al., 2005). This process detects the changes in a land-cover over time by analysing pairs of remotely sensed images of a geographical area captured at different time instants. The changes might occur due to several reasons, e.g., natural hazards, urban growth, deforestation. Process of change detection is applied to various domains like analysing land-use change (Zhang et al., 2002), monitoring urban growth (Merril and Jiajun, 1998), identifying burned area (Bruzzone and Prieto, 2000).

One can view the problem of change detection as an image segmentation one, where two groups of pixels are formed, one for the changed class and the other for the unchanged one. Methodology of change detection is broadly classified into two categories: supervised (Camps-Valls et al., 2008) and unsupervised (Ghosh et al., 2007, 2009, 2011; Mishra et al., 2012). Supervised techniques enjoy several advantages, e.g., the same can explicitly recognise the kinds of changes occurred and is robust to variation of atmospheric and light conditions at two acquisition dates. Various methods exist in the literature to carry out supervised change detection, e.g., post classification (Singh, 1989), direct multidate classification (Singh, 1989), kernel-based methods (Camps-Valls et al., 2008), etc. However, exploration of supervised methods in change detection is limited owing to the constraint of the requirement of sufficient amount of ground truth information and such information is expensive, hard and monotonous to collect. On the contrary, in unsupervised approach, such additional information is not required. Due to the insufficient number of labelled patterns, unsupervised techniques seem to be the only option for change detection.

Unsupervised change detection process is of two types: context insensitive (Singh, 1989) and context sensitive (Ghosh et al., 2007, 2009, 2011; Mishra et al., 2012; Kasetkasem and Varshney, 2002). Histogram thresholding (Singh, 1989; Patra et al., 2011) is the simplest unsupervised context insensitive change detection method which suffers from the problem of not considering the spatial correlation among neighbourhood pixels in the decision process. To overcome this difficulty, context sensitive methods using Markov random fields (MRF) (Kasetkasem and Varshney, 2002) are developed. These techniques also posses certain difficulties like requirement of the selection of a suitable model for statistical distribution of the changed and the unchanged class pixels. Change detection methodologies based on neural networks, both using supervised and unsupervised learning (Ghosh et al., 2007, 2009), exist in the literature and they are free from such limitations. Generally, the following three steps are carried out, one after another, for unsupervised change detection:

1 image preprocessing

2 image comparison and

3 image analysis (Singh, 1989).

Images of the same geographical area, captured at different time instants, constitute the input of the change detection process. In preprocessing step, these images are made compatible by operations like radiometric and geometric corrections, co-registration, noise reduction (Singh, 1989; Ghosh et al., 2009), etc. After preprocessing, image comparison is done, pixel by pixel, to generate a difference image (DI) which, in turn, is used for change detection. The difference image can be generated in many ways, e.g., univariate image differencing, change vector analysis (CVA), image rationing (Singh, 1989). In the present work, CVA technique is adopted to generate DI.

In change detection, the class labels of some labelled patterns may be made available by experts. But, if the number of these labelled samples is less, then the available information may be insufficient for developing any supervised method. Under this circumstance, knowledge of labelled patterns will be completely unutilised if unsupervised approach is used. This motivated us to use the semi-supervised approach

(Zhu, 2008; Chapelle et al., 2006) instead of an unsupervised or supervised one. Semi-supervision has been explored successfully for improving the performance of clustering and classification tasks (Basu et al., 2002, 2003; Wagstaff et al., 2001; Yeung and Chang, 2007; Hou et al., 2011) when sufficient amount of labelled data is absent. In semi-supervised classification, training is performed using abundant unlabelled patterns along with a few labelled patterns; whereas semi-supervised clustering utilises a few labelled samples for determining more accurate clusters. Two different approaches are used for semisupervised clustering: search-based approach and similarity-based approach (Basu et al., 2003). Search-based approaches use the labelled patterns to search for an accurate optimum partitioning. In similarity-based approaches, the labelled patterns are utilised to adopt the underlying similarity metrics.

In the present work, five semi-supervised K-Means algorithms namely, COP-KMeans (Wagstaff et al., 2001), constrained-KMeans (Basu et al., 2002) and seeded-KMeans (Basu et al., 2002), semi-supervised-HMRF-KMeans (Basu et al., 2004) and semi-supervised-kernel-KMeans (Kulis et al., 2005) have been studied in the domain of change detection. As per the knowledge of the authors, there are no such applications of these techniques in the area of change detection by using these semi-supervised variants of K-Means algorithm. This motivated us to pursue the present study. Comparative analysis between these techniques and standard K-Means algorithm (MacQueen, 1967) has been done for two multi-temporal and multi-spectral remotely sensed images. Comparison reveals that constrained-KMeans, for both the datasets, is more applicable for change detection than COP-KMeans and Seeded-KMeans; whereas semi-supervised-HMRF-KMeans and semi-supervised-kernel-KMeans algorithms are not robust for all the datasets.

The rest of the article is organised into six sections. Section 2 describes the process of generation of input patterns using CVA technique. Section 3 describes the procedure for collecting labelled patterns (or constraints) for experimental purpose. A brief description of the standard K-Means algorithm and the five semi-supervised variants of K-Means algorithm, used in the present article, is given in Section 4. Description of the datasets used for experimentation is provided in Section 5. In Section 6, implementation details and experimental results are given. Conclusion is drawn in Section 7.

2 Generation of input pattern

The difference image $D = \{l_{mn}, 1 \le m \le p, 1 \le n \le q\}$ is produced by the CVA technique (Singh, 1989) from the two co-registered and radiometrically corrected γ-spectral band images Y_1 and Y_2 of size $p \times q$ of the same geographical area captured at different times T1 and T2. Here, grey value of the difference image, D, at spatial position (m, n), denoted as l_{mn}, is calculated as,

$$l_{mn} = (\text{int})\sqrt{\sum_{\alpha=1}^{\gamma}\left(l_{mn}^{\alpha}(Y_1) - l_{mn}^{\alpha}(Y_2)\right)^2},$$

where $l_{mn}^{\alpha}(Y_1)$ and $l_{mn}^{\alpha}(Y_2)$ are the grey values of the pixels at the spatial position (m, n) in the α^{th} band of the images Y_1 and Y_2, respectively.

From the difference image, the input pattern for a particular pixel position is generated by considering the grey value of the said pixel as well as those of its neighbouring ones to exploit (spatial) contextual information from neighbours. In the present methodology, 2nd order neighbourhood system is used. Here, each input pattern consists of nine features, one grey value of the concerned pixel and eight grey values corresponding to its eight neighbours. Here, the y-dimensional input pattern of the $(m, n)^{\text{th}}$ pixel position of DI is denoted by $\overrightarrow{X_{mn}} = [x_{mn,1}, x_{mn,2}, \ldots, x_{mn,y}]$.

3 Labelled patterns (or constraints) collection

Semi-supervised variants of K-Means algorithm utilise insufficient labelled information either in the form of seed data (labelled pattern) or constraint while performing the partitioning (iterative) by the standard K-Means algorithm. In the present work, for experimental purpose, labelled patterns for both the classes are picked up from the ground truth with equal percentage. These labelled information can be used for obtaining constraints. There are mainly two types of constraints: 'must-link' and 'cannot-link'. 'Must-link' constraint ensures that a pair of patterns must be in the same group while 'cannot-link' constraint specifies that the said two patterns can not belong to the same group. For generating the constraints, each combination of the pattern pair is considered at a time. Now, if they are in the same class, then 'must-link' constraint is generated; else, 'cannot-link' constraint is generated.

4 Description of background methodologies

As mentioned earlier, in the present investigation standard K-Means, COP-KMeans, constrained-KMeans, seeded-KMeans, semi-supervised-HMRF-KMeans and semi-supervised-kernel-KMeans algorithms are used. These algorithms are described in brief in the following subsections.

4.1 K-Means algorithm

In K-Means algorithm (MacQueen, 1967), initially, K number of patterns (for K number of clusters) are randomly selected from a set of unlabelled patterns and they correspond to initial cluster centres. Let, v_1, v_2, …, v_K represent these K cluster centres. In each iteration, the Euclidean distance of the unlabelled patterns from each of the cluster centres is computed one by one. Then, an unlabelled pattern is assigned to the cluster for which this distance is minimum. After that, for each (i^{th}) cluster, its centre (v_i, i = 1, 2, …, K) is updated by the arithmetic mean of the patterns (feature wise) assigned to the said cluster. This process (partitioning and assignment) is repeated until the following objective function is minimised:

$$O_{kmeans} = \sum_{j=1}^{K} \sum_{X_{mn} \in \chi_j} \left\| X_{mn} - v_j \right\|; \tag{1}$$

where χ_j is the set of patterns assigned to cluster j. The algorithm terminates when no more changes occur from the partitioning point of view.

4.2 COP-KMeans algorithm

In COP-KMEANS, proposed by Wagstaff et al. (2001), the labelled information is used in the form of 'must-link' and 'cannot-link' constraints during the partitioning process of K-Means algorithm. In this algorithm, both the constraints are to be satisfied for assigning a pattern to a particular cluster. As a result, unlike standard K-Means algorithm, where an unlabelled pattern is assigned to the nearest cluster, in COPK-Means, a sorted list of clusters (in ascending order based on the distance of the pattern from each of the cluster centres) is generated for each of the unlabelled patterns before assigning it to a specific cluster.

Initially, the first one from the sorted list is selected and the unlabelled pattern is assigned to the said nearest cluster if no constraints are violated. This reflects that the patterns, which were already assigned in that cluster, are not associated with the 'cannot-link' constraint for the said unlabelled pattern. Likewise, the patterns which were already assigned in different clusters, are not in the 'must-link' constraint with the concerned pattern. If any of the constraints is violated, the next cluster from the sorted list is checked for assignment. This process continues until a cluster is found where the pattern could be assigned or the list is exhausted. If no such cluster is found then it can be said that the partitioning is not possible with the initial cluster centres without violating the constraints. Once the assignment is done, the rest of the steps are similar to those of standard K-Means algorithm.

4.3 Seeded-KMeans algorithm (Wagstaff et al., 2001)

As mentioned earlier, in standard K-Means algorithm, the initial cluster centres (K in number) are randomly chosen from a set of patterns. While, in seeded-KMeans, labelled patterns (seed data) are utilised for selecting the initial cluster centres. The i^{th} cluster centre is initialised by computing the arithmetic mean of all the labelled patterns belonging to the i^{th} cluster. In this algorithm, labelled patterns are only used in the initialisation phase of the algorithm. Afterwards, this class label information may get changed while executing the other steps (partitioning and assignment) of the algorithm in a repetitive fashion.

4.4 Constrained-KMeans algorithm (Wagstaff et al., 2001)

In constrained-KMeans algorithm, selection step of initial cluster centres is the same as it is done in seeded-KMeans algorithm. But, unlike seeded-KMeans, here the labelling of seed data is not reestimated (changed) during the remaining phase of the algorithm.

4.5 Semi-supervised K-Means algorithm based on hidden MRF (Basu et al., 2004)

Basu et al. (2004) proposed a semi-supervised K-Means algorithm based on Hidden Markov Model (HMRF). HMRF has two components:

1 a set of hidden variables, L, i.e., unobserved cluster labels for the patterns

2 a set of observed random variables, Ψ, i.e., the patterns.

Here, maximum a-posteriori probability (MAP) configuration of the HMRF model is the same as maximising the posteriori probability $Pr(L|\Psi)$. The neighbourhood of the MRF is considered using the generalised Potts potential function to incorporate the constraints in the iterative learning of the K-Means algorithm. Here, the objective function of the semi-supervised K-Means clustering is modified as,

$$
\begin{aligned}
O_{obj} = & \sum_{j=1}^{K} \sum_{X_{mn} \in \chi_j} D\left(X_{mn} - v_j\right) + \sum_{(X_{mn}, X_{m'n'}) \epsilon M} w_{(mn,m'n')} 1\left[label_{mn} \neq label_{m'n'}\right] \\
& + \sum_{(X_{mn}, X_{m'n'}) \epsilon C} w_{(mn,m'n')} 1\left[label_{mn} = label_{m'n'}\right];
\end{aligned}
\tag{2}
$$

where M is the set of 'must-link' constraints and C is the set of 'cannot-link' constraints. $w_{(mn,m'n')}$ is the penalty cost for violating a constraint between the patterns. X_{mn} and $X_{m'n'}$ and the class assignment of the $(m, n)^{th}$ pattern denotes by $label_{mn}$. 1 is the indicator function, where $(1[true] = 1, 1[False] = 0)$. $D(X_{mn}, v_j)$ is the distortion measure between $(m, n)^{th}$ pattern and the j^{th} cluster centre. Here, Euclidean 7 distance as well as non-Euclidean distance can be used for distortion measure. In the present investigation, parameterised I-divergence is used as the distortion measure and gradient decent approach is used for iterative learning of the distortion measure.

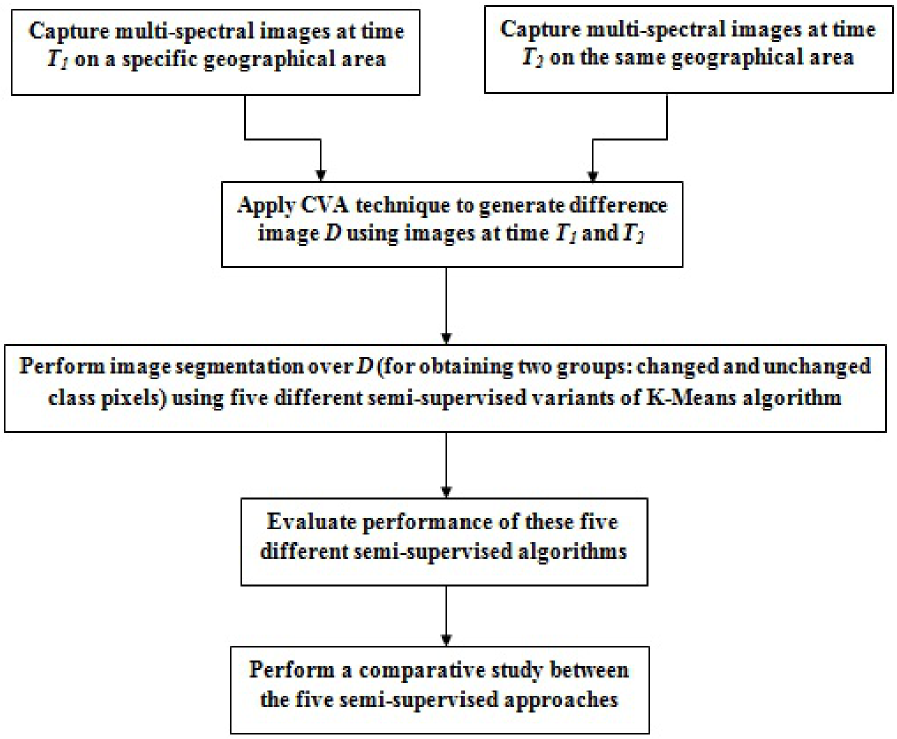
4.6 Semi-supervised kernel-based K-Means algorithm (Kulis et al., 2005)

Kulis et al. proposed a weighted kernel-based approach in semi-supervised learning framework. Here, the objective function of the K-Means algorithm is expressed as following:

$$
\begin{aligned}
O_{obj} = & \sum_{j=1}^{K} \sum_{X_{mn} \in \chi_j} \alpha_{mn} \left\|\Phi\left(X_{mn}\right) - v_j\right\|^2 - \sum_{(X_{mn}, X_{m'n'}) \epsilon M} w_{(mn,m'n')} 1\left[label_{mn} = label_{m'n'}\right] \\
& + \sum_{(X_{mn}, X_{m'n'}) \epsilon C} w_{(mn,m'n')} 1\left[label_{mn} = label_{m'n'}\right];
\end{aligned}
\tag{3}
$$

where $\Phi(X_{mn})$ is a function that maps X_{mn} to a (generally) higher dimensional space and α_{mn} is a non-negative weight associated with each of the $(m, n)^{th}$ patterns. Here, instead of giving the penalty for constraint violation, the term 'reward' is introduced for obeying the constraints. In practice, this is done by subtracting the penalty term when the patterns, those are in the must-link constraint, are also assigned in the same cluster. In the present investigation, the exponential kernel is used for non-linear mapping. The block diagram of the present study is given in Figure 1.

Figure 1 Block diagram of the present study



5 Description of datasets

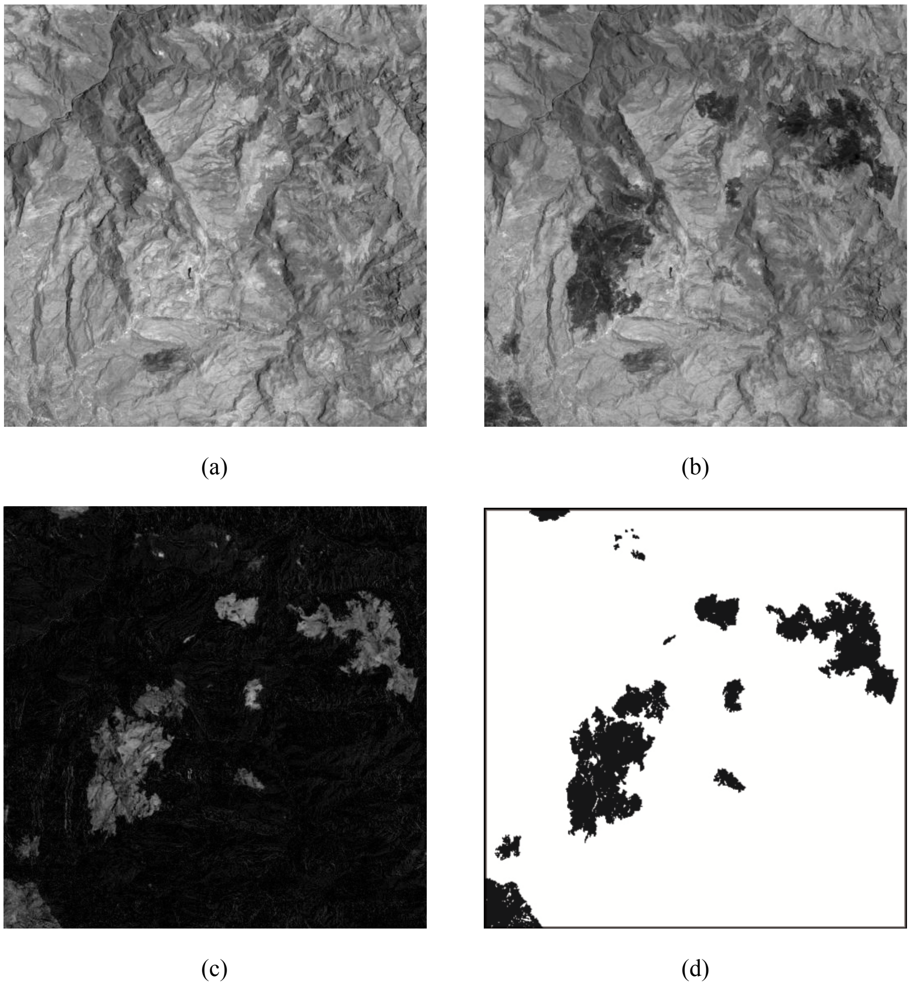
To evaluate the effectiveness of the proposed methodology, experiments are conducted on two multitemporal remotely sensed images corresponding to the geographical areas of Mexico and Sardinia Island of Italy.

5.1 Dataset related to Mexico area (Ghosh et al., 2009)

This dataset consists of two multi-spectral images of the Landsat-7 satellite captured by the Landsat enhanced thematic mapper plus (ETM+) sensor over an area of Mexico acquired on 18th April, 2000 and 20th May, 2002. From the entire available Landsat scene, a section of 512 × 512 pixels are selected as test site. A fire destroyed a large portion of the vegetation in the considered region between two acquisition dates. Figures 2(a) and 2(b), respectively show the band 4 images of April, 2000 and May, 2002. The difference image [Figure 2(c)] created by spectral band 4 using CVA technique is only used for further analysis. To evaluate the performance of the algorithms, a reference map [Figure 2(d)] is used.

The reference map contains 25,599 changed and 236,545 unchanged pixels.

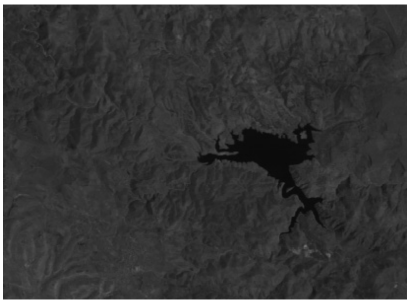
Figure 2 Images of Mexico area, (a) band 4 image acquired in April 2000, (b) band 4 image acquired in May 2002, (c) corresponding difference image, and (d) a reference map of the changed area



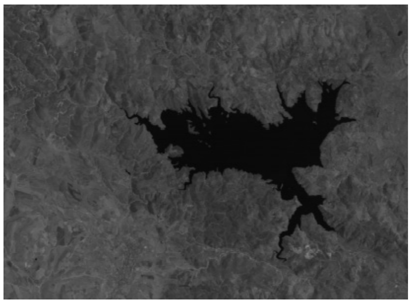
5.2 *Dataset related to Sardinia Island, Italy (Ghosh et al., 2009)*

Two multi-spectral images are acquired in September, 1995 and July, 1996 by the Landsat thematic mapper (TM) sensor of the Landsat-5 satellite. The test site of 412 × 300 pixels of a scene includes the lake Mulargia on the Sardinia Island, Italy. The water level of the lake increased between two acquisition dates (lower central part of the image reflects the same). Figures 3(a) and 3(b), respectively, show the 1995 and 1996 images of band 4. CVA technique has been applied on spectral bands 1, 2, 4, and 5 of these two images to obtain the difference image [Figure 3(c)]. 7,480 changed and 116,120 unchanged pixels are present in the reference map [Figure 3(d)].

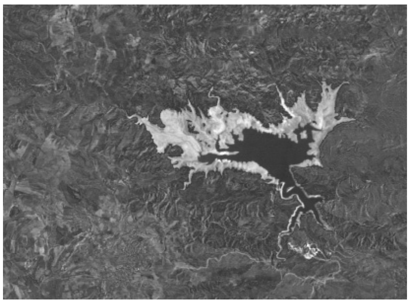
Figure 3 Images of Sardinia Island, Italy, (a) band 4 image acquired in September 1995, (b) band 4 image acquired in July 1996, (c) difference image generated by CVA technique using bands 1, 2, 4, and 5, and (d) a reference map of the changed area



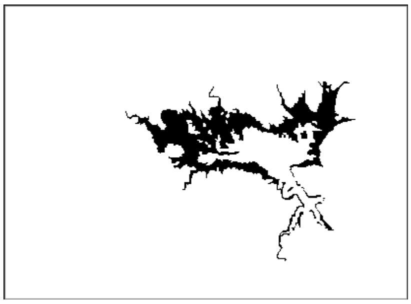
(a)



(b)



(c)



(d)

6 Results and analysis

As mentioned in Section I, to investigate the effectiveness of the algorithms considered in our study [i.e., standard K-Means, COP-KMeans, constrained-KMeans, seeded-KMeans, semi-supervised-HMRFK-Means (SS-HMRF-KMeans), and semi-supervised-kernel-KMeans (SS-kernel-KMeans)], experiments are carried out on two multi-temporal remotely sensed images. Results of Mexico dataset and Sardinia dataset are given in Tables 1 and 2, respectively. For experimental purpose, three different percentages of training patterns (0.1%, 0.5%, and 1%) are considered and ten simulations are performed in each case. The performance of each of the algorithms is assessed using various performance measuring indices, e.g., the number of missed alarms (*MA*), the number of false alarms (*FA*), the number of overall error (*OE*), micro averaged F_1 measure (Micro F_1) (Halder et al., 2009), macro averaged F_1 measure (MacroF_1) (Halder et al., 2009), Kappa measure (*Kappa*) (Congalton and Green, 2009), and error probability (P_E). The best result in terms of minimum overall error (among ten simulations) is depicted in the tables. Experiments have been conducted using Visual C++ on a machine with Intel(R) Core(TM)i5 CPU 2.50 GHz and 4.0 GB RAM. Average execution time (over ten simulations) is also considered for assessment.

Results of the K-Means algorithm and five semi-supervised variants of K-Means algorithm (over ten simulations) are also statistically validated using paired t-test (Kreyszig, 1970). The paired t-test is performed at 5% level of significance. Here, for typical illustration, results of t-test in terms of p-score using Kappa measure with 1% training patterns are reported in Tables 3 and 4, for Mexico dataset and Sardinia dataset, respectively.

From Table 1, it is observed that for Mexico dataset, COP-KMeans algorithm is significantly better than the standard K-Means algorithm (considering all the percentages of training patterns used) in terms of all the measuring indices taken into consideration except the execution time. It has also been found that the required execution time mostly increases by increasing the number of labelled patterns or the constraints. This may be due to the fact that a significant amount of time is required for checking the constraints in each case before assigning it to any cluster. This might be a bottleneck of considering COP-KMeans algorithm for change detection under semi-supervised framework. Similar findings have been observed for COP-KMeans algorithm while experimenting with Sardinia dataset (Table 2).

From Tables 1 and 2, it has also been noticed that for both the datasets seeded-KMeans algorithm is performing better than standard K-Means in terms of time requirement. While executing the standard K-Means algorithm (for different simulations with different initial cluster centres), it has been observed that the results are not much sensitive to the choice of the initial cluster centres. This happens for both the datasets. In seeded-KMeans algorithm, since the labelled patterns are only used for selecting the initial cluster centres and if the performance of the said algorithm does not depend much on initialisation part, its performance cannot be better than its standard K-Means counterpart. Though a little, some amount of knowledge (obtained from the labelled patterns) is being used in the seeded-KMeans algorithm during the initialisation phase. On the contrary, K-Means algorithm uses randomisation for initialisation. As a result, rate of convergence of seeded-KMeans algorithm is higher than that of standard K-Means one and the results (shown in Tables 1 and 2) also corroborate to it.

By analysing the tables, it can be concluded that constrained-KMeans algorithm is exploiting the superiority of both COP-KMeans algorithm (in terms of different performance measuring indices used) and seeded-KMeans algorithm (w.r.t. CPU time requirement). From the results, it is found that constrained-KMeans algorithm is able to produce better output by consuming comparable amount of CPU time. In this algorithm, constraints or labels are fixed for labelled patterns during iterative partitioning process. So, time for checking the violation of the constraints is not required as it is needed in COP-KMeans algorithm resulting in lower CPU time requirement. Values of the performance measuring indices are found to be either similar (for Mexico dataset) or better (for Sardinia dataset) in case of constrained-KMeans algorithm as compared to COP-KMeans one.

Table 1 Results on Mexico dataset

Techniques used	*Training patterns*	*MA*	*F A*	*OE*	*Micro F_1*	*Macro F_1*	*KM*	P_E	*Time (in second)*
K-Means	-	3,107	665	3,772	0.9581	0.9574	0.9147	0.0144	3.7
COP-KMeans	0.1%	3,099	664	3,763	0.9582	0.9575	0.9149	0.0144	5.8
	0.5%	3,077	661	3,738	0.9585	0.9578	0.9155	0.0143	11.0
	1.0%	3,063	657	3,720	0.9587	0.9580	0.9160	0.0142	20.4
Seeded-KMeans	0.1%	3,107	665	3,772	0.9581	0.9574	0.9147	0.0144	3.9
	0.5%	3,107	665	3,772	0.9581	0.9574	0.9147	0.0144	2.7
	1.0%	3,107	665	3,772	0.9581	0.9574	0.9147	0.0144	2.6
Constrained-KMeans	0.1%	3,099	664	3,763	0.9582	0.9575	0.9149	0.0144	2.8
	0.5%	3,077	661	3,738	0.9585	0.9578	0.9155	0.0143	2.5
	1.0%	3,063	657	3,720	0.9587	0.9580	0.9160	0.0142	2.5
SS-HMRF-K-Means	0.1%	671	2,428	3,099	0.9678	0.9674	0.9349	0.0118	5.6
	0.5%	666	2,416	3,082	0.9680	0.9676	0.9353	0.0118	22.6
	1%	662	2,415	3,077	0.9680	0.9677	0.9354	0.0117	56.7
SS-kernel-K-Means	0.1%	9,109	225	9,334	0.8916	0.8800	0.7610	0.0356	6.1
	0.5%	9,030	226	9,256	0.8926	0.8812	0.7634	0.0353	20.1
	1%	8,928	227	9,155	0.8939	0.8828	0.7664	0.0349	34.6

Table 2 Results on Sardinia dataset

Techniques used	*Training patterns*	*MA*	*F A*	*OE*	*Micro F_1*	*Macro F_1*	*KM*	*P_E*	*Time (in second)*
K-Means	-	637	1879	2516	0.9184	0.9169	0.8339	0.0204	2.7
COP-KMeans	0.1%	637	1876	2513	0.9185	0.9170	0.8341	0.0203	4
	0.5%	635	1865	2500	0.9189	0.9174	0.8348	0.0202	6.5
	1.0%	631	1839	2470	0.9197	0.9183	0.8366	0.0200	10.1
Seeded-KMeans	0.1%	637	1879	2516	0.9184	0.9169	0.8339	0.0204	2.1
	0.5%	637	1879	2516	0.9184	0.9169	0.8339	0.0204	1.8
	1.0%	637	1879	2516	0.9184	0.9169	0.8339	0.0204	1.8
Constrained-KMeans	0.1%	637	1876	2513	0.9185	0.9170	0.8341	0.0203	2
	0.5%	635	1858	2493	0.9191	0.9176	0.8352	0.0202	1.9
	1.0%	631	1839	2470	0.9197	0.9183	0.8366	0.0200	1.8
SS-HMRF-K-Means	0.1%	210	9857	10067	0.8111	0.7728	0.5532	0.0814	5.8
	0.5%	207	9749	9956	0.8122	0.7745	0.5564	0.0805	21.9
	1%	211	9592	9803	0.8136	0.7766	0.5604	0.0793	42
SS-kernel-K-Means	0.1%	1137	540	1677	0.9385	0.9380	0.8761	0.0136	0.9
	0.5%	1119	547	1666	0.9389	0.9385	0.8771	0.0135	2.1
	1%	1134	525	1659	0.9391	0.9386	0.8773	0.0134	3.1

Table 3 Results of paired t-test performed with different clustering techniques in terms of *p*-score for Mexico dataset (using 1% training patterns)

Methods used	*K-Means*	*COP-KMeans*	*Seeded-KMeans*	*Constrained-KMeans*	*SS-HMRF-KMeans*	*SS-kernel-KMeans*
K-Means	-	**2.0104×10^{-11}**	-	**1.9822×10^{-11}**	**1.9028×10^{-20}**	**2.6199×10^{-23}**
COP-KMeans	**2.0104×10^{-11}**	-	**2.0104×10^{-11}**	0.3434	**1.0538×10^{-19}**	**1.6403×10^{-23}**
Seeded-KMeans	-	**2.0104×10^{-11}**	-	**1.9822×10^{-11}**	**1.9028×10^{-20}**	**2.6199×10^{-23}**
Constrained-KMeans	**1.9822×10^{-11}**	0.3434	**1.9822×10^{-11}**	-	**1.0709×10^{-19}**	**1.5117×10^{-23}**
SS-HMRF-KMeans	**1.9028×10^{-20}**	**1.0538×10^{-19}**	**1.9028×10^{-20}**	**1.0709×10^{-19}**	-	**8.0710×10^{-24}**
SS-kernel-KMeans	**2.6199×10^{-23}**	**1.6403×10^{-23}**	**2.6199×10^{-23}**	**1.5117×10^{-23}**	**8.0710×10^{-24}**	-

Table 4 Results of paired t-test performed with different clustering techniques in terms of *p*-score for Sardinia dataset (using 1% training patterns)

Methods used	*K-Means*	*COP-KMeans*	*Seeded-KMeans*	*Constrained-KMeans*	*SS-HMRF-KMeans*	*SS-kernel-KMeans*
K-Means	-	**6.3852×10^{-8}**	**8.5381×10^{-6}**	**2.1517×10^{-8}**	**1.3427×10^{-9}**	**2.2032×10^{-22}**
COP-KMeans	**6.3852×10^{-8}**	-	**1.1479×10^{-7}**	0.2153	**1.3049×10^{-9}**	**7.9425×10^{-20}**
Seeded-KMeans	**8.5381×10^{-6}**	**1.1479×10^{-7}**	-	**3.5053×10^{-8}**	**1.3383×10^{-9}**	**3.0354×10^{-22}**
Constrained-KMeans	**2.1517×10^{-8}**	0.2153	**3.5053×10^{-8}**	-	**1.3010×10^{-9}**	**6.6390×10^{-20}**
SS-HMRF-KMeans	**1.3427×10^{-9}**	**1.3049×10^{-9}**	**1.3383×10^{-9}**	**1.3010×10^{-9}**	-	**3.8902×10^{-10}**
SS-kernel-KMeans	**2.2032×10^{-22}**	**7.9425×10^{-20}**	**3.0354×10^{-22}**	**6.6390×10^{-20}**	**3.8902×10^{-10}**	-

From Table 1, it is noticed for Mexico dataset that SS-HMRF-KMeans outperforms standard K-Means algorithm as well as the other semi-supervised variants of K-Means algorithm in terms of all the performance measuring indices except the case of average execution time. Increase in time requirement may be due to the same reason as COP-KMeans. It has been also observed that the performance of semi-supervised kernel K-Means algorithm is worse in term of all the performance measuring indices. The probable reason behind this may be the use of exponential kernel for mapping.

By analysing the results in Table 2, it is found that semi-supervised HMRF-based K-Means algorithm performed very bad on Sardinia dataset. This may be due to the fact that parameterised I-divergence may not be applicable for this data. It is seen that semi-supervised kernel-based K-Means algorithm is well-applicable for Sardinia dataset than all other algorithms (used for comparison). From Tables 1 and 2, it has been noticed that the values for most of the different performance measuring indices, are steadily improving with increase in the percentage of labelled patterns for almost all the algorithms except seeded-KMeans algorithm. It is noticed that with increase in labelled patterns, in most of the cases, for Mexico dataset, the number of missed alarms decreases more than the number of false alarms, whereas the reverse situation occurs for Sardinia dataset.

In Tables 3 and 4, statistically significant results in terms of *p*-score of the paired t-test (at 5% level of significance) are marked as bold. It is also found for both the datasets that results of the semi-supervised variants (over ten simulations) are significantly different (at 5% level) from each other for most of the pairs except the pair between COP-KMeans and constrained-KMeans. This also corroborates our earlier findings.

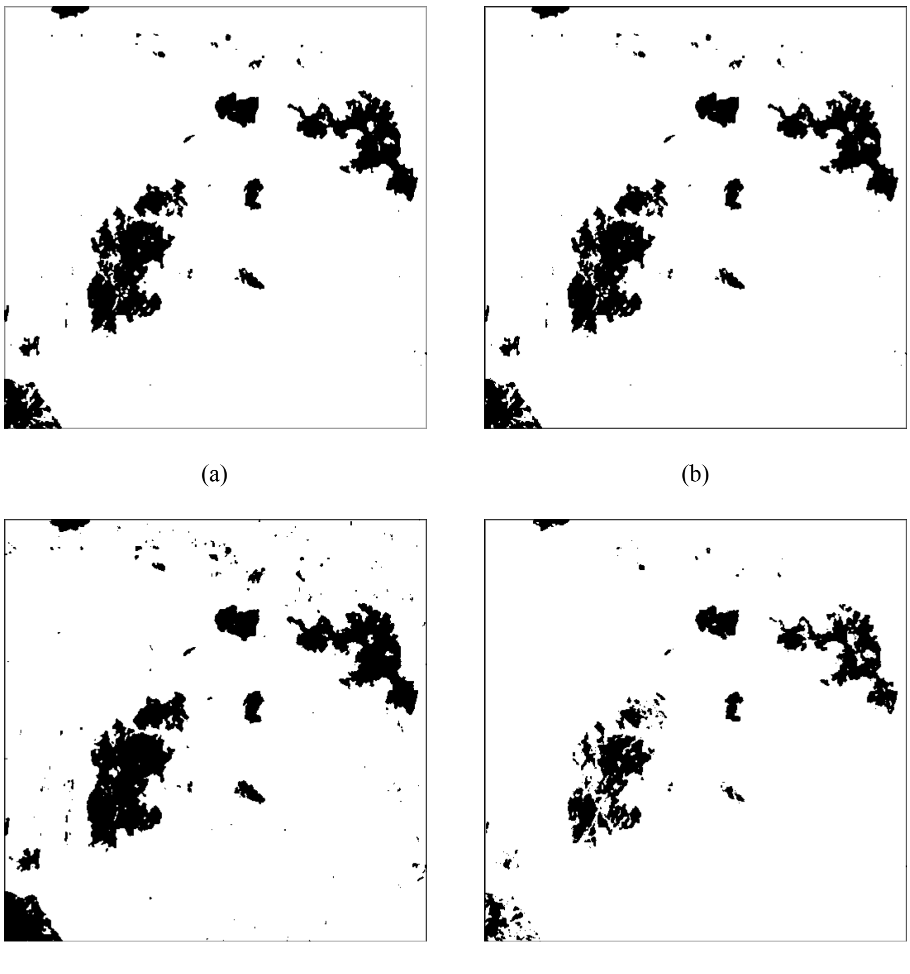
It has been observed from the tables that the overall error in the best case, obtained by seeded-KMeans and standard K-Means algorithms are similar. The performance of COP-KMeans and constrained-KMeans are also similar for both the datasets. For this reason, the change detection maps corresponding to minimum overall error (obtained over ten simulations), using standard K-Means algorithm and constrained-KMeans algorithm are only shown for visual illustration. The change detection maps are also depicted for semi-supervised-HMRF-KMeans and semi-supervised-kernel-KMeans algorithm. Figures 4 and 5 show the corresponding maps for Mexico and Sardinia dataset, respectively. For Mexico dataset, Figure 4(a) shows the map obtained using K-Means algorithm while Figures 4(b) to 4(d) show the maps using constrained-KMeans, SS-HMRF-KMeans and SS-kernel-KMeans algorithms with 1% training patterns, respectively. Corresponding change detection maps for Sardinia dataset are displayed in Figures 5(a) to 5(d). It has been observed that for Mexico dataset the maps obtained using semi-supervised HMRF-based K-Means algorithm and for Sardinia dataset, the maps obtained using semi-supervised kernel-based K-Means algorithm are more accurate resemblance of the reference map. This corroborates to our earlier findings regarding the superiority of the algorithms used in our investigation.

7 Conclusions

In this article, performance of some of the semi-supervised K-Means clustering algorithms (namely, COP-KMeans, constrained-KMeans, seeded-KMeans, SS-HMRF-

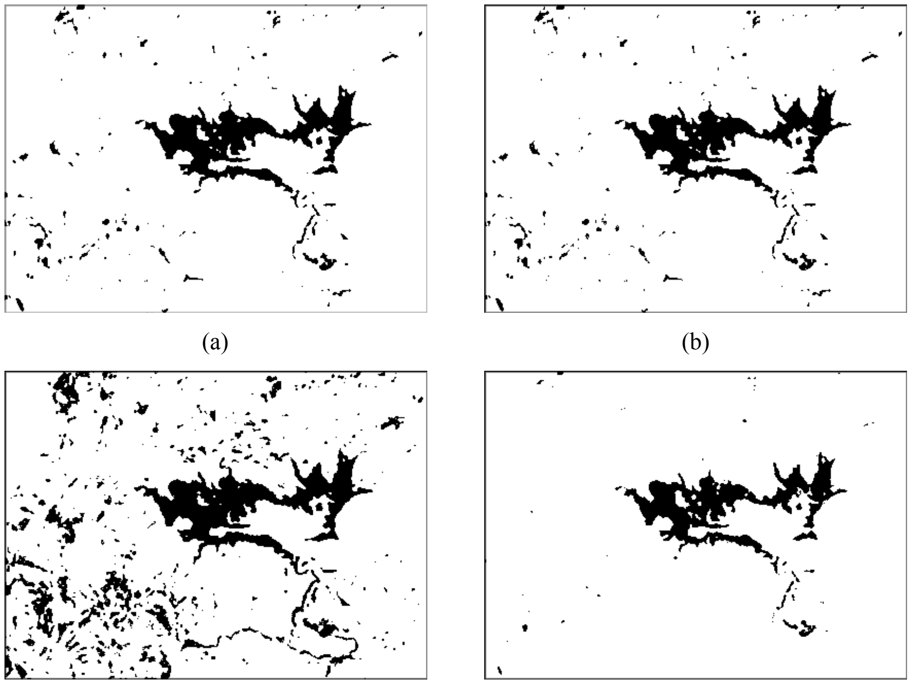
KMeans, and SS-kernel-KMeans) is studied for change detection. To assess the effectiveness of these algorithms, experiments are conducted on two multi-temporal remotely sensed images. By analysing the results, it can be concluded that constrained-KMeans algorithm for both the datasets is more applicable for change detection than COP-KMeans and seeded-KMeans in terms of both execution time requirement and quality of results (i.e., w.r.t other performance measuring indices); while HMRF-based semi-supervised K-Means algorithm and kernel-based semi-supervised K-Means algorithm achieve better performance than other semi-supervised variants of K-Means algorithm in case of only one dataset. In future work, we plan to carry out a similar study in the domain of change detection using semi-supervised graph-based algorithms.

Figure 4 Change detection maps obtained for Mexico dataset, (a) using K-Means algorithm, (b) using constrained-KMeans algorithm, (c) using SS-HMRF-KMeans algorithm, and (d) using SS-kernel-KMeans algorithm (with 1% training pattern)



(a) (b) (c) (d)

Figure 5 Change detection maps obtained for Sardinia dataset, (a) using K-Means algorithm, (b) using constrained-KMeans algorithm, c) using SS-HMRF-KMeans algorithm, and (d) using SS-kernel-KMeans algorithm (with 1% training pattern)



(a) (b)

(c) (d)

Acknowledgements

The authors like to thank the reviewers for their thorough and constructive comments which helped to enhance the quality of the article. The authors are also grateful to the Department of Science and Technology (DST), Government of India and University of Trento, Italy, the sponsors of the ITPAR programme and Prof. L. Bruzzone for providing the data. Moumita Roy is grateful to Council of Scientific and Industrial Research (CSIR), India for providing her a Senior Research Fellowship [No. 09/096(0684)2k11-EMR-I].

References

Basu, S., Banerjee, A. and Mooney, R.J. (2002) 'Semi-supervised clustering by seeding', in C. Sammut and A.G. Hoffmann (Eds.): *Proceedings 19th International Conference on Machine Learning (ICML-2002)*, pp.27–34, Sydney, Australia: Morgan Kaufmann, San Fransisco, USA.

Basu, S., Bilenko, M. and Mooney, R.J. (2004) 'A probabilistic framework for semi-supervised clustering', in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*, pp.59–68, Seattle, WA.

Basu, S., Bilenko, M. and Mooney, R.J. (2003) 'Comparing and unifying search-based and similarity-based approaches to semi-supervised clustering', in *Proceedings of the ICML-2003 Workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining Systems*, pp.42–49, Washington DC, USA.

Bishop, C.M. (2006) *Pattern Recognition and Machine Learning*, Springer, New York, USA.

Bruzzone, L. and Prieto, D.F. (2000) 'An adaptive parcel-based technique for unsupervised change detection', *International Journal of Remote Sensing*, Vol. 21, No. 4, pp.817–822.

Camps-Valls, G., Gomez-Chova, L., Muñoz-Mari, J., Rojo-Alvarez, J.L. and Martinez-Ramon, M. (2008) 'Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection', *IEEE Transactions on Geoscience & Remote Sensing*, Vol. 46, No. 6, pp.1822–1835.

Canty, M.J. (2007) *Image Analysis, Classification and Change Detection in Remote Sensing: with Algorithms for ENVI/IDL*, CRC Press, Taylor & Francis Group, Boca Raton.

Chapelle, O., Scholkopf, B. and Zien, A. (2006) *Semi-supervised Learning*, MIT Press, Cambridge.

Congalton, R.G. and Green, K. (2009) *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed., CRC Press, Boca Raton, USA.

Ghosh, A., Mishra, N.S. and Ghosh, S. (2011) 'Fuzzy clustering algorithms for unsupervised change detection in remote sensing images', *Information Sciences*, Vol. 181, No. 4, pp.699–715.

Ghosh, S., Bruzzone, L., Patra, S., Bovolo, F. and Ghosh, A. (2007) 'A context-sensitive technique for unsupervised change detection based on Hopfield-type neural networks', *IEEE Transactions on Geoscience & Remote Sensing*, Vol. 45, No. 3, pp.778–789.

Ghosh, S., Patra, S. and Ghosh, A. (2009) 'An unsupervised context-sensitive change detection technique based on modified self-organizing feature map neural network', *International Journal of Approximate Reasoning*, Vol. 50, No. 1, pp.37–50.

Halder, A., Ghosh, A. and Ghosh, S. (2009) 'Aggregation pheromone density based pattern classification', *Fundamenta Informaticae*, Vol. 92, No. 4, pp.345–362.

Hou, C., Nie, F., Wang, F., Zhang, C. and Wu, Y. (2011) 'Semisupervised learning using negative labels', *IEEE Transactions on Neural Networks*, Vol. 22, No. 3, pp.420–432.

Kasetkasem, T. and Varshney, P.K. (2002) 'An image change detection algorithm based on Markov random field models', *IEEE Transactions on Geoscience & Remote Sensing*, Vol. 40, No. 8, pp.1815–1823.

Kreyszig, E. (1970) *Introductory Mathematical Statistics*, John Wiley, New York.

Kulis, B., Basu, S., Dhillon, I. and Mooney, R.J. (2005) 'Semi-supervised graph clustering: a kernel approach', in *Proceedings of the 22nd International Conference on Machine Learning*, Germany.

MacQueen, J. (1967) 'Some methods for classification and analysis of multivariate observations', in L.M.L. Cam and J. Neyman (Eds.): *Proceedings 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp.281–297, Statistical Laboratory of the University of California, Berkeley: University of California Press, Berkeley, Calif.

Merril, K.R. and Jiajun, L. (1998) 'A comparison of four algorithms for change detection in an urban environment', *Remote Sensing of Environment*, Vol. 63, No. 2, pp.95–100.

Mishra, N.S., Ghosh, S. and Ghosh, A. (2012) 'Fuzzy clustering algorithms incorporating local information for change detection in remotely sensed images', *Applied Soft Computing*, Vol. 12, No. 8, pp.2683–2692.

Patra, S., Ghosh, S. and Ghosh, A. (2011) 'Histogram thresholding for unsupervised change detection of remote sensing images', *International Journal of Remote Sensing*, Vol. 32, No. 21, pp.6071–6089.

Radke, R.J., Andra, S., Al-Kofahi, O. and Roysam, B. (2005) 'Image change detection algorithms: a systematic survey', *IEEE Transactions on Image Processing*, Vol. 14, No. 3, pp.294–307.

Singh, A. (1989) 'Digital change detection techniques using remotely-sensed data', *International Journal of Remote Sensing*, Vol. 10, No. 6, pp.989–1003.

Wagstaff, K., Cardie, C., Rogers, S. and Schroedl, S. (2001) 'Constrained K-Means clustering with background knowledge', in C.E. Brodley and A.P. Danyluk (Eds.): *Proceedings 18th International Conference on Machine Learning (ICML-2001)*, pp.577–584, Morgan Kaufmann, Williamstown, MA, USA.

Yeung, D-Y. and Chang, H. (2007) 'A kernel approach for semisupervised metric learning', *IEEE Transactions on Neural Networks*, Vol. 18, No. 1, pp.141–149.

Zhang, Q., Wang, J., Peng, X., Gong, P. and Shi, P. (2002) 'Urban built-up land change detection with road density and spectral information from multi-temporal Landsat TM data', *International Journal of Remote Sensing*, Vol. 23, No. 15, pp.3057–3078.

Zhu, X. (2008) *Semi-supervised Learning Literature Survey*, Computer Sciences TR1530, University of Wisconsin, Madison.