



HAL
open science

Detection of Urban Areas using Genetic Algorithms and Kohonen Maps on Multispectral images

Djelloul Mokadem, Abdelmalek Amine, Zakaria Elberrichi, David Helbert

► **To cite this version:**

Djelloul Mokadem, Abdelmalek Amine, Zakaria Elberrichi, David Helbert. Detection of Urban Areas using Genetic Algorithms and Kohonen Maps on Multispectral images. *International Journal of Organizational and Collective Intelligence (IJOCI)*, 2018, 8 (1), pp.46 - 62. 10.4018/IJOCI.2018010104 . hal-01696710

HAL Id: hal-01696710

<https://hal.science/hal-01696710>

Submitted on 11 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Detection of Urban Areas using Genetic Algorithms and Kohonen Maps on Multispectral images

Djelloul Mokadem, GeCoDe Laboratory, Tahar Moulay University of Saida, Algeria

Abdelmalek Amine, GeCoDe Laboratory, Department of Computer Science, Tahar Moulay University of Saida, Algeria

Zakaria Elberrichi, Djilali Liabes University of SidiBelAbbes, Algeria

David Helbert, Xlim, CNRS, University of Poitiers, Chasseneuil-Futuroscope, France

ABSTRACT

In this article, the detection of urban areas on satellite multispectral Landsat images. The goal is to improve the visual interpretations of images from remote sensing experts who often remain subjective. Interpretations depend deeply on the quality of segmentation which itself depends on the quality of samples. A remote sensing expert must actually prepare these samples. To enhance the segmentation process, this article proposes to use genetic algorithms to evolve the initial population of samples picked manually and get the most optimal samples. These samples will be used to train the Kohonen maps for further classification of a multispectral satellite image. Results are obtained by injecting genetic algorithms in sampling phase and this paper proves the effectiveness of the proposed approach.

KEYWORDS

Genetic Algorithms, Kohonen Maps, Multispectral Satellite Image, Segmentation, Urban Areas

1. INTRODUCTION

With the recent innovations in Earth observation techniques (sensors, satellites, and data), “urban” remote sensing or “remote sensing urban applications” have rapidly gained popularity among urban planners and, more generally, administrations in charge of territorial planning.

Indeed, remote sensing makes it possible to increase our understanding of urban areas in different ways, although the real potential of this technique is often challenged by the complexity of the urban environment itself. (Xiaojun, 2011)

Previous studies have been carried out on the urban theme, notably the work of (Antoine, 2009), he tried to develop a method able to detect and qualify changes in small areas from Very High spatial Resolution (VHR) remote sensing data acquired at different dates and from different sources by comparing the textural properties of objects of interest. First, he tried to extract objects by a region growing segmentation and then, their texture are compared.

Another work has been carried out in this direction, (Ashish, 2010) used a technique based on fuzzy clustering approach which takes care of spatial correlation between neighboring pixels of the difference image produced by comparing two images acquired on the same geographical area at different times.

DOI: 10.4018/IJOICI.2018010104

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Using always the VHR satellite data, (Tobias, 2017) proposed a novel object-based approach for unsupervised change detection with focus on individual buildings, he first applied a principal component analysis together with a unique procedure to determine the number of relevant principal components is performed as a predecessor for change detection; then he used k-means clustering for discrimination of changed and unchanged buildings.

(Andrew, 2017) used a different technique called Import Vector Machine (IVM) which builds upon the popular Support Vector Machine (SVM) methodology (Ribana, 2012). To obtain the optimum classification, the IVM algorithm explores all possible subsets of training data for optimal selection (termed import vectors) which are derived through successively adding training data samples until a given convergence criterion is met (Ribana, 2012). Data samples are selected according to their contribution to the classification solution.

1.1. Other Literature on Image Segmentation and Visualization on Medical Science

(Chang, 2018) has used computational intelligence and proposed his MapReduce framework with fusion algorithm as techniques of visualization to simulate medical imaging in order to explore areas that cannot be easily achieved by conventional ways.

He (Chang, 2017) also used data analytics and visualization to study how cells and genes can become malignant and harmful to human bodies, the approach he adopted consist in presenting by analytics all complex concepts and sequence of events (taking place every second) in order to facilitate the understanding the complex biological transformations.

On weather studies:

A demonstration using Cloud Computing technologies, MapReduce and optimization techniques was exposed by (Chang, 2017) to simulate temperature distributions, analyze weather data and forecast temperatures based on studying trends from the historical data. Finally, numerical computing is transformed into visualization after classifying results into clusters in order to facilitate the interpretation of output results.

On Social networks:

In (Chang, 2017), the author presented his architecture for the Social Network Analysis Platform (SNAP) and developed his SocialNetwork API with which, he presented how to extract information from Facebook, which processes Big Data of the author's network. The details of network have been discussed and the techniques to analyze and visualize data have been presented. These techniques aim to understand for example the behavior of customers and thus the latest trend in the market.

However, in this work, we do not study the urban change on multi-temporal satellite images as it is mentioned in previous related works, but we focus on the problem of the detection of urban areas on multispectral Landsat satellite images (composite of bands 4, 3 and 1). Picking good samples requires the engagement of an expert in remote sensing image analysis to get an optimal classification according to an optimal training set; the suburban spaces are very often very heterogeneous and subject to abrupt and irregular changes in time and space (Antoine,2009).

So, given that the quality of the classification depends closely on the quality of sampling, we have the idea to improve the quantity and the quality of the samples from an initial and credible group of samples manually picked using the concept of the natural evolution of the species on which the idea of the genetic algorithms (GA in this paper) is founded (Holland, 1975). These samples mostly fictitious that we consider as the best representatives of different classes are then injected in the classification process to train the Kohonen maps. The adoption of Kohonen maps as a classification tool is justified simply by the use of maps in the self-adaptation of weights, which ultimately converge to an optimal configuration (Thomas, 2001).

The Kohonen maps have proved their effectiveness, especially in case of classification of weakly homogeneous images in which we note several weaknesses such as the lack of contrast and the presence of noise and blurring which feature the data to be.

This work plan also reduces significantly the complexity of the data analysis from 7 thematic mappers to only 3 thematic mappers from previous studies that reduced the size of the Landsat data space from 7 to 3 using the Principle Component Analysis (PCA).

In the end we validate the proposed approach that we adopted by applying two known measures (Validity Measure (Legany, 2006) and Davies-Bouldin Index (Davies, 1979)) to prove the contribution of the GA in the sampling phase of the segmentation process.

2. THE GENETIC ALGORITHMS

2.1. Introduction

Genetic algorithms are designed using Darwin's natural selection mechanism and Mendel's combination methods to solve problems and deal with difficult situations such as unforeseen situations, unknown environments, changing constraints induced by the environment.

This method works on a population of many different potential solutions. It eliminates the weakest elements to promote the conservation and reproduction of the most "performing" individuals (the fittest ones, the best adapted ones).

The recombination (reproduction by genetic hybridization) of the strongest individuals allows to give birth to even better individuals from one generation to the next.

Our strategy is based on taking advantage of these characteristics which make the strength of genetic algorithms to prepare good samples to the classification process. (Holland, 1975)

2.2. Design of Genetic Algorithms

Simplicity and efficiency are two of the most attractive features of GA approach. The implementation of a genetic algorithm needs (Harrat, 2003), (Goldberg, 1994):

1. A genetic representation of the problem, which means an appropriate coding of the solutions into chromosomes. This step associate to each point of the search space a data structure. It usually takes place after a phase of mathematical modeling of the problem. The quality of the data coding conditions the success of the genetic algorithms;
2. A mechanism to generate the initial population. This mechanism must be able to produce a non-homogeneous population that will serve as a basis for future generations. Choosing the initial population is important because it influences the speed of the convergence to the global optimum;

In the case where nothing is known about the problem to be solved, it is essential that the initial population should be distributed over the entire research area.

3. A dedicated fitness function to measure the force of each chromosome;
4. A mode of selection of the chromosomes to be reproduced;
5. Operators to diversify the population over generations and to explore the research space; the crossing operator recomposes the genes of the existing individuals in the population; the mutation operator aims to guarantee the exploration of the research space.
6. Values of the parameters used by the algorithm such as: Size of population, total number of generations or stopping criterion, crossing and mutation probabilities;

2.3. Genetic Algorithms Operating Steps

A genetic algorithm typically operates through a simple cycle of four steps (Harrat, 2003)

1. Creation of a population of chromosomes

2. Assessment of each chromosome
3. Selection of the best chromosomes
4. Genetic manipulation, to create a new population of chromosomes

The cycle described in Figure 1 is inspired by genetic terminology. During each cycle, a new generation of solutions of the problem is obtained. Initially, an initial population is generated where each individual-solution of the population is coded as a string of characters (chromosomes).

Then, an assessment of each chromosome will be established. This assessment is done by measuring the quality of the chromosomes using the assessment function (fitness). This it allows to select the most suitable chromosomes and thus apply the genetic operators (crossing and mutation), which creates a new generation.

At the end of the cycle, a new population is acquired, opening the way for a new generation and consequently a new cycle.

2.4. Algorithm Details

Preparation of samples is a very delicate operation, it must be carried out by an expert in remote sensing domain, because the quality of these samples influences without doubt the quality of the segmentation process. Bad specimens certainly lead to erroneous results despite the efficiency and potency of the classification method.

Our aim is to take full advantage of genetic algorithms to prepare the best and most optimal samples by minimizing the distance between pixels of the same population (candidate class) and maximizing the distance between the centroids of these populations.

These populations thus generated must always keep the same centroid resulting from the manual selection of the samples, it should be noted that the pixels generated during this phase are mostly fictitious.

The algorithm we have adopted uses the following real coding:

The individual is composed of 30 genes, each gene has a red or green or blue pixel value, 15 genes for urban samples (5 pixels) and 15 genes for non-urban samples. Each pixel has a value between 0 and 255. Figure 2 shows coding individuals.

2.5. The Stages of The Proposed GA Process

1. **Initialization:** Generate randomly a population of n individuals (pixels);
2. **Evaluation:** Calculate the fitness function $f(x)$ for each individual x of the population;
3. **Selection:** In our case we use the tournament selection: we choose two or more individuals and select the strongest. This process is repeated several times until obtaining N individuals ($N=50$).

The advantage of such selection is to prevent a very strong individual from being selected several times (Harrat, 2003).

Figure 1. Genetic cycle

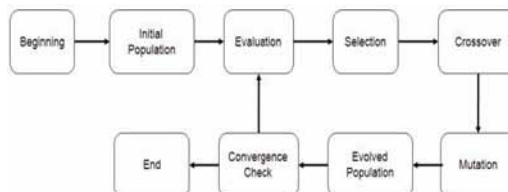
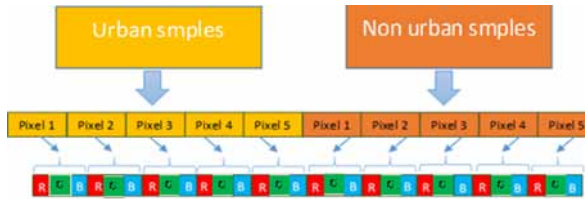


Figure 2. Coding individuals



4. **Crossover:** Apply the real cyclic crossover search operator on the parents with the probability of crossover P_c associated to give children. If there is no crossing, the children are the same copy of the parents (Hang, 2016)
5. **Mutation:** Apply the mutation operator P_m on children with the associated mutation probability
6. **Accept:** Place the children in the new population
7. **Replace:** Use the generated population to run the algorithm
8. **Test:** If the stop condition is satisfied, stop, and return the best solution. Otherwise go to step 2

2.6. Parameters Values

The values of such parameters depend strongly on the problem to be studied. Thus, there are no parameters that are suitable for solving all the problems that can be posed to a genetic algorithm. However, some values are often used (defined in the literature) and can be good starting points for starting a solution search using a GA.

Based on the observation that the parameter values of the various operators are themselves unknown and indefinitely improved only through experiments. Some authors, such as (Novkovic et al., 1997) have proposed to use a kind of Meta-GA: one to find the optimal individual and the other to find the optimal value of the parameters. These two algorithms would then rotate simultaneously or sequentially. However, it is inevitable that the calculation time will increase accordingly. (Thomas, 2004)

In 2003, the same authors (Nokovic, 2003) proposed a novel algorithm called Genetic Algorithm with self-generated Random parameters (GAR), rather than the programmer predetermine the parameter values, chromosome portions which do not translate into fitness “genetic residual” are given function to diversify control parameters for the GA, providing random parameter setting along the way, and doing away with fine-tuning of probabilities of crossover and mutation. (Nokovic, 2003)

1. **The Probability of Crossover:** Choosing the suitable value for the probability of crossover P_c as well as the probability of mutation P_m depends on the fitness function. Its choice is generally heuristic, more p_c is higher, and more the population undergoes major changes. In general, accepted values are between 0.5 and 0.9 (Thomas, 2001)
2. The mutation probability is selected in the interval (0.001, 0.01). P_m must be low since a high rate may lead to a suboptimal solution.
3. **The Fitness Function:** It is defined as the Euclidean distance between the two centroids or urban and non-urban regions $centroid1 (R1, G1, B1)$ and $centroid2 (R2, G2, B2)$ representing urban and non-urban population.

$$F = \sqrt{(R2 - R1)^2 + (G2 - G1)^2 + (B2 - B1)^2}$$

In this approach we empirically fix the crossover probability to 0.5 and the mutation rate at 0.0015.

3. THE KOHONEN MAPS

The Kohonen map also called “Self Organizing Maps: SOM” are an automatic classification technique (clustering, unsupervised learning), they were defined by Tuevo Kohonen in the late 1980’s (Kohonen, 1990). They aim to produce a grouping so that the individuals in the same box are similar, the individuals in different boxes are different. On closer inspection, we realize that the learning algorithm is a sophisticated version of the K-Means method.

The aim of unsupervised classification is to organize the data into homogeneous classes without supervised learning, each class regrouping individuals have similar features.

These networks are inspired by the biological observations of the functioning of the nervous systems of perception of mammals. Figure 3 indicates that the network consists of two layers, the first contains the inputs, and the second consists of the associated outputs. Each neuron of the latter is connected to all input neurons through n connections that carry the weights.

3.1. Operating Principle

This model uses competitive learning, it means that a single neuron will be activated, after the propagation, the latter will then be encouraged (it means that only the winning neuron according to a certain function, as well as in some cases, a certain neighborhood of this node, will see its weight modified). The winning neuron is calculated by choosing the one whose weight vector is closest to the input vector along Euclidean distance. Learning involves bringing the weights of the winning neurons closer to the input values.

Kohonen maps are useful in the case of automatic classification. Indeed, the neurons of the Kohonen layer are used as indicators of belonging to a given class.

For a certain neuron, the weight vector W associated with it indicates the representative of the class that this neuron has the task of symbolizing. This neuron will be activated only if the element presented to the network is closer to the representative of this class and the representatives of the other classes are more or less far away.

Of course, the problem of Kohonen maps to perform the automatic classification consists in fixing the number of classes and, at the same time, the number of neurons in the Kohonen layer.

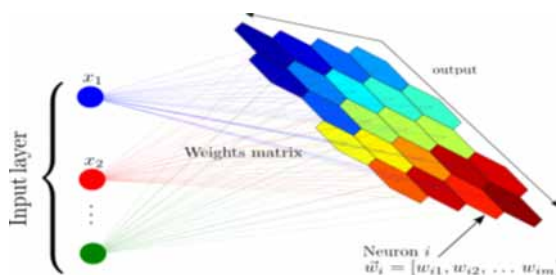
Concerning the dynamics, activation of the Kohonen network by applying the input vector X , amounts to finding the winning neuron j (search for the minimum Euclidean distance):

$$j(p) = \min_i \{ \|X - W_i(p)\| \}; i = 1, 2, \dots, n$$

The i runs through the output layer and n represents the number of neurons in the decision layer (output).

The learning is carried out in the following way (Jadouin, 1994)

Figure 3. Kohonen maps



$$\nabla W_{ik}(p+1) = \begin{cases} \alpha * [X_k - W_{ik}(p)], & i \in A_j(p) \\ 0, & i \notin A_j(p) \end{cases}$$

with:

$A_j(P)$ the neighborhood of the winning neuron j ,
 P : the number of the iteration,
 K : runs through all input neurons,
 α : the learning rate,
 W_i : the weight vector associated with the output neuron I .

3.2. The steps of Designing a Network

Beginners are often surprised to learn that to build a neural network, the first thing to do is not to choose the type of network but to choose its samples of learning, testing and validation data. Only then will the choice of the type of the network be made. In order to clarify the ideas, we present in the following the four steps that should guide the creation of a neural network.

3.2.1. Choosing Samples

The process of developing a neural network always begins with the selection and the preparation of data samples. As in data analysis, this step is crucial and will help the designer to determine the most appropriate network type to solve are problem. The way in which the sample is presented affects: the type of network, the number of input neurons, the number of output neurons, and how learning, testing and validation will be conducted.

3.2.2. Network Topology

The basic principles of the network process, inspired by the biological observations on the ordering of neurons, were introduced in 1981 by Teuvo Kohonen (Kohonen, 1990).

The neighborhood concept is often used to have a regular map of changes and to avoid noise phenomena. Thus, each entry of the neural network is not an isolated pixel, but a pixel accompanied by its neighborhood, (Alexandre, 2008)

The network we have adopted use a linear topology neighborhood radius equal to 1 in which each neuron (pixel) is connected to 2 neighbors.

3.2.3. Machine Learning

Learning consists first of all of calculating the optimal weights of the different links, using data samples.

3.2.4. Validation and Testing

The tests concerning the performance verification of the neuron network out of samples and its capacity of generalization, the validation is sometimes used during the learning. Once the network has been calculated, it is always necessary to check if our network reacts correctly. There are several methods of validation: cross validation, bootstrapping... but for the tests, generally, a part of the sample is simply removed from the learning sample and kept for testing out of sample. For example, 60% of the samples can be used for learning, 20% for validation and 20% for testing. In the case of small samples, such a distinction cannot always be used simply because it is not always possible to have sufficient data in each of the groups thus created. Procedures such as cross-validation are sometimes used to establish the optimal network structure.

3.3. The Stages of Kohonen Map Algorithm

1. **Initialization:** Synaptic weights should be randomly initiated;
Give a small positive value to the learning rate α ;
Set the number of iterations (stop criterion)
2. **Presentation:** Random selection of an entry x (pixel) in the learning database to present it to the network;
3. **Similarity:** Search for the winning neuron (x) using the minimal Euclidean distance criterion (nearest neuron or pixel);
4. **Updating:** Update the weights of neurons belonging to the topological neighborhood $\Lambda_{(x)}(n)$ following the adaptation step $\eta(n)$; where n is the iteration number;
Adjust the learning parameters $\Lambda_{(x)}(n)$ and $\eta(n)$.
5. **Continuation:** Return to step 2 until the iteration number criterion is reached.

3.4. The Advantages and Disadvantages of the SOM

For the advantages of neural networks, we can mention: (Hadjila, 2003)

- Automatic learning of weights (learning rules);
 - Generalization power (the testing phase);
 - Possibility of parallelism (the elements of each layer can operate in parallel);
 - Resistance to failure (if a neuron no longer works, the network does not disturb);
 - Modeling of unknown functions;
 - Good approximation.
- On the other hand, neural networks have some disadvantages: (Hadjila, 2003)
- Complex representation (usually the neural network architecture is complex, and even if it is schematized, it remains unreadable);
 - Difficulty to interpret parameters (black box);
 - Non-existent of rules defining the architecture of the network according to the problem to be solved;
 - The learning algorithms take a considerable amount of computing time.

4. REMOTE SENSING DATA SET

Reduce the volume of data by selecting the most useful data. The channels (7 for Landsat TM) of the same scene are often correlated with each other, resulting in information redundancy. The transformation is used to synthesize the information.

The computing of these new synthetic information layers is based on statistical analysis methods such as principal component analysis (PCA) and arithmetic combinations of channels (also called spectral ratios) leading to the creation of new channels.

The PCA is carried out according to traditional statistical methods, the particularity for remote sensing lies in the very big quantity of data to be processed (the pixels). The general approach consists in defining, in the multispectral space, new channels summarizing the information contained in the image. This method aims to maximize (statistically) the quantity of information (or variance) of the original data in a restricted number of components. (Thomas, 2001)

In our case we use Landsat5 thematic mapper images which have seven spectral bands with 30m spatial resolution. Subsequent studies have shown that the TM4, TM3 and TM1 channels hold the overwhelming majority of the information contained in the whole image (Christopher, 2004). Table 1 shows the Landsat5 TM bands.

Table 1. Landsat5 TM bands (Landsat, 2017)

	Spectral band	Resolution	Utilization
TM1	0,45-0,52 μm visible	30m x 30m	Differentiation ground / plants, coastal areas
TM2	0,52-0,60 μm visible	30m x 30m	Vegetation
TM3	0,63-0,69 μm Visible	30m x 30m	Differentiation of plant species
TM4	0,76-0,90 μm near infrared	30m x 30m	Biomass
TM5	1,55-1,75 μm near-infrared	30m x 30m	Differentiation snow/ clouds
TM6	10,4-12,5 μm Thermal	30m x 30m	Thermal
TM7	2,08-2,35 μm mid- infrared	30m x 30m	Lithology

4.1. Preprocessing and Reduction of the Size of the Data Space

In this study, the Landsat images used did not require radiometric or geometric corrections, insofar as they were previously selected for their quality and that the USGS automatically proceeds to the improvements of the images. Thus, the steps of correcting the distortions induced by the rotation of the Earth, its curvature, the pitch of the sensor, the orientation and the aspect ratio were avoided.

Our choice focused on TM4, TM3 and TM1 bands to highlight the urban subject on the satellite image, we judged that these three bands are the most useful and the most suitable to achieve our goal,

On one side we will reduce the space to be studied from 7 to 3 without losing a significant spectral information and on the other side it will facilitate the task of visualization in colored composition by affecting TM4 to Red, TM3 to Green and TM1 to Blue.

5. IMPLEMENTATION OF THE APPROACH

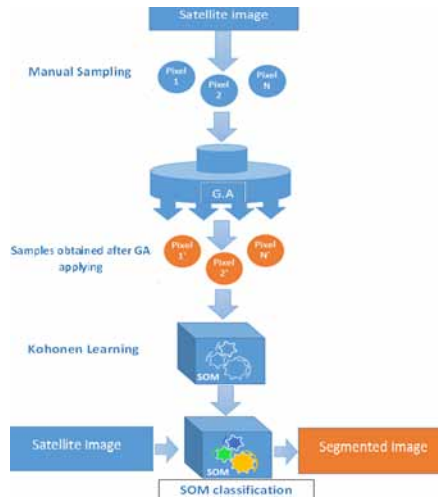
Our approach is based on the optimization of the samples by the genetic algorithms, these samples which are mostly fictitious and descending from real existing parents will be used to launch the classification by the Kohonen maps in order to highlight and isolate the urban theme on satellite images. Preparing samples of good quality and quantity is sometimes difficult; using GA technique has given a great benefit; we are not obliged to take a large quantity of pixels from each region to learn the kohonen maps. Taking some pixels (5 from urban region and 5 from non-urban region) was sufficient for us generate a good population of individuals which are mostly not real but they hold their credibility from the GA design, the latter technique is inspired from the natural evolution of the species.

Once the samples are ready, we will use them to learn the Kohonen maps.

5.1. The Diagram

The diagram below (Figure 4) shows the different phases of the clustering process of the satellite image from the input data (composite colored image) and the final clustered image.

Figure 4. General scheme of the proposed approach



5.2. Evaluation Tools

5.2.1. Validity Measure (VM)

VM is one of the clues to test the validity of clustering algorithms. VM is commonly used in the application of image segmentation based on clustering (Ariana, 2014; Chandok, 2012). VM is calculated using the following equation:

$$VM = y \left(\frac{intra}{inter} \right)$$

where *intra* is the intracluster distance, *inter* is the intercluster distance, and *y* is a function of the number of clusters. The following equation is used to find the value of the intra-cluster distance.

$$intra = \frac{1}{N} \sum_{i=1}^k \sum_{x \in c_i} x - z_{-i}^2$$

where *N* is the total number of pixels in the image, *k* is the cluster number, and *Z_i* is the cluster center *C_i*.

In addition to calculate *VM*, we take the minimum value of the inter-cluster distance (Ray, 1999). The following equation is used to find the inter-cluster distance.

$$inter = \min \left(\|z_i - z_j\| \right)$$

where *i* = 1, 2 ... *k*, and *j* = *i*+1, ..., *k*

y is multiplied by the quotient between the intra-cluster and inter-cluster distance. The following equation is used to calculate *y*.

$$y = c.N(2,1) + 1$$

where c is a constant value in the range 15 to 25. $N(2.1)$ is a Gaussian function for the number of clusters (k). The Gaussian is represented by the following equation.

$$N(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[-\frac{k-\mu^2}{2\sigma^2}\right]}$$

VM has to be at minimum to get an optimal result and therefore get well separated clusters (Palus, 2003).

5.2.2. Davies-Bouldin Index (DBI)

DBI was introduced in 1979 by David L. Davies and Donald W. DBI is used as a metric to evaluate the clustering algorithms. (Davies, 1979)

DBI is a function to measure the ratio of total dispersion within the cluster and separation between cluster (distance between clusters) (Maulik, 2002). The following equation is used to calculate cluster disparity.

$$S_i = \frac{1}{T} \sum_{x \in C_i} \|x - z_i\|$$

where T_i is the number of members in group (C_i), and Z_i is the center of i th cluster.

Distance between clusters is calculated by the Euclidean distance between the center of i th cluster and j th cluster, the following equation is used to calculate the distance.

$$d_{ij} = \|z_i - z_j\|$$

R_{ij} is the ratio value between i th and j th cluster which is computed by equation

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

Finding the maximum value of the ratio (D_i), will be used to find the value of DBI. The following equation explain how to calculate the value of D_i .

$$D_i = \max_{j:j \neq i} R_{ij}$$

Then the DBI value can be calculated using the next equation;

$$DBI = \frac{1}{k} \sum_{i=1}^k D_i$$

where K is the number of clusters.

In fact, the DBI indicates the optimality score of a clustering, lower it is, and more the clustering is optimal and gives well separated clusters.

6. RESULTS AND DISCUSSIONS

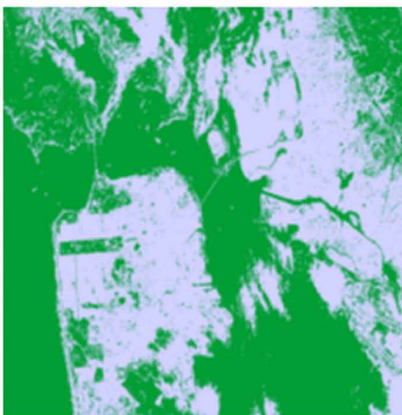
In the image segmentation N ° 1 (Figure 5), the use of Kohonen maps without the GA (image (b) in figure 5) gives a somewhat erroneous segmentation, and it is found that some forest zones have been classified as being urban area; and another part of the sea right at the edge of the ports was also allocated to urban areas; this is due to their direct proximity to urban areas, forests and the pollution of the sea in the proximity of the ports.

Image (c) (Figure 5) shows the result of the segmentation of the same image by coupling the GA and the Kohonen maps: the defects detected in the image (b) (Figure 5) are repaired, the rate of good classification has clearly increased, and it is proved by the calculated VM and DBI indices and shown in Table 2. This proves the effectiveness of the genetic algorithms coupling Kohonen's maps in the detection of urban areas on noisy and fuzzy satellite images.

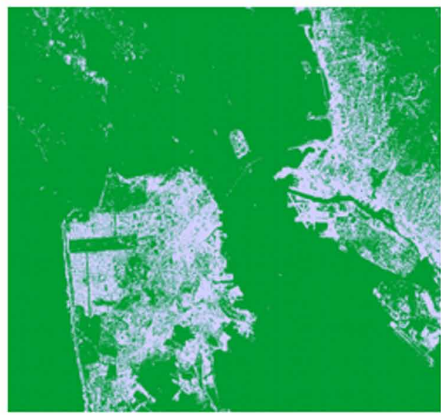
Figure 5. Segmentation of image N°1 (San Francisco city, USA)



(a) Original image.



(b) Segmented image using Kohonen maps without using GA.



(c) Segmented image using GA and Kohonen maps.

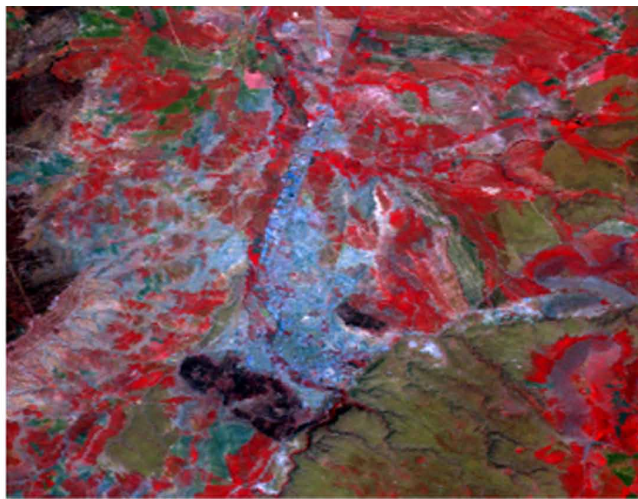
Concerning the image N ° 2 (Figure 6), which shows the result of the segmentation by the two techniques without and with the GA, we visually observe the clear failure of the first technique in the detection of various urban areas, false and large non-urban areas appear to be urban.

As for the approach we propose, it has considerably succeeded in minimizing the rate of misclassification without giving good results, this is due to the nature of the satellite image itself.

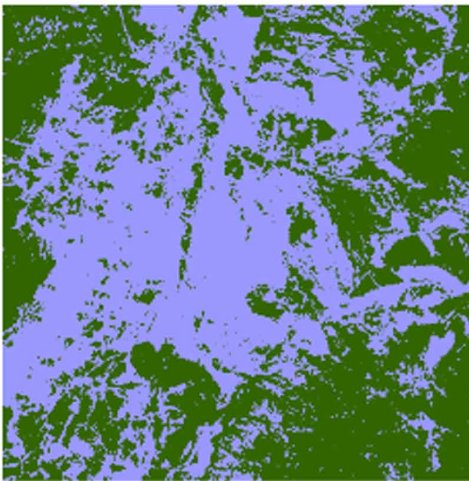
In fact, there has been a huge problem in acquiring Landsat satellite imagery of very good quality and updated on the region of SAIDA (ALGERIA). The measurements in Table 2 argue this comment.

Finally, the image segmentation N ° 3 (Figure 7) consolidates the choice we made, the couple GA-Kohonen maps have always succeeded in minimizing the classification errors, a simple visual analysis can confirm it and measurements are performed and displayed in Table 2.

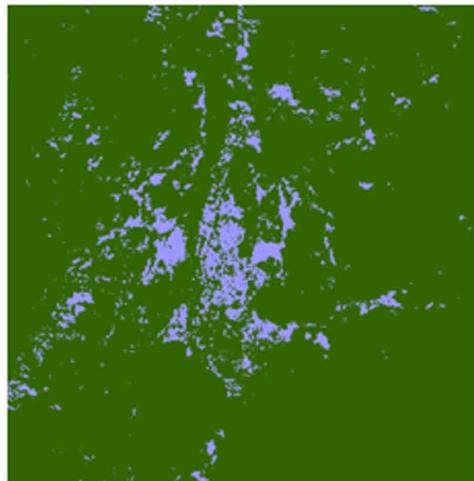
Figure 6. Segmentation of image N ° 2



(a) Original image



(b) Segmented image using Kohonen maps without using GA.

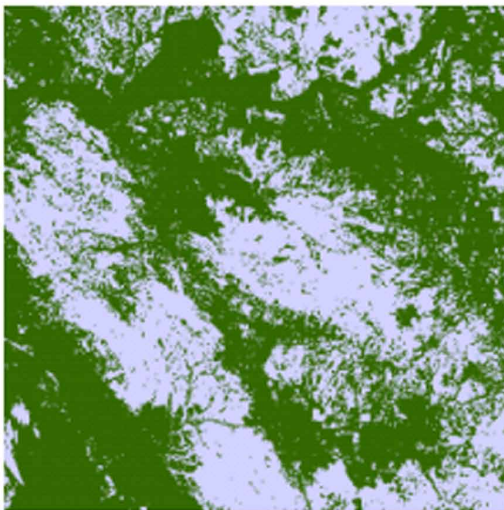


(c) Segmented image using GA and Kohonen maps.

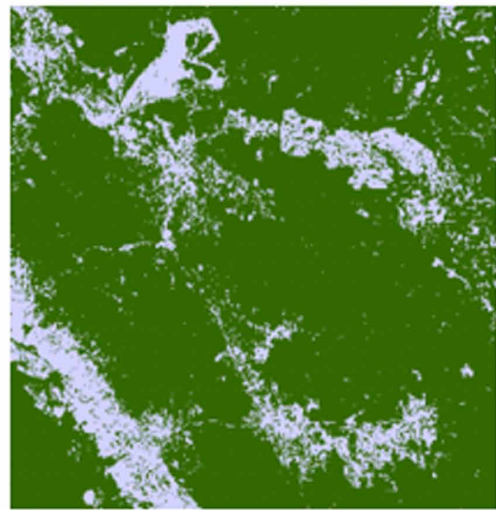
Figure 7. Segmentation of image N°3



(a) Original image



(b) segmented image using Kohonen maps without using GA.



(c) segmented image using GA and Kohonen maps.

7. CONCLUSION

In this paper we propose a combination of two techniques to highlight the urban zone on remote sensing images; the first one is the genetic algorithms (GA) in order to prepare an optimal sample population generated from real sample pixels picked the both urban and non-urban areas. This population mostly fictitious will be introduced as training samples for the second Kohonen maps technique.

Our work consists in improving the quality of the samples which strongly determine the quality of the classification by using the genetic algorithms since the selection of the first real samples are done manually. In this phase, we increase the sets of sample pixels by creating new fictitious pixels without modifying the initial centroid of the samples manually prepared. This operation aims to

Table 2. Segmentation assessment using VM and DBI indices

Image	Validity Measure (VM)		Davies-Bouldin Index (DBI)	
	Kohonen without GA	GA and Kohonen	Kohonen without GA	GA and Kohonen
N° 1	2.63	2.58	0.74	0.53
N° 2	4.68	3.71	0.91	0.71
N° 3	9.84	2.87	2.47	0.59

prepare a solid and homogeneous ground on which we have propel a classificatory process using Kohonen maps known by their self-adapting force.

We prove the effectiveness of the approach by applying it to satellite images of two cities (San Francisco, USA and Saida, Algeria). The obtained results show the contribution of the introduction of the GA into the segmentation process. To argue our comments, we used Validity Measure and Davies-Bouldin Index; the analysis of these indices shows the effectiveness of the proposed approach. In fact, each time we use the GA to prepare the samples, we obtain the best results.

Highlighting urban areas on multispectral images using the proposed approach aims to help decision makers and specialists in territorial planning to plan and asses the extension of the urban areas; we use satellite images as data sources because they offer a global vision on the ground in real and very short time.

REFERENCES

- Ariana, I., Sari Hartati, R., Putra, I., & Wirdiani, N. (2014). Color Image Segmentation using Kohonen Self-Organizing Map (SOM). *IACSIT International Journal of Engineering and Technology*.
- Chandok, C. (2012). Color Image Segmentation using K-Means Clustering. *International Journal of VLSI & Digital Signal Processing Application*.
- Chang, V. (2017). Data analytics and visualization for inspecting cancers and genes. *Multimedia Tools and Applications*.
- Chang, V. (2017). (in press). A proposed social network analysis platform for big data analytics. *Technological Forecasting and Social Change*. doi:10.1016/j.techfore.2017.11.002
- Chang, V. (2017). Towards data analysis for weather cloud computing. *Knowledge-Based Systems*, 127, 29–45. doi:10.1016/j.knsys.2017.03.003
- Chang, V. (2018). Computational Intelligence for Medical Imaging Simulations. *Journal of Medical Systems*, 42(1), 10. doi:10.1007/s10916-017-0861-x PMID:29177790
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1*(2), 224–227. doi:10.1109/TPAMI.1979.4766909 PMID:21868852
- Forgy, E. W. (1965). *Cluster analysis of multivariate data: efficiency versus interpretability of classifications*.
- Fournier, A. (2008). Détection et classification de changements sur des scènes urbaines en Télédétection. Interface homme-machine. Institut Supérieur de l’Aéronautique et de l’Espace, thèse doctorat.
- Goldberg, E. (1994). *Algorithmes génétiques Exploitation, optimisation et apprentissage automatique*. Edition Addison-Wesley.
- Gosh, A., & Shekhar, N. (2010). Fuzzy clustering algorithms for unsupervised change detection in remote sensing images. *Information Sciences*, 181(4), 699–715. doi:10.1016/j.ins.2010.10.016
- F. Hadjila et R. Bouabdellah. (2003). Reconnaissance des visages par les réseaux de neurones. mémoire Magister de l’université Abou Bakr Belkaid Tlemcen.
- Harrat, Y. (2003). Contribution à l’ordonnancement conjoint de la production et de la maintenance: Application au cas d’un job Shop [Thèse de Doctorat de L’U.F.R des Sciences et Techniques]. Université de Franche-comté.
- Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: Univ. of Michigan Press.
- Jadouin, J. F. (1994). Les réseaux de neurones: Principes et définitions “. *Hermes*.
- Kohonen, T. (1990). The self organizing map. *IEEE proceedings*, 78(9). doi:10.1109/5.58325
- Lefebvre, A., Corpetti, T., & Moy, L. H. (2009). *Détection de changements dans des images à très haute résolution spatiale par analyse de texture: application en milieu urbain*. Besancon: Neuvièmes Rencontres de ThéoQuant.
- Legany, C., & Juhasz, S., Babos. (2006). A.: Cluster validity measurement techniques. In *AIKED’06: Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases* (pp. 388-393).
- Leichtle, T., Geiß, C., Wurm, M., Lakes, T., & Taubenböck, H. (2017). Unsupervised change detection in VHR remote sensing imagery – an object-based clustering approach in a dynamic urban environment. *International Journal of Applied Earth Observation and Geoinformation*, 54, 15–27. doi:10.1016/j.jag.2016.08.010
- MacLachlan, A., & Biggs, E. (2017). *Urban Growth Dynamics in Perth*. Western Australia: Using Applied Remote Sensing for Sustainable Future Planning.
- Maulik, U., & Bandyopadhyay, S. (2002). Performance Evaluation of Some Clustering Algorithms and Validity Indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12), 1650–1654. doi:10.1109/TPAMI.2002.1114856
- Munyati, C. (2004). Use of Principal Component Analysis (PCA) of Remote Sensing Images in Wetland Change Detection on the Kafue Flats, Zambia. Geocarto International.

- Novkovic, S. & Sverko, D. (2003). A Genetic Algorithm With Self-Generated Random Parameters. *Journal of Computing and Information Technology*, 4, 271–283.
- Palus, H., & Bogdanski, M. (2003). Clustering techniques in color image segmentation. *In Proceedings of the 4th Symposium on Methods of Artificial Intelligence*.
- Ray, S., & Turi, R. (1999). Determination of number of clusters in k-means clustering and application in color image segmentation. *In Proceedings of the 4th international conference on advances in pattern recognition and digital techniques*.
- Roscher, R., & Waske, B. (2012). Incremental Import Vector Machines for classifying hyperspectral Data. *IEEE Transactions On Geoscience And Remote Sensing*, 50(9).
- Seiffert, U., & Jain, L. C. (Eds.), *Self-Organizing Maps*.
- Vallée, T. & Yildizoglu, M. (2004). Présentation des algorithmes génétiques et de leurs applications en économie.
- Villmann, T., & Merényi, E. (2001). Extensions and Modifications of the Kohonen-SOM and Applications in Remote Sensing Image Analysis.
- Xiajoun. (2011). Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment. Landsat. Retrieved 23/03/2017 from <https://landsat.usgs.gov/landsat-5-history> .
- Zhao, H., & Fansen, K. (2016). Research and Applications of Shop Scheduling Based on Genetic Algorithms. *Braz. arch. biol. technol.*, 59.

Mokadem Djelloul is a teacher researcher in Remote sensing, He works especially on remote sensing image information retrieval.

Abdelmalek Amine received an engineering degree in Computer Science from the Computer Science department of Djillali Liabes University of Sidi-Belabbes-Algeria, received the Magister diploma in Computational Science and PhD from Djillali Liabes University in collaboration with Joseph Fourier University of Grenoble. His research interests include big data, data mining, text mining, ontology, classification, clustering, neural networks, and biomimetic optimization methods. He participates in the program committees of several international conferences and on the editorial boards of international journals. Prof. Amine is the head of GeCoDe-knowledge management and complex data-laboratory at UTM University of Saida, Algeria; he also collaborates with the “knowledge base and database” team of TIMC laboratory at Joseph Fourier University of Grenoble.

David Helbert received his Master degree in Information Processing in 2002 and his Ph.D. degree in Signal and Image Processing and Computer Science at the University of Poitiers, France in 2005. He is Associate Professor at the University of Poitiers in the Xlim Institute, CNRS and he is the head of the image processing team ICONES. His main research interests are in image processing, multiresolution analysis, compressed sensing in various field such as image restoration and multispectral image analysis.